APPLIED PROBABILITY



Paul Pfeiffer
Rice University

LibreTexts™

Rice University

Applied Probability

Paul Pfeiffer

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden borne from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and were supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 10/25/2026

Preface

Preface to Pfeiffer Applied Probability

The course

This is a "first course" in the sense that it presumes no previous course in probability. The units are modules taken from the unpublished text: Paul E. Pfeiffer, ELEMENTS OF APPLIED PROBABILITY, USING MATLAB. The units are numbered as they appear in the text, although of course they may be used in any desired order. For those who wish to use the order of the text, an outline is provided, with indication of which modules contain the material.

The mathematical prerequisites are ordinary calculus and the elements of matrix algebra. A few standard series and integrals are used, and double integrals are evaluated as iterated integrals. The reader who can evaluate simple integrals can learn quickly from the examples how to deal with the iterated integrals used in the theory of expectation and conditional expectation. Appendix B provides a convenient compendium of mathematical facts used frequently in this work. And the symbolic toolbox, implementing MAPLE, may be used to evaluate integrals, if desired.

In addition to an introduction to the essential features of basic probability in terms of a precise mathematical model, the work describes and employs user defined MATLAB procedures and functions (which we refer to as *m-programs*, or simply *programs*) to solve many important problems in basic probability. This should make the work useful as a stand alone exposition as well as a supplement to any of several current textbooks.

Most of the programs developed here were written in earlier versions of MATLAB, but have been revised slightly to make them quite compatible with MATLAB 7. In a few cases, alternate implementations are available in the Statistics Toolbox, but are implemented here directly from the basic MATLAB program, so that students need only that program (and the symbolic mathematics toolbox, if they desire its aid in evaluating integrals).

Since machine methods require precise formulation of problems in appropriate mathematical form, it is necessary to provide some supplementary analytical material, principally the so-called *minterm analysis*. This material is not only important for computational purposes, but is also useful in displaying some of the structure of the relationships among events.

A probability model

Much of "real world" probabilistic thinking is an amalgam of intuitive, plausible reasoning and of statistical knowledge and insight. Mathematical probability attempts to to lend precision to such probability analysis by employing a suitable *mathematical model*, which embodies the central underlying principles and structure. A successful model serves as an aid (and sometimes corrective) to this type of thinking.

Certain concepts and patterns have emerged from experience and intuition. The mathematical formulation (the mathematical model) which has most successfully captured these essential ideas is rooted in measure theory, and is known as the *Kolmogorov model*, after the brilliant Russian mathematician A.N. Kolmogorov (1903-1987).

One cannot prove that a model is *correct*. Only experience may show whether it is *useful* (and not incorrect). The usefulness of the Kolmogorov model is established by examining its structure and showing that patterns of uncertainty and likelihood in any practical situation can be represented adequately. Developments, such as in this course, have given ample evidence of such usefulness.

The most fruitful approach is characterized by an interplay of

- A formulation of the problem in precise terms of the model and careful mathematical analysis of the problem so formulated.
- A grasp of the problem based on experience and insight. This underlies both problem formulation and interpretation of analytical results of the model. Often such insight suggests approaches to the analytical solution process.

MATLAB: A tool for learning

In this work, we make extensive use of MATLAB as an aid to analysis. I have tried to write the MATLAB programs in such a way that they constitute useful, ready-made tools for problem solving. Once the user understands the problems they are designed to solve, the solution strategies used, and the manner in which these strategies are implemented, the collection of programs should provide a useful resource.

However, my primary aim in exposition and illustration is to *aid the learning process* and to deepen insight into the structure of the problems considered and the strategies employed in their solution. Several features contribute to that end.

1. Application of machine methods of solution requires precise formulation. The data available and the fundamental assumptions must be organized in an appropriate fashion. The requisite *discipline* for such formulation often contributes to enhanced understanding of the problem.

2. The development of a MATLAB program for solution requires careful attention to possible solution strategies. One cannot instruct the machine without a clear grasp of what is to be done.

3. I give attention to the tasks performed by a program, with a general description of how MATLAB carries out the tasks. The reader is not required to trace out all the programming details. However, it is often the case that available MATLAB resources suggest alternative solution strategies. Hence, for those so inclined, attention to the details may be fruitful. I have included, as a separate collection, the m-files written for this work. These may be used as patterns for extensions as well as programs in MATLAB for computations. Appendix A provides a directory of these m-files.

4. Some of the details in the MATLAB script are presentation details. These are refinements which are not essential to the solution of the problem. But they make the programs more readily usable. And they provide illustrations of MATLAB techniques for those who may wish to write their own programs. I hope many will be inclined to go beyond this work, modifying current programs or writing new ones.

An Invitation to Experiment and Explore

Because the programs provide considerable freedom from the burden of computation and the tyranny of tables (with their limited ranges and parameter values), standard problems may be approached with a new spirit of experiment and discovery. When a program is selected (or written), it embodies one method of solution. There may be others which are readily implemented. The reader is invited, even urged, to explore! The user may experiment to whatever degree he or she finds useful and interesting. The possibilities are endless.

Acknowledgments

After many years of teaching probability, I have long since lost track of all those authors and books which have contributed to the treatment of probability in this work. I am aware of those contributions and am most eager to acknowledge my indebtedness, although necessarily without specific attribution.

The power and utility of MATLAB must be attributed to to the long-time commitment of Cleve Moler, who made the package available in the public domain for several years. The appearance of the professional versions, with extended power and improved documentation, led to further appreciation and utilization of its potential in applied probability.

The Mathworks continues to develop MATLAB and many powerful "tool boxes," and to provide leadership in many phases of modern computation. They have generously made available MATLAB 7 to aid in checking for compatibility the programs written with earlier versions. I have not utilized the full potential of this version for developing professional quality user interfaces, since I believe the simpler implementations used herein bring the student closer to the formulation and solution of the problems studied.

CONNEXIONS

The development and organization of the CONNEXIONS modules has been achieved principally by two people: C.S.(Sid) Burrus a former student and later a faculty colleague, then Dean of Engineering, and most importantly a long time friend; and Daniel Williamson, a music major whose keyboard skills have enabled him to set up the text (especially the mathematical expressions) with great accuracy, and whose dedication to the task has led to improvements in presentation. I thank them and others of the CONNEXIONS team who have contributed to the publication of this work.

Paul E. Pfeiffer

Rice University

This page titled Preface is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

TABLE OF CONTENTS

Preface

Licensing

1: Probability Systems

1.1: Likelihood

1.2: Probability Systems

1.3: Interpretations

1.4: Problems on Probability Systems

2: Minterm Analysis

2.1: Minterms

2.2: Minterms and MATLAB Calculations

2.3: Problems on Minterm Analysis

3: Conditional Probability

3.1: Conditional Probability

3.2: Problems on Conditional Probability

4: Independence of Events

4.1: Independence of Events

4.2: MATLAB and Independent Classes

4.3: Composite Trials

4.4: Problems on Independence of Events

5: Conditional Independence

5.1: Conditional Independence

5.2: Patterns of Probable Inference

5.3: Problems on Conditional Independence

6: Random Variables and Probabilities

6.1: Random Variables and Probabilities

6.2: Problems on Random Variables and Probabilities

7: Distribution and Density Functions

7.1: Distribution and Density Functions

7.2: Distribution Approximations

7.3: Problems on Distribution and Density Functions

8: Random Vectors and Joint Distributions

8.1: Random Vectors and Joint Distributions

8.2: Random Vectors and MATLAB

8.3: Problems on Random Vectors and Joint Distributions

9: Independent Classes of Random Variables

9.1: Independent Classes of Random Variables

9.2: Problems on Independent Classes of Random Variables

10: Functions of Random Variables

10.1: Functions of a Random Variable

10.2: Function of Random Vectors

10.3: The Quantile Function

10.4: Problems on Functions of Random Variables

11: Mathematical Expectation

11.1: Mathematical Expectation- Simple Random Variables

11.2: Mathematical Expectation and General Random Variables

11.3: Problems on Mathematical Expectation

12: Variance, Covariance, and Linear Regression

12.1: Variance

12.2: Covariance and the Correlation Coefficient

12.3: Linear Regression

12.4: Problems on Variance, Covariance, Linear Regression

13: Transform Methods

13.1: Transform Methods

13.2: Convergence and the Central Limit Theorem

13.3: Simple Random Samples and Statistics

13.4: Problems on Transform Methods

14: Conditional Expectation, Regression

14.1: Conditional Expectation, Regression

14.2: Problems on Conditional Expectation, Regression

15: Random Selection

15.1: Random Selection

15.2: Some Random Selection Problems

15.3: Problems on Random Selection

16: Conditional Independence, Given a Random Vector

16.1: Conditional Independence, Given a Random Vector

16.2: Elements of Markov Sequences

16.3: Problems on Conditional Independence, Given a Random Vector

17: Appendices

17.1: Appendix A to Applied Probability- Directory of m-functions and m-procedures

17.2: Appendix B to Applied Probability- some mathematical aids

17.3: Appendix C- Data on some common distributions

17.4: Appendix D to Applied Probability- The standard normal distribution

17.5: Appendix E to Applied Probability - Properties of Mathematical Expectation
17.6: Appendix F to Applied Probability- Properties of conditional expectation, given a random vector
17.7: Appendix G to Applied Probability- Properties of conditional independence, given a random vector
17.8: Matlab files for "Problems" in "Applied Probability"

Index

Glossary

Detailed Licensing

This page titled Applied Probability (Pfeiffer) is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

This page titled Licensing is shared under a license and was authored, remixed, and/or curated by LibreTexts.

1: PROBABILITY SYSTEMS

1.1: Likelihood
1.2: Probability Systems
1.3: Interpretations
1.4: Problems on Probability Systems

This page titled 1: Probability Systems is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Likelihood

Introduction

Probability models and techniques permeate many important areas of modern life. A variety of types of random processes, reliability models and techniques, and statistical considerations in experimental work play a significant role in engineering and the physical sciences. The solutions of management decision problems use as aids decision analysis, waiting line theory, inventory theory, time series, cost analysis under uncertainty — all rooted in applied probability theory. Methods of statistical analysis employ probability analysis as an underlying discipline.

Modern probability developments are increasingly sophisticated mathematically. To utilize these, the practitioner needs a sound conceptual basis which, fortunately, can be attained at a moderate level of mathematical sophistication. There is need to develop a feel for the structure of the underlying mathematical model, for the role of various types of assumptions, and for the principal strategies of problem formulation and solution.

Probability has roots that extend far back into antiquity. The notion of "chance" played a central role in the ubiquitous practice of gambling. But chance acts were often related to magic or religion. For example, there are numerous instances in the Hebrew Bible in which decisions were made "by lot" or some other chance mechanism, with the understanding that the outcome was determined by the will of God. In the New Testament, the book of Acts describes the selection of a successor to Judas Iscariot as one of "the Twelve." Two names, Joseph Barsabbas and Matthias, were put forward. The group prayed, then drew lots, which fell on Matthias.

Early developments of probability as a mathematical discipline, freeing it from its religious and magical overtones, came as a response to questions about games of chance played repeatedly. The mathematical formulation owes much to the work of Pierre de Fermat and Blaise Pascal in the seventeenth century. The game is described in terms of a well defined trial (a play); the result of any trial is one of a specific set of distinguishable outcomes. Although the result of any play is not predictable, certain "statistical regularities" of results are observed. The possible results are described in ways that make each result seem equally likely. If there are N such possible "equally likely" results, each is assigned a probability $1/N$.

The developers of mathematical probability also took cues from early work on the analysis of statistical data. The pioneering work of John Graunt in the seventeenth century was directed to the study of "vital statistics," such as records of births, deaths, and various diseases. Graunt determined the fractions of people in London who died from various diseases during a period in the early seventeenth century. Some thirty years later, in 1693, Edmond Halley (for whom the comet is named) published the first life insurance tables. To apply these results, one considers the selection of a member of the population on a chance basis. One then assigns the probability that such a person will have a given disease. The trial here is the selection of a person, but the interest is in certain characteristics. We may speak of the event that the person selected will die of a certain disease– say "consumption." Although it is a person who is selected, it is death from consumption which is of interest. Out of this statistical formulation came an interest not only in probabilities as fractions or relative frequencies but also in averages or expectatons. These averages play an essential role in modern probability.

We do not attempt to trace this history, which was long and halting, though marked by flashes of brilliance. Certain concepts and patterns which emerged from experience and intuition called for clarification. We move rather directly to the mathematical formulation (the "mathematical model") which has most successfully captured these essential ideas. This is the model, rooted in the mathematical system known as measure theory, is called the *Kolmogorov model*, after the brilliant Russian mathematician A.N. Kolmogorov (1903-1987). Kolmogorov succeeded in bringing together various developments begun at the turn of the century, principally in the work of E. Borel and H. Lebesgue on measure theory. Kolmogorov published his epochal work in German in 1933. It was translated into English and published in 1956 by Chelsea Publishing Company.

Outcomes and events

Probability applies to situations in which there is a well defined trial whose possible outcomes are found among those in a given basic set. The following are typical.

- A pair of dice is rolled; the outcome is viewed in terms of the numbers of spots appearing on the top faces of the two dice. If

the outcome is viewed as an ordered pair, there are thirty six equally likely outcomes. If the outcome is characterized by the total number of spots on the two die, then there are eleven possible outcomes (not equally likely).
- A poll of a voting population is taken. Outcomes are characterized by responses to a question. For example, the responses may be categorized as positive (or favorable), negative (or unfavorable), or uncertain (or no opinion).
- A measurement is made. The outcome is described by a number representing the magnitude of the quantity in appropriate units. In some cases, the possible values fall among a finite set of integers. In other cases, the possible values may be any real number (usually in some specified interval).
- Much more sophisticated notions of outcomes are encountered in modern theory. For example, in communication or control theory, a communication system experiences only one signal stream in its life. But a communication system is not designed for a single signal stream. It is designed for one of an infinite *set* of possible signals. The likelihood of encountering a certain kind of signal is important in the design. Such signals constitute a subset of the larger set of all possible signals.

These considerations show that our probability model must deal with

- A *trial* which results in (selects) an *outcome* from a *set* of conceptually possible outcomes. The trial is not successfully completed until one of the outcomes is realized.
- Associated with each outcome is a certain characteristic (or combination of characteristics) pertinent to the problem at hand. In polling for political opinions, it is a person who is selected. That person has many features and characteristics (race, age, gender, occupation, religious preference, preferences for food, etc.). But the primary feature, which characterizes the outcome, is the political opinion on the question asked. Of course, some of the other features may be of interest for analysis of the poll.

Inherent in informal thought, as well as in precise analysis, is the notion of an *event* to which a *probability* may be assigned as a measure of the *likelihood* the event will *occur* on any trial. A successful mathematical model must formulate these notions with precision. An *event* is identified in terms of the characteristic of the outcome observed. The event “a favorable response” to a polling question *occurs* if the outcome observed has that characteristic; i.e., iff (if and only if) the respondent replies in the affirmative. A hand of five cards is drawn. The event “one or more aces” *occurs* iff the hand actually drawn has at least one ace. If that same hand has two cards of the suit of clubs, then the event “two clubs” has *occurred*. These considerations lead to the following definition.

Definition. The *event* determined by some characteristic of the possible outcomes is the set of those outcomes having this characteristic. The event *occurs* iff the outcome of the trial is a member of that set (i.e., has the characteristic determining the event).

- The event of throwing a “seven” with a pair of dice (which we call the event SEVEN) consists of the set of those possible outcomes with a total of seven spots turned up. The event SEVEN occurs iff the outcome is one of those combinations with a total of seven spots (i.e., belongs to the event SEVEN). This could be represented as follows. Suppose the two dice are distinguished (say by color) and a picture is taken of each of the thirty six possible combinations. On the back of each picture, write the number of spots. Now the event SEVEN consists of the set of all those pictures with seven on the back. Throwing the dice is equivalent to selecting randomly one of the thirty six pictures. The event SEVEN occurs iff the picture selected is one of the set of those pictures with seven on the back.
- Observing for a very long (theoretically infinite) time the signal passing through a communication channel is equivalent to selecting one of the conceptually possible signals. Now such signals have many characteristics: the maximum peak value, the frequency spectrum, the degree of differentibility, the average value over a given time period, etc. If the signal has a peak absolute value less than ten volts, a frequency spectrum essentially limited from 60 herz to 10,000 herz, with peak rate of change 10,000 volts per second, then it is *one* of the *set* of signals with those characteristics. The event "the signal has these characteristics" has occured. This set (event) consists of an uncountable infinity of such signals.

One of the advantages of this formulation of an event as a subset of the basic set of possible outcomes is that we can use elementary set theory as an aid to formulation. And tools, such as Venn diagrams and indicator functions for studying event combinations, provide powerful aids to establishing and visualizing relationships between events. We formalize these ideas as follows:

- Let Ω be the set of all possible outcomes of the basic trial or experiment. We call this the *basic space* or the *sure event*, since if the trial is carried out successfully the outcome will be in Ω; hence, the event Ω is sure to occur on any trial. We must specify unambiguously what outcomes are “possible.” In flipping a coin, the only accepted outcomes are “heads” and “tails.” Should the coin stand on its edge, say by leaning against a wall, we would ordinarily consider that to be the result of an improper trial.
- As we note above, each outcome may have several characteristics which are the basis for describing events. Suppose we

are drawing a single card from an ordinary deck of playing cards. Each card is characterized by a "face value" (two through ten, jack, queen, king, ace) and a "suit" (clubs, hearts, diamonds, spades). An ace is drawn (the event ACE occurs) iff the outcome (card) belongs to the set (event) of four cards with ace as face value. A heart is drawn iff the card belongs to the set of thirteen cards with heart as suit. Now it may be desirable to specify events which involve various logical combinations of the characteristics. Thus, we may be interested in the event the face value is jack *or* king *and* the suit is heart *or* spade. The set for jack *or* king is represented by the union $J \cup K$ and the set for heart *or* spade is the union $H \cup S$. The occurrence of both conditions means the outcome is in the intersection (common part) designated by $\cap$. Thus the event referred to is

$$E = (J \cup K) \cap (H \cup S)$$

The notation of set theory thus makes possible a precise formulation of the event E.

Sometimes we are interested in the situation in which the outcome does *not* have one of the characteristics. Thus the set of cards which does not have suit heart is the set of all those outcomes not in event *H* . In set theory, this is the *complementary* set (event) H^c.
Events are *mutually exclusive* iff not more than one can occur on any trial. This is the condition that the sets representing the events are disjoint (i.e., have no members in common).
The notion of the *impossible event* is useful. The impossible event is, in set terminology, the *empty set* $\emptyset$. Event $\emptyset$ cannot occur, since it has no members (contains no outcomes). One use of $\emptyset$ is to provide a simple way of indicating that two sets are mutually exclusive. To say $AB = \emptyset$ (here we use the alternate AB for $A \cap B$) is to assert that events A and B have no outcome in common, hence cannot both occur on any given trial.
The language and notaton of sets provide a precise language and notation for events and their combinations. We collect below some useful facts about logical (often called Boolean) combinations of events (as sets). The notion of Boolean combinations may be applied to arbitrary classes of sets. For this reason, it is sometimes useful to use an index set to designate membership. We say the index J is countable if it is finite or countably infinite; otherwise it is uncountable. In the following it may be arbitrary.

$A_i : i \in J$ is the class of sets A_i, one for each index i in the index set J

For example, if $J = 1, 2, 3$ then $A_i : i \in J$ is the class A_1, A_2, A_3, and

$$\bigcup_{i \in J} A_i = A_1 \cup A_2 \cup A_3, \bigcup_{i \in J} A_i = A_1 \cap A_2 \cap A_3,$$

If $J = 1, 2, \cdots$ then $A_i : i \in J$ is the sequence $A_1 : 1 \leq i$, and

$$\bigcup_{i \in J} A_i = \bigcup_{i=1}^{\infty} A_i, \bigcap_{i \in J} A_i = \bigcap_{i=1}^{\infty} A_i$$

If event *E* is the *union* of a class of events, then event *E* occurs iff *at least one* event in the class occurs. If *F* is the *intersection* of a class of events, then event *F* occurs iff *all* events in the class occur on the trial.

The role of disjoint unions is so important in probability that it is useful to have a symbol indicating the union of a disjoint class. We use the big V to indicate that the sets combined in the union are disjoint. Thus, for example, we write

$A = \bigvee_{i=1}^{n} A_i$ to signify $A = \bigcup_{i=1}^{n} A_i$ with the proviso that the A_i form a disjoint class

Events derived from a class

Consider the class E_1, E_2, E_3 of events. Let A_k be the event that exactly k occur on a trial and B_k be the event that k or more occur on a trial. Then

$$A_0 = E_1^c E_2^c E_3^c, A_1 = E_1 E_2^c E_3^c \bigvee E_1^c E E_3^c \bigvee E_1^c E_2^c E_3, A_2 = E_1 E_2 E_3^c \bigvee E_1 E_2^c E_3 \bigvee E_1^c E_2 E_3, A_3 = E_1 E_2 E_3$$

The unions are disjoint since each pair of terms has E_i in one and E_i^c in the other, for at least one i. Now the B_k can be expressed in terms of the \(A_k\. For example

$$V_2 = A_2 \bigvee A_3$$

The union in this expression for B_2 is disjoint since we cannot have exactly two of the E_i occur *and* exactly three of them occur on the same trial. We may express B_2 directly in terms of the E_i as follows:

$$B_2 = E_1E_2 \cup E_1E_3 \cup E_2E_3$$

Here the union is not disjoint, in general. However, if one pair, say E_1, E_3 is disjoint, then $E_1E_3 = \emptyset$ and the pair E_1E_2, E_2E_3 is disjoint (draw a Venn diagram). Suppose C is the event the first two occur or the last two occur but no other combination. Then

$$C = E_1E_2E_3^c \bigvee E_1^cE_2E_3$$

Let D be the event that one or three of the events occur,

$$D = A_1 \bigvee A_3 = E_1E_2^cE_3^c \bigvee E_1^cE_2E_3^c \bigvee E_1^cE_2^cE_3 \bigvee E_1E_2E_3$$

The important patterns in set theory known as DeMorgan's rules are useful in the handing of events. For an arbitrary class $A_i : i \in J$ of events,

$$[\bigcup_{i\in J} A_i]^c = \bigcap_{i\in J} A_i^c \text{ and } [\bigcap_{i\in J} A_i]^c = \bigcup_{i\in J} A_i^c$$

An outcome is not in the union (i.e., not in at least one) of the A_i iff it fails to be in all A_i, and it is not in the intersection (i.e. not in all) iff it fails to be in at least one of the A_i.

continuation of example

Express the event of no more than one occurrence of the events in E_1, E_2, E_3 as B_2^c.

$$B_2^c = [E_1E_2 \cup E_1E_3 \cup E_2E_3]^c = (E_1^c \cup E_2^c)(E_1^c \cup E_3^c)(E_2^3 \cup E_3^c) = E_1^cE_2^c \cup E_1^cE_3^c \cup E_2^cE_3^c$$

The last expression shows that not more than one of the E_i occurs iff at least two of them fail to occur.

This page titled 1.1: Likelihood is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Probability Systems

Probability measures

In the module "Likelihood" we introduce the notion of a basic space ΩΩ of all possible outcomes of a trial or experiment, events as subsets of the basic space determined by appropriate characteristics of the outcomes, and logical or Boolean combinations of the events (unions, intersections, and complements) corresponding to logical combinations of the defining characteristics.

Occurrence or nonoccurrence of an event is determined by characteristics or attributes of the outcome observed on a trial. Performing the trial is visualized as selecting an outcome from the basic set. An event occurs whenever the selected outcome is a member of the subset representing the event. As described so far, the selection process could be quite deliberate, with a prescribed outcome, or it could involve the uncertainties associated with “chance.” Probability enters the picture only in the latter situation. Before the trial is performed, there is *uncertainty* about which of these latent possibilities will be realized. *Probability* traditionally is a number assigned to an event indicating the *likelihood* of the occurrence of that event on any trial.

We begin by looking at the *classical* model which first successfully formulated probability ideas in mathematical form. We use modern terminology and notation to describe it.

Definition: Classical Probability

1. The basic space Ω consists of a finite number *N* of possible outcomes.

   -There are thirty six possible outcomes of throwing two dice.
   -There are $C(52,5) = \frac{52!}{5!47!} = 2598960$ different hands of five cards (order not important).
   -There are $2^5 = 32$ results (sequences of heads or tails) of flipping five coins.
2. Each possible outcome is assigned a probability 1/N
3. If event (subset) A has N_A elements, then the probability assigned event A is

$$P(A) = N_A/N \text{ (i.e., the fraction favorable to } A)$$

With this definition of probability, each event A is assigned a unique probability, which may be determined by counting N_A, the number of elements in A (in the classical language, the number of outcomes "favorable" to the event) and N the total number of possible outcomes in the *sure* event Ω.

Probabilities for hands of cards

Consider the experiment of drawing a hand of five cards from an ordinary deck of 52 playing cards. The number of outcomes, as noted above, is $N = C(52,5) = 2598960N = C(52,5) = 2598960$. What is the probability of drawing a hand with exactly two aces? What is the probability of drawing a hand with two or more aces? What is the probability of not more than one ace?

Solution

Let A be the event of exactly two aces, B be the event of exactly three aces, and C be the event of exactly four aces. In the first problem, we must count the number N_A of ways of drawing a hand with two aces. We select two aces from the four, and select the other three cards from the 48 non aces. Thus

$$N_A = C(4,2)C(48,3) = 103776, \text{ so that } P(A) = \frac{N_A}{N} = \frac{103776}{2598960} \approx 0.0399$$

There are two or more aces iff there are exactly two or exactly three or exactly four. Thus the event D of two or more is $D = A \bigvee B \bigvee C$, since A, B, C are mutually exclusive,

$$N_D = N_A + N_b + N_c = C(4,2)C(48,3) + C(4,3)C(48,2) + C(4,4)C(48,1) = 103776 + 4512 + 48 = 108336$$

so that $P(D) \approx 0.0417$. There is one ace or none iff there are not two or more aces. We thus want $P(D^c)$. Now the number in D_c is the number not in D which is $N - N_D$, so that

$$P(D^c) = \frac{N - N_D}{N} = 1 - \frac{N_D}{N} = 1 - P(D) = 0.9583$$

This example illustrates several important properties of the classical probability.

$P(A) = N_A/N$ is a nonnegative quantity.
$P(\Omega) = N/N = 1$
If A, B, C are *mutually exclusive*, then the number in the disjoint union is the sum of the numbers in the individual events, so that

$$P(A \bigvee B \bigvee C) = P(A) + P(B) + P(C)$$

Several other elementary properties of the classical probability may be identified. It turns out that they can be derived from these three. Although the classical model is highly useful, and an extensive theory has been developed, it is not really satisfactory for many applications (the communications problem, for example). We seek a more general model which includes classical probability as a special case and is thus an extension of it. We adopt the *Kolmogorov model* (introduced by the Russian mathematician A. N. Kolmogorov) which captures the essential ideas in a remarkably successful way. Of course, no model is ever completely successful. Reality always seems to escape our logical nets.

The Kolmogorov model is grounded in abstract measure theory. A full explication requires a level of mathematical sophistication inappropriate for a treatment such as this. But most of the concepts and many of the results are elementary and easily grasped. And many technical mathematical considerations are not important for applications at the level of this introductory treatment and may be disregarded. We borrow from measure theory a few key facts which are either very plausible or which can be understood at a practical level. *This enables us to utilize a very powerful mathematical system for representing practical problems in a manner that leads to both insight and useful strategies of solution.*

Our approach is to begin with the notion of events as sets introduced above, then to introduce probability as a number assigned to events subject to certain conditions which become definitive properties. Gradually we introduce and utilize additional concepts to build progressively a powerful and useful discipline. The *fundamental properties needed* are just those illustrated in Example for the classical case.

Definition

A probability system consists of a basic set Ω of elementary outcomes of a trial or experiment, a class of events as subsets of the basic space, and a *probability measure* $P(\cdot)$ which assigns values to the events in accordance with the following rules

(P1): For any event A, the probability $P(A) \geq 0$.
(P2): The probability of the sure event $P(\Omega) = 1$.
(P3): *Countable additivity*. If $A_i : 1 \in J$ is a mutually exclusive, countable class of events, then the probability of the disjoint union is the sum of the individual probabilities.

The necessity of the mutual exclusiveness (disjointedness) is illustrated in Example. If the sets were not disjoint, probability would be counted more than once in the sum. A probability, as defined, is abstract—simply a number assigned to each set representing an event. But we can give it an interpretation which helps to visualize the various patterns and relationships encountered. We may think of *probability* as *mass* assigned to an event. The total unit mass is assigned to the basic set Ω. The additivity property for disjoint sets makes the mass interpretation consistent. We can use this interpretation as a precise representation. Repeatedly we refer to the probability mass assigned a given set. The mass is proportional to the weight, so sometimes we speak informally of the weight rather than the mass. Now a mass assignment with three properties does not seem a very promising beginning. But we soon expand this rudimentary list of properties. We use the mass interpretation to help visualize the properties, but are primarily concerned to interpret them in terms of likelihoods.

(P4): $P(A^c) = 1 - P(A)$. The follows from additivity and the fact that

$$1 = P(\Omega) = P(A \bigvee A^c) = P(A) + P(A^c)$$

(P5): $P(\emptyset) = 0$. The empty set represents an impossible event. It has no members, hence cannot occur. It seems reasonable that it should be assigned zero probability (mass). Since $\emptyset = \Omega^c$, this follows logically from P(4) and (P2).

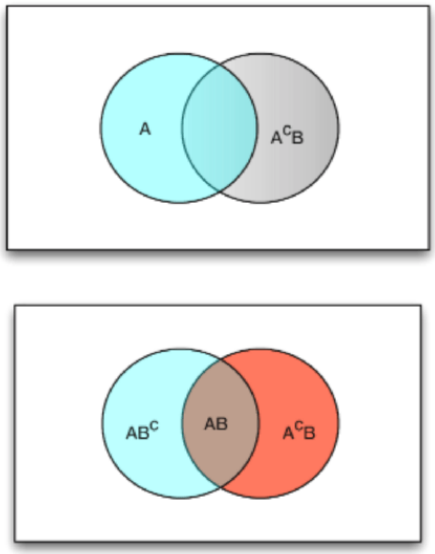


Figure 1.2.1: Partitions of the union $A \cup B$

(P6): If $A \subset B$, then $P(A) \leq P(B)$. From the mass point of view, every point in A is also in B, so that B must have at least as much mass as A. Now the relationship $A \subset B$ means that if A occurs, B must also. Hence B is at least as likely to occur as A. From a purely formal point of view, we have

$$B = A \bigvee A^c B \text{ so that } P(B) = P(A) + P(A^c B) \geq P(A) \text{ since } P(A^c B) \geq 0$$

(P7):$P(A \cup B) = P(A) + P(A^c B) = P(B) + P(AB^c) = P(AB^c) + P(AB) + P(A^c B)$

$= P(A) + P(B) - P(AB)$

The first three expressions follow from additivity and partitioning of $A \cup B$ as follows (see Figure 1.2.1).

$$A \cup B = A \bigvee A^c B = B \bigvee AB^c = AB^c \bigvee AB \bigvee A^c B$$

If we add the first two expressions and subtract the third, we get the last expression. In terms of probability mass, the first expression says the probability in $A \cup B$ is the probability mass in A plus the additional probability mass in the part of B which is not in A. A similar interpretation holds for the second expression. The third is the probability in the common part plus the extra in A and the extra in B. If we add the mass in A and B we have counted the mass in the common part twice. The last expression shows that we correct this by taking away the extra common mass.

(P8): If $B_i : i \in J$ is a countable, disjoint class and A is contained in the union, then

$$A = \bigvee_{i \in J} AB_i \text{ so that } P(A) = \sum_{i \in J} P(AB_i)$$

(P9): *Subadditivity*. If $A = \bigcup_{i=1}^{\infty} A_i$, then $P(A) \leq \sum_{i=1}^{\infty} P(A_i)$. This follows from countable additivity, property (P6), and the fact

(Partitions)

$A = \bigcup_{i=1}^{\infty} A_i = \bigvee_{i=1}^{\infty} B_i$, where $B_i = A_i A_1^c A_2^c \cdots A_{i-1}^c \subset A_i$

This includes as a special case the union of a finite number of events.

Some of these properties, such as (P4), (P5), and (P6), are so elementary that it seems they should be included in the defining statement. This would not be incorrect, but would be inefficient. If we have an assignment of numbers to the events, we need only establish (P1), (P2), and (P3) to be able to assert that the assignment constitutes a probability measure. And the other properties follow as logical consequences.

Flexibility at a price

In moving beyond the classical model, we have gained great flexibility and adaptability of the model. It may be used for systems in which the number of outcomes is infinite (countably or uncountably). It does not require a uniform distribution of the probability mass among the outcomes. For example, the dice problem may be handled directly by assigning the appropriate probabilities to the various numbers of total spots, 2 through 12. As we see in the treatment of conditional probability, we make new probability assignments (i.e., introduce new probability measures) when partial information about the outcome is obtained.

But this freedom is obtained at a price. In the classical case, the probability value to be assigned an event is clearly defined (although it may be very difficult to perform the required counting). In the general case, we must resort to experience, structure of the system studied, experiment, or statistical studies to assign probabilities.

The existence of uncertainty due to “chance” or “randomness” does not necessarily imply that the act of performing the trial is haphazard. The trial may be quite carefully planned; the contingency may be the result of factors beyond the control or knowledge of the experimenter. The mechanism of chance (i.e., the source of the uncertainty) may depend upon the nature of the actual process or system observed. For example, in taking an hourly temperature profile on a given day at a weather station, the principal variations are not due to experimental error but rather to unknown factors which converge to provide the specific weather pattern experienced. In the case of an uncorrected digital transmission error, the cause of uncertainty lies in the intricacies of the correction mechanisms and the perturbations produced by a very complex environment. A patient at a clinic may be self selected. Before his or her appearance and the result of a test, the physician may not know which patient with which condition will appear. In each case, from the point of view of the experimenter, the cause is simply attributed to “chance.” Whether one sees this as an “act of the gods” or simply the result of a configuration of physical or behavioral causes too complex to analyze, the situation is one of uncertainty, before the trial, about which outcome will present itself.

If there were complete uncertainty, the situation would be chaotic. But this is not usually the case. While there is an extremely large number of possible hourly temperature profiles, a substantial subset of these has very little likelihood of occurring. For example, profiles in which successive hourly temperatures alternate between very high then very low values throughout the day constitute an unlikely subset (event). One normally expects trends in temperatures over the 24 hour period. Although a traffic engineer does not know exactly how many vehicles will be observed in a given time period, experience provides some idea what range of values to expect. While there is uncertainty about which patient, with which symptoms, will appear at a clinic, a physician certainly knows approximately what fraction of the clinic's patients have the disease in question. In a game of chance, analyzed into “equally likely” outcomes, the assumption of equal likelihood is based on knowledge of symmetries and structural regularities in the mechanism by which the game is carried out. And the number of outcomes associated with a given event is known, or may be determined.

In each case, there is some basis in statistical data on past experience or knowledge of structure, regularity, and symmetry in the system under observation which makes it possible to assign likelihoods to the occurrence of various events. It is this ability to assign likelihoods to the various events which characterizes applied probability. However determined, *probability is a number assigned to events to indicate their likelihood of occurrence*. The assignments must be consistent with the defining properties (P1), (P2), (P3) along with derived properties (P4) through (P9) (plus others which may also be derived from these). Since the probabilities are not “built in,” as in the classical case, a prime role of probability theory is to *derive other probabilities* from a set of *given probabilites*.

This page titled 1.2: Probability Systems is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Interpretations

What is Probability?

The formal probability system is a *model* whose usefulness can only be established by examining its structure and determining whether patterns of uncertainty and likelihood in any practical situation can be represented adequately. With the exception of the sure event and the impossible event, the model does not tell us how to assign probability to any given event. The formal system is consistent with many probability assignments, just as the notion of mass is consistent with many different mass assignments to sets in the basic space.

The defining properties (P1), (P2), (P3) and derived properties provide *consistency rules* for making probability assignments. One cannot assign negative probabilities or probabilities greater than one. The sure event is assigned probability one. If two or more events are mutually exclusive, the total probability assigned to the union must equal the sum of the probabilities of the separate events. Any assignment of probability consistent with these conditions is allowed.

One may not know the probability assignment to every event. Just as the defining conditions put constraints on allowable probability assignments, they also provide important structure. *A typical applied problem provides the probabilities of members of a class of events (perhaps only a few) from which to determine the probabilities of other events of interest*. We consider an important class of such problems in the next chapter.

There is a variety of points of view as to how probability should be interpreted. These impact the manner in which probabilities are assigned (or assumed). One important dichotomy among practitioners.

- One group believes probability is *objective* in the sense that it is something inherent in the nature of things. It is to be discovered, if possible, by analysis and experiment. Whether we can determine it or not, "it is there."
- Another group insists that probability is a *condition of the mind* of the person making the probability assessment. From this point of view, the laws of probability simply impose rational consistency upon the way one assigns probabilities to events. Various attempts have been made to find objective ways to measure the *strength of one*'s *belief* or *degree of certainty* that an event will occur. The probability $P(A)$ expresses the degree of certainty one feels that event *A* will occur. One approach to characterizing an individual's degree of certainty is to equate his assessment of $P(A)$ with the amount *a* he is willing to pay to play a game which returns one unit of money if *A* occurs, for a gain of $(1 - a)$, and returns zero if *A* does not occur, for a gain of $-a$. Behind this formulation is the notion of a *fair game*, in which the "expected" or "average" gain is zero.

The early work on probability began with a study of *relative frequencies* of occurrence of an event under repeated but independent trials. This idea is so imbedded in much intuitive thought about probability that some probabilists have insisted that it must be built into the definition of probability. This approach has not been entirely successful mathematically and has not attracted much of a following among either theoretical or applied probabilists. In the model we adopt, there is a fundamental limit theorem, known as *Borel*'s *theorem*, which may be interpreted "if a trial is performed a large number of times in an independent manner, the *fraction* of times that event A occurs *approaches as a limit the value* $P(A)$. Establishing this result (which we do not do in this treatment) provides a formal validation of the intuitive notion that lay behind the early attempts to formulate probabilities. Inveterate gamblers had noted long-run statistical regularities, and sought explanations from their mathematically gifted friends. From this point of view, probability is meaningful only in repeatable situations. Those who hold this view usually assume an objective view of probability. It is a number determined by the nature of reality, to be discovered by repeated experiment.

There are many applications of probability in which the relative frequency point of view is not feasible. Examples include predictions of the weather, the outcome of a game or a horse race, the performance of an individual on a particular job, the success of a newly designed computer. These are unique, nonrepeatable trials. As the popular expression has it, "You only go around once." Sometimes, probabilities in these situations may be quite subjective. As a matter of fact, those who take a subjective view tend to think in terms of such problems, whereas those who take an objective view usually emphasize the frequency interpretation.

Subjective probability and a football game

The probability that one's favorite football team will win the next Superbowl Game may well be only a subjective probability of the bettor. This is certainly not a probability that can be determined by a large number of repeated trials. The game is only played once. However, the subjective assessment of probabilities may be based on intimate knowledge of relative strengths and weaknesses of the teams involved, as well as factors such as weather, injuries, and experience. There may be a considerable objective basis for the subjective assignment of probability. In fact, there is often a hidden “frequentist” element in the subjective evaluation. There is an assessment (perhaps unrealized) that in *similar situations* the frequencies tend to coincide with the value subjectively assigned.

The probabilty of rain

Newscasts often report that the probability of rain of is 20 percent or 60 percent or some other figure. There are several difficulties here.

- To use the formal mathematical model, there must be precision in determining an event. An event either occurs or it does not. How do we determine whether it has rained or not? Must there be a measurable amount? Where must this rain fall to be counted? During what time period? Even if there is agreement on the area, the amount, and the time period, there remains ambiguity: one cannot say with logical certainty the event did occur or it did not occur. Nevertheless, in this and other similar situations, use of the concept of an event may be helpful even if the description is not definitive. There is usually enough practical agreement for the concept to be useful.
- What does a 30 percent probability of rain mean? Does it mean that if the prediction is correct, 30 percent of the area indicated will get rain (in an agreed amount) during the specified time period? Or does it mean that 30 percent of the occasions on which such a prediction is made there will be significant rainfall in the area during the specified time period? Again, the latter alternative may well hide a frequency interpretation. Does the statement mean that it rains 30 percent of the times when conditions are *similar* to current conditions?

Regardless of the interpretation, there is some ambiguity about the event and whether it has occurred. And there is some difficulty with knowing how to interpret the probability figure. While the precise meaning of a 30 percent probability of rain may be difficult to determine, it is generally useful to know whether the conditions lead to a 20 percent or a 30 percent or a 40 percent probability assignment. And there is no doubt that as weather forecasting technology and methodology continue to improve the weather probability assessments will become increasingly useful.

Another common type of probability situation involves determining the distribution of some characteristic over a population—usually by a survey. These data are used to answer the question: What is the probability (likelihood) that a member of the population, chosen “at random” (i.e., on an equally likely basis) will have a certain characteristic?

Empirical probability based on survey data

A survey asks two questions of 300 students: Do you live on campus? Are you satisfied with the recreational facilities in the student center? Answers to the latter question were categorized “reasonably satisfied,” “unsatisfied,” or “no definite opinion.” Let C be the event “on campus;” O be the event “off campus;” S be the event “reasonably satisfied;” U be the event ”unsatisfied;” and N be the event “no definite opinion.” Data are shown in the following table.

Survey Data

	S	U	N
C	127	31	42
O	46	43	11

If an individual is selected on an equally likely basis from this group of 300, the probability of any of the events is taken to be the relative frequency of respondents in each category corresponding to an event. There are 200 on campus members in the population, so $P(C) = 200/300$ and $P(O) = 100/300$. The probability that a student selected is on campus and satisfied is taken to be $P(CS) = 127/300$. The probability a student is either on campus and satisfied or off campus and not satisfied is

$$P(CS \bigvee OU) = P(CS) + P(OU) = 127/300 + 43/300 = 170/300$$

If there is reason to believe that the population sampled is representative of the entire student body, then the same probabilities would be applied to any student selected at random from the entire student body.

It is fortunate that we do not have to declare a single position to be the "correct" viewpoint and interpretation. The formal model is consistent with *any* of the views set forth. We are free in any situation to make the *interpretation most meaningful and natural to the problem at hand*. It is not necessary to fit all problems into one conceptual mold; nor is it necessary to change mathematical model each time a different point of view seems appropriate.

Probability and odds

Often we find it convenient to work with a ratio of probabilities. If A and B are events with positive probability the *odds* favoring A over B is the probability ratio $P(A)P(B)$. If not otherwise specified, B is taken to be A^c and we speak of the *odds* favoring A

$$O(A) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

This expression may be solved algebraically to determine the probability from the odds

$$P(A) = \frac{O(A)}{1 + O(A)}$$

In particular, if $O(A) = a/b$ then $P(A) = \dfrac{a/b}{1 + a/b} = \dfrac{a}{a + b}$.

$O(A) = 0.7/0.3 = 7/3$. If the odds favoring A is 5/3, then $P(A) = 5/(5 + 3) = 5/8$.

Partitions and Boolean combinations of events

The countable additivity property (P3) places a premium on appropriate partitioning of events.

Definition

A *partition* is a mutually exclusive class

$$A_i : i \in J \text{ such that } \Omega = \bigvee_{i \in J} A_i$$

A *partition* of event A is a mutually exclusive class

$$A_i : i \in J \text{ such that } A = \bigvee_{i \in J} A_i$$

Remarks.

- A partition is a mutually exclusive class of events such that one (and only one) must occur on each trial.
- A partition of event A is a mutually exclusive class of events such that A occurs iff one (and only one) of the A_i occurs.
- A partition (no qualifier) is taken to be a partition of the sure event Ω.
- If class $B_i : i \in J$ is mutually exclusive and $A \subset B = \bigvee_{i \in J} B_i$, then the class $AB_i : i \in J$ is a partition of A and $A = \bigvee_{i \in J} AB_i$.

We may begin with a sequence $A_1 : 1 \leq i$ and determine a mutually exclusive (disjoint) sequence $B_1 : 1 \leq i$ as follows:

$B_1 = A_1$, and for any $i > 1$, $B_i = A_i A_1^c A_2^c \cdots A_{i-1}^c$

Thus each B_i is the set of those elements of A_i not in any of the previous members of the sequence.

This representation is used to show that subadditivity (P9) follows from countable additivity and property (P6). Since each $B_i \subset A_i$, by (P6) $P(B_i) \leq P(A_i)$. Now

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigvee_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i)$$

The representation of a union as a disjoint union points to an important *strategy* in the solution of probability problems. If an event can be expressed as a countable disjoint union of events, each of whose probabilities is known, then the probability of the combination is the sum of the individual probailities. In in the module on Partitions and Minterms, we show that *any* Boolean combination of a *finite*class of events can be expressed as a disjoint union in a manner that often facilitates systematic determination of the probabilities.

The indicator function

One of the most useful tools for dealing with set combinations (and hence with event combinations) is the *indicator function* I_E for a set $E \subset \Omega$. It is defined very simply as follows:

$$I_E(\omega) = \begin{cases} 1 & \text{for } \omega \in E \\ 0 & \text{for } \omega \in E^c \end{cases}$$

Remark. Indicator fuctions may be defined on any domain. We have occasion in various cases to define them on the real line and on higher dimensional Euclidean spaces. For example, if M is the interval $[a, b]$ on the real line then $I_M(t) = 1$ for each t in the interval (and is zero otherwise). Thus we have a step function with unit value over the interval M. In the abstract basic space Ω we cannot draw a graph so easily. However, with the representation of sets on a Venn diagram, we can give a schematic representation, as in Figure 1.3.1.

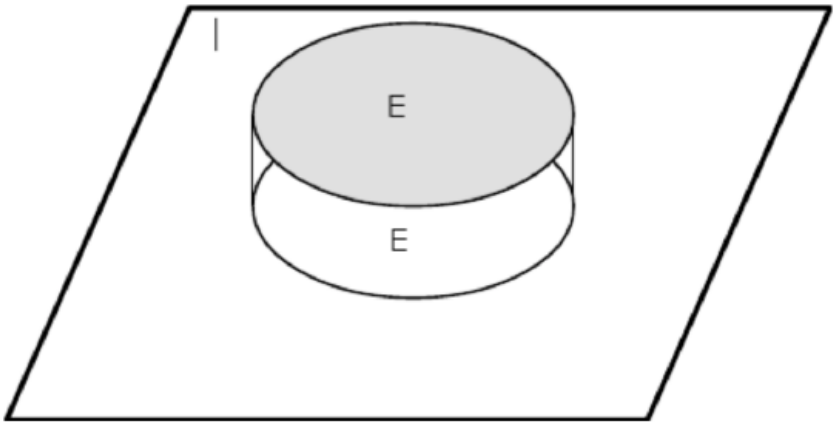


Figure 1.3.1. Representation of the indicator function I_E for event E.

Much of the usefulness of the indicator function comes from the following properties.

(IF1): $I_A \leq I_B$ iff $A \subset B$. If $I_A \leq I_B$, then $\omega \in A$ implies $I_A(\omega) = I_B(\omega) = 1$, so $\omega \in B$, then $I_A(\omega) = 1$ implies $\omega \in A$ implies $\omega \in B$ implies $I_B(\omega) = 1$.
(IF2): $I_A = I_B$ iff $A = B$

$$A = B \text{ iff both } A \subset B \text{ and } B \subset A \text{ iff } I_A \leq I_B \text{ and } I_B \leq I_A \text{ iff } I_A = I_B$$

(IF3): $I_{A^c} = 1 - I_A$ This follows from the fact $I_{A^c}(\omega) = 1$ iff $I_A(\omega) = 0$.
(IF4): $I_{AB} = I_A I_B = \min I_A, I_B$ (extends to any class) An element ω belongs to the intersection iff it belongs to all iff the indicator function for each event is one iff the product of the indicator functions is one.
(IF5): $I_{A \cup B} = I_A + I_B - I_A I_B = \min I_A, I_B$ (the maximum rule extends to any class) The maximum rule follows from the fact

that ω is in the union iff it is in any one or more of the events in the union iff any one or more of the individual indicator function has value one iff the maximum is one. The sum rule for two events is established by DeMorgan's rule and properties (IF2), (IF3), and (IF4).

$I_{A \cup B} = 1 - I_{A^c B^c} = 1 - [1 - I_A][1 - I_B] = 1 - 1 + I_B + I_A - I_A I_B$

(IF6): If the pair A, B is disjoint, $I_{A \bigvee B} = I_A + I_B$ (extends to any disjoint class)

The following example illustrates the use of indicator functions in establishing relationships between set combinations. Other uses and techniques are established in the module on Partitions and Minterms.

Indicator functions and set combinations

Suppose $A_i : 1 \leq i \leq n$ is a partition.

$$\text{If } B = \bigvee_{i=1}^{n} A_i C_i, \text{ then } B^c = \bigvee_{i=1}^{n} A_i C_i^c$$

Proof

Utilizing properties of the indicator function established above, we have

$$I_B = \sum_{i=1}^{n} I_{A_i} I_{C_i}$$

Note that since the A_i form a partition, we have $\sum_{i=1}^{n} I_{A_i} = 1$, so that the indicator function for the complementary event is

$$I_{B^c} = 1 - \sum_{i=1}^{n} I_{A_i} I_{C_i} = \sum_{i=1}^{n} I_{A_i} - \sum_{i=1}^{n} I_{A_i} I_{C_i} = \sum_{i=1}^{n} [1 - I_{C_i}] = \sum_{i=1}^{n} I_{A_i} I_{C_i^c}$$

The last sum is the indicator function for $\bigvee_{i=1}^{n} A_i C_i^c$

A technical comment on the class of events

The class of events plays a central role in the intuitive background, the application, and the formal mathematical structure. Events have been modeled as subsets of the basic space of all possible outcomes of the trial or experiment. In the case of a finite number of outcomes, *any* subset can be taken as an event. In the general theory, involving infinite possibilities, there are some technical mathematical reasons for limiting the class of subsets to be considered as events. The practical needs are these:

1. If A is an event, its complementary set must also be an event.
2. If $A_i : i \in J$ is a finite or countable class of events, the union and the intersection of members of the class need to be events.

A simple argument based on DeMorgan's rules shows that if the class contains complements of all its sets and countable unions, then it contains countable intersections. Likewise, if it contains complements of all its sets and countable intersections, then it contains countable unions. A class of sets closed under complements and countable unions is known as a *sigma algebra* of sets. In a formal, measure-theoretic treatment, a basic assumption is that the class of events is a sigma algebra and the probability measure assigns probabilities to members of that class. Such a class is so general that it takes very sophisticated arguments to establish the fact that such a class does not contain all subsets. But precisely because the class is so general and inclusive *in ordinary applications we need not be concerned about which sets are permissible as events*

A primary task in formulating a probability problem is identifying the appropriate events and the relationships between them. The theoretical treatment shows that we may work with great freedom in forming events, with the assurrance that in most applications a set so produced is a mathematically valid event. The so called *measurability* question only comes into play in dealing with random processes with continuous parameters. Even there, under reasonable assumptions, the sets produced will be events.

This page titled 1.3: Interpretations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

1.4: Problems on Probability Systems

Exercise 1.4.1

Let Ω consist of the set of positive integers. Consider the subsets

$A = \{\omega : \omega \leq 12\}$ $B = \{\omega : \omega < 8\}$ $C = \{\omega : \omega \text{ is even}\}$

$D = \{\omega : \omega \text{ is a multiple of } 3\}$ $E = \{\omega : \omega \text{ is a multiple of } 4\}$

Describe in terms of A, B, C, D, E and their complements the following sets:

a. {1, 3, 5, 7}
b. {3, 6, 9}
c. {8, 10}
d. The even integers greater than 12
e. The positive integers which are multiples of six.
f. The integers which are even and no greater than 6 or which are odd and greater than 12.

Answer available at the end of this section.

Exercise 1.4.2

Let Ω be the set of integers 0 through 10. Let $A = \{5, 6, 7, 8\}$, $B =$ the odd integers in Ω, and $C =$ the integers in Ω which are even or less than three. Describe the following sets by listing their elements.

a. AB
b. AC
c. $AB^c \cup C$
d. ABC^c
e. $A \cup B^c$
f. $A \cup BC^c$
g. ABC
h. A^cBC^c

Answer available at the end of this section.

Exercise 1.4.3

Consider fifteen-word messages in English. Let $A =$ the set of such messages which contain the word "bank" and let $B =$ the set of messages which contain the word "bank" *and* the word "credit." Which event has the greater probability? Why?

Answer available at the end of this section.

Exercise 1.4.4

A group of five persons consists of two men and three women. They are selected one-by-one in a random manner. Let E_i be the event a man is selected on the ith selection. Write an expression for the event that both men have been selected by the third selection.

Answer available at the end of this section.

Exercise 1.4.5

Two persons play a game consecutively until one of them is successful or there are ten unsuccessful plays. Let E_i be the event of a success on the ith play of the game. Let A, B, C be the respective events that player one, player two, or neither wins. Write an expression for each of these events in terms of the events E_i, $1 \leq i \leq 10$.

Answer available at the end of this section.

Exercise 1.4.6

Suppose the game in Exercise 1.4.5 could, in principle, be played an unlimited number of times. Write an expression for the event D that the game will be terminated with a success in a finite number of times. Write an expression for the event F that the game will never terminate.

Answer available at the end of this section.

Exercise 1.4.7

Find the (classical) probability that among three random digits, with each digit (0 through 9) being equally likely and each triple equally likely:

a. All three are alike.
b. No two are alike.
c. The first digit is 0.
d. Exactly two are alike.

Answer available at the end of this section.

Exercise 1.4.8

The classical probability model is based on the assumption of equally likely outcomes. Some care must be shown in analysis to be certain that this assumption is good. A well known example is the following. Two coins are tossed. One of three outcomes is observed: Let ω_1 be the outcome both are "heads," ω_2 the outcome that both are "tails," and ω_3 be the outcome that they are different. Is it reasonable to suppose these three outcomes are equally likely? What probabilities would you assign?

Answer available at the end of this section.

Exercise 1.4.9

A committee of five is chosen from a group of 20 people. What is the probability that a specified member of the group will be on the committee?

Answer available at the end of this section.

Exercise 1.4.10

Ten employees of a company drive their cars to the city each day and park randomly in ten spots. What is the (classical) probability that on a given day Jim will be in place three? There are $n!$ equally likely ways to arrange n items (order important).

Answer available at the end of this section.

Exercise 1.4.11

An extension of the classical model involves the use of areas. A certain region L (say of land) is taken as a reference. For any subregion A, define $P(A) = area(A)/area(L)$. Show that $P(\cdot)$ is a probability measure on the subregions of L.

Answer available at the end of this section.

Exercise 1.4.12

John thinks the probability the Houston Texans will win next Sunday is 0.3 and the probability the Dallas Cowboys will win is 0.7 (they are not playing each other). He thinks the probability both will win is somewhere between—say, 0.5. Is that a reasonable assumption? Justify your answer.

Answer available at the end of this section.

Exercise 1.4.13

Suppose $P(A) = 0.5$ and $P(B) = 0.3$. What is the largest possible value of $P(AB)$? Using the maximum value of $P(AB)$, dertermine $P(AB^c)$, $P(A^cB)$, $P(A^cB^c)$ and $P(A \cup B)$. Are these values determined uniquely?

Answer available at the end of this section.

Exercise 1.4.14

For each of the following probability "assignments", fill out the table. Which assignments are not permissible? Explain why, in each case.

$P(A)$	$P(B)$	$P(AB)$	$P(A \cup B)$	$P(AB^c)$	$P(A^cB)$	$P(A) + P(B)$
0.3	0.7	0.4				
0.2	0.1	0.4				
0.3	0.7	0.2				
0.3	0.5	0				
0.3	0.8	0				

Answer available at the end of this section.

Exercise 1.4.15

The class $\{A, B, C\}$ of events is a partition. Event A is twice as likely as C and event B is as likely as the combination A or C. Determine the probabilities $P(A)$, $P(B)$, $P(C)$.

Answer available at the end of this section.

Exercise 1.4.16

Determione the probability $P(A \cup B \cup C)$ in terms of the probabilities of the events A, B, C and their intersections.

Answer available at the end of this section.

Exercise 1.4.17

If occurrence of event A implies occurrence of B, show that $P(A^cB) = P(B) - P(A)$.

Answer available at the end of this section.

Exercise 1.4.18

Show that $P(AB) \geq P(A) + P(B) - 1$.

Answer available at the end of this section.

Exercise 1.4.19

The set combination $A \oplus B = AB^c \bigvee A^cB$ is known as the *disjunctive union* or the *symetric difference* of A and B. This is the event that only one of the events A or B occurs on a trial. Determine $P(A \oplus B)$ in terms of $P(A)$, $P(B)$, and $P(AB)$

Answer available at the end of this section.

Exercise 1.4.20

Use fundamental properties of probability to show

a. $P(AB) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$

b. $P(\bigcap_{j=1}^{\infty} E_j) \leq P(E_i) \leq P(\bigcup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} P(E_j)$

Answer available at the end of this section.

Exercise 1.4.21

Suppose P_1, P_2 are probability measures and c_1, c_2 are positive numbers such that $c_1 + c_2 = 1$. Show that the assignment $P(E) = c_1P_1(E) + c_2P_2(E)$ to the class of events is a probability measure. Such a combination of probability measures is known as a *mixture*. Extend this to

$P(E) = \sum_{i=1}^{n} c_iP_i(E)$, where the P_i are probabilities measures, $c_i > 0$, and $\sum_{i=1}^{n} c_i = 1$

Answer available at the end of this section.

Exercise 1.4.22

Suppose $\{A_1, A_2, \cdots, A_n\}$ is a partition and $\{c_1, c_2, \cdots, c_n\}$ is a class of positive constants. For each event E, let

$$Q(E) = \sum_{i=1}^{n} c_iP(EA_i) / \sum_{i=1}^{n} c_iP(A_i)$$

Show that $Q(\cdot)$ us a probability measure.

Answer available at the end of this section.

Answers

Exercise 1.4.1

$a = BC^c$
$b = DAE^c$
$c = CAB^cD^c$
$d = CA^c$
$e = CD$
$f = BC \bigvee A^cC^c$

Exercise 1.4.2

a. $AB = 5, 7$
b. $AC = 6, 8$
c. $AB^c \cup C = C$
d. $ABC^c = AB$
e. $A \cup B^c = 0, 2, 4, 5, 6, 7, 8, 10$
f. $ABC = \emptyset$
g. $A^cBC^c = 3, 9$

Exercise 1.4.3

$B \subset A$ implies $P(B) \leq P(A)$.

Exercise 1.4.4

$A = E_1E_2 \bigvee E_1E_2^cE_3 \bigvee E_1^cE_2E_3$

Exercise 1.4.5

$A = E_1 \bigvee E_1^cE_2^cE_3 \bigvee E_1^cE_2^cE_3^cE_4^cE_5 \bigvee E_1^cE_2^cE_3^cE_4^cE_5^cE_6^cE_7 \bigvee E_1^cE_2^cE_3^cE_4^cE_5^cE_6^cE_7^cE_8^cE_9$
$B = E_1^cE_2 \bigvee E_1^cE_2^cE_3^cE_4 \bigvee E_1^cE_2^cE_3^cE_4^cE_5^cE_6 \bigvee E_1^cE_2^cE_3^cE_4^cE_5^cE_6^cE_7^cE_8 \bigvee E_1^cE_2^cE_3^cE_4^cE_5^cE_6^cE_7^cE_8^cE_9E_{10}$
$C = \bigcap_{i=1}^{10} E_i^c$

Exercise 1.4.6

Let $F_0 = \Omega$ and $F_k = \bigcap_{i=1}^{k} E_i^c$ for $k \geq 1$. Then

$$D = \bigvee_{n=1}^{\infty} F_{n-1}E_n \text{ and } F = D^c = \bigcap_{i=1}^{\infty} E_i^c$$

Exercise 1.4.7

Each triple has probability $1/10^3 = 1/1000$

a. Ten triples, all alike: $P = 10/1000$.
b. $10 \times 9 \times 8$ triples all different: $P = 720/1000$.
c. 100 triples with first one zero: $P = 100/1000$
d. $C(3, 2) = 3$ ways to pick two positions alike; 10 ways to pick the common value; 9 ways to pick the other. $P = 270/1000$.

Exercise 1.4.8

$P(\{\omega_1\}) = P(\{\omega_2\}) = 1/4$, $P(\{\omega_3\}) = 1/2$

Exercise 1.4.9

$C(20, 5)$ committees; $C(19, 4)$ have a designated member.

$$P = \frac{19!}{4!15!} \cdot \frac{5!15!}{20!} = 5/20 = 1/4$$

Exercise 1.4.10

10! permutations,$1 \times 9!$ permutations with Jim in place 3. $P = 9!/10! = 1/10$.

Exercise 1.4.11

Additivity follows from additivity of areas of disjoint regions.

Exercise 1.4.12

$P(AB) = 0.5$ is not reasonable. It must no greater than the minimum of $P(A) = 0.3$ and $P(B) = 0.7$.

Exercise 1.4.13

Draw a Venn diagram, or use algebraic expressions $P(AB^c) = P(A) - P(AB) = 0.2$
$P(A^cB) = P(B) - P(AB) = 0$ $P(A^cB^c) = P(A^c) - P(A^cB) = 0.5$ $P(A \cup B) = 0.5$

Exercise 1.4.14

$P(A)$	$P(B)$	$P(AB)$	$P(A \cup B)$	$P(AB^c)$	$P(A^cB)$	$P(A) + P(B)$
0.3	0.7	0.4	0.6	-0.1	0.3	1.0
0.2	0.1	0.4	-0.1	-0.2	-0.3	0.3
0.3	0.7	0.2	0.8	0.1	0.5	1.0
0.3	0.5	0	0.8	0.3	0.5	0.8
0.3	0.8	0	1.1	0.3	0.8	1.1

Only the third and fourth assignments are permissible.

Exercise 1.4.15

$P(A) + P(B) + P(C) = 1$, $P(A) = 2P(C)$, and $P(B) = P(A) + P(C) = 3P(C)$, which implies

$$P(C) = 1/6,\ P(A) = 1/3,\ P(B) = 1/2$$

Exercise 1.4.16

$P(A \cup B \cup C) = P(A \cup B) + P(C) - P(AC \cup BC)$
$= P(A) + P(B) - P(AB) + P(C) - P(AC) - P(BC) + P(ABC)$

Exercise 1.4.17

$P(AB) = P(A)$ and $P(AB) + P(A^cB) = P(B)$ implies $P(A^cB) = P(B) - P(A)$.

Exercise 1.4.18

Follows from $P(A) + P(B) - P(AB) = P(A \cup B) \leq 1$.

Exercise 1.4.19

A Venn diagram shows $P(A \oplus B) = P(AB^c) + P(AB^c) = P(A) + P(B) - 2P(AB)$.

Exercise 1.4.20

$AB \subset A \subset A \cup B$ implies $P(AB) \leq P(A) \leq P(A \cup B) = P(A) + P(B) - P(AB) \leq P(A) + P(B)$. The general case follows similarly, with the last inequality determined by subadditivity.

Exercise 1.4.21

Clearly $P(E) \geq 0$. $P(\Omega) = c_1P_1(\Omega) + c_2P_2(\Omega) = 1$.

$$E = \bigvee_{i=1}^{\infty} E_i \text{ implies } P(E) = c_1 \sum_{i=1}^{\infty} P_1(E_i) + c_2 \sum_{i=1}^{\infty} P_2(E_i) = \sum_{i=1}^{\infty} P(E_i)$$

The pattern is the same for the general case, except that the sum of two terms is replaced by the sum of n terms $c_iP_i(E)$.

Exercise 1.4.22

Clearly $Q(E) \geq 0$ and since $A_i\Omega = A_i$ we have $Q(\Omega) = 1$. If

$$E = \bigvee_{k=1}^{\infty} E_k, \text{ then } P(EA_i) = \sum_{k=1}^{\infty} P(E_kA_i)\ \forall i$$

Interchanging the order of summation shows that Q is countably additive.

This page titled 1.4: Problems on Probability Systems is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

2: MINTERM ANALYSIS

A *fundamental problem* in elementary probability is to find the probability of a logical (Boolean) combination of a finite class of events, when the probabilities of certain other combinations are known. If we partition an event F into component events whose probabilities can be determined, then the additivity property implies the probability of F is the sum of these component probabilities. Frequently, the event F is a Boolean combination of members of a finite class– say, $\{A, B, C\}$ or $\{A, B, C, D\}$. For each such finite class, there is a *fundamental partition* determined by the class. The members of this partition are called *minterms*. Any Boolean combination of members of the class can be expressed as the disjoint union of a unique subclass of the minterms. If the probability of every minterm in this subclass can be determined, then by additivity the probability of the Boolean combination is determined. We examine these ideas in more detail.

2.1: Minterms

2.2: Minterms and MATLAB Calculations

2.3: Problems on Minterm Analysis

This page titled 2: Minterm Analysis is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Minterms

Partitions and minterms

To see how the fundamental partition arises naturally, consider first the partition of the basic space produced by a single event A.

$$\Omega = A \bigvee A^c \tag{2.1.1}$$

Now if B is a second event, then

$$A = AB \bigvee AB^c \tag{2.1.2}$$

and

$$A^c = A^cB \bigvee A^cB^c \tag{2.1.3}$$

so that

$$\Omega = A^cB^c \bigvee A^cB \bigvee AB^c \bigvee AB \tag{2.1.4}$$

The pair $\{A, B\}$ has partitioned Ω into $\{A^cB^c, A^cB, AB^c, AB\}$. Continuation is this way leads systematically to a partition by three events $\{A, B, C\}$, four events $\{A, B, C, D\}$, etc.

We illustrate the fundamental patterns in the case of four events $\{A, B, C, D\}$. We form the *minterms* as intersections of members of the class, with various patterns of complementation. For a class of four events, there are $2^4 = 16$ such patterns, hence 16 minterms. These are, in a systematic arrangement,

$A^cB^cC^cD^c$	$A^cBC^cD^c$	$AB^cC^cD^c$	ABC^cD^c
$A^cB^cC^cD$	A^cBC^cD	AB^cC^cD	ABC^cD
$A^cB^cCD^c$	A^cBCD^c	AB^cCD^c	$ABCD^c$
A^cB^cCD	A^cBCD	AB^cCD	$ABCD$

No element can be in more than one minterm, because each differs from the others by complementation of at least one member event. Each element ω is assigned to exactly one of the minterms by determining the answers to four questions:

Is it in A? Is it in B? Is it in C? Is it in D?

Suppose, for example, the answers are: Yes, No, No, Yes. Then ω is in the minterm AB^cC^cD. In a similar way, we can determine the membership of each ω in the basic space. Thus, the minterms form a *partition*. That is, the minterms represent mutually exclusive events, one of which is sure to occur on each trial. The membership of any minterm depends upon the membership of each generating set A, B, C or D, and the relationships between them. For some classes, one or more of the minterms are empty (impossible events). As we see below, this causes no problems.

An examination of the development above shows that if we begin with a class of *n* events, there are 2^n minterms. To aid in systematic handling, we introduce a simple numbering system for the minterms, which we illustrate by considering again the four events A, B, C, D, in that order. The answers to the four questions above can be represented numerically by the scheme

No ~ 0 and Yes ~ 1

Thus, if ω is in $A^cB^cC^cD^c$, the answers are tabulated as 0 0 0 0. If ω is in AB^cC^cD, then this is designated 1 0 0 1. With this scheme, the minterm arrangement above becomes

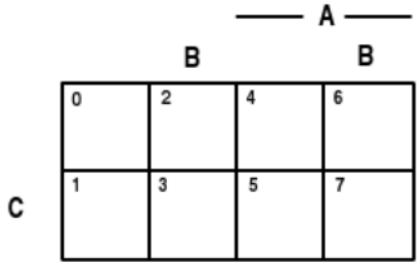
0000 ∼ 0	0100 ∼ 4	1000 ∼ 8	1100 ∼ 12
0001 ∼ 1	0101 ∼ 5	1001 ∼ 9	1101 ∼ 13
0010 ∼ 2	0110 ∼ 6	1010 ∼ 10	1110 ∼ 14
0011 ∼ 3	0111 ∼ 7	1011 ∼ 11	1111 ∼ 15

We may view these quadruples of zeros and ones as binary representations of integers, which may also be represented by their decimal equivalents, as shown in the table. Frequently, it is useful to refer to the minterms by number. If the members of the generating class are treated in a fixed order, then each minterm number arrived at in the manner above specifies a minterm uniquely. Thus, for the generating class $\{A, B, C, D\}$, in that order, we may designate

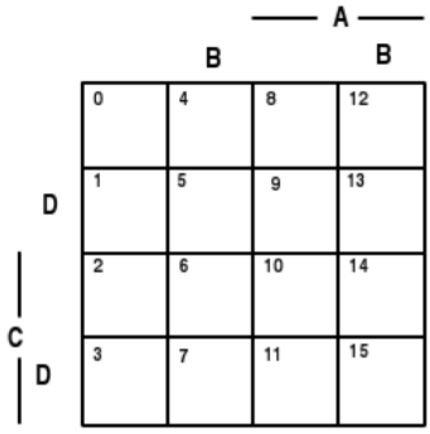
$$A^cB^cC^cD^c = M_0 \text{ (minterm 0) } AB^cC^cD = M_9 \text{ (minterm 9), etc.}$$

We utilize this numbering scheme on special Venn diagrams called *minterm maps*. These are illustrated in Figure, for the cases of three, four, and five generating events. Since the actual content of any minterm depends upon the sets A, B, C, and D in the generating class, it is customary to refer to these sets as *variables*. In the three-variable case, set A is the right half of the diagram and set C is the lower half; but set *B* is split, so that it is the union of the second and fourth columns. Similar splits occur in the other cases.

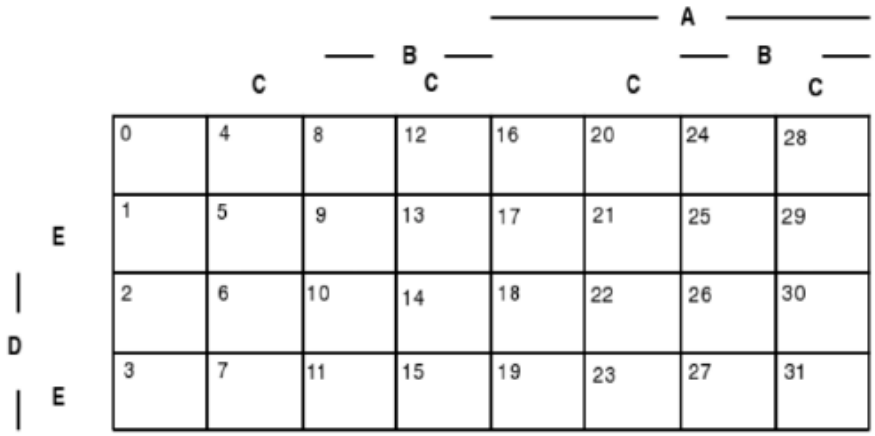
Remark. Other useful arrangements of minterm maps are employed in the analysis of switching circuits.



Three variables



Four variables



Five variables

Figure 2.1.1. Minterm maps for three, four, or five variables

Minterm maps and the minterm expansion

The significance of the minterm partition of the basic space rests in large measure on the following fact.

Minterm expansion

Each Boolean combination of the elements in a generating class may be expressed as the disjoint union of an appropriate subclass of the minterms. This representation is known as the *minterm expansion* for the combination.

In deriving an expression for a given Boolean combination which holds for any class $\{A, B, C, D\}$ of four events, we include all possible minterms, whether empty or not. If a minterm is empty for a given class, its presence does not modify the set content or probability assignment for the Boolean combination.

The existence and uniqueness of the expansion is made plausible by simple examples utilizing minterm maps to determine graphically the minterm content of various Boolean combinations. Using the arrangement and numbering system introduced above, we let M_i represent the ith minterm (numbering from zero) and let $p(i)$ represent the probability of that minterm. When we deal with a union of minterms in a minterm expansion, it is convenient to utilize the shorthand illustrated in the following.

$$M(1,3,7) = M_1 \bigvee M_3 \bigvee M_7 \text{ and } p(1,3,7) = p(1) + p(3) + p(7)$$

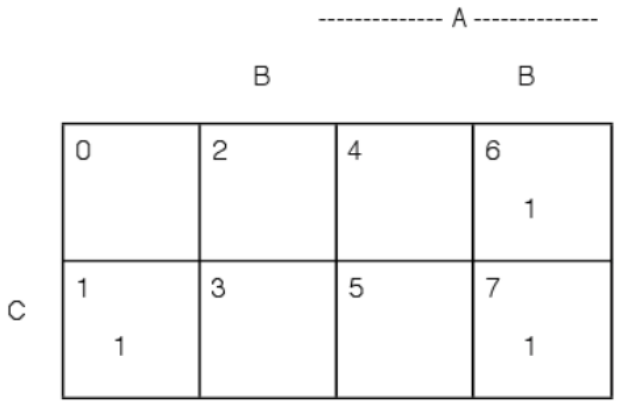


Figure 2.1.2. $E = AB \cup A^c(B \cup C^c)^c = M(1:6,7)$ Minterm expansion for Example 2.1.1

Consider the following simple example.

Example 2.1.1 Minterm expansion

Suppose $E = AB \cup A^c(B \cup C^c)^c$. Examination of the minterm map in Figure 2.1.2 show that AB consists of the union of minterms M_6, M_7, which we designate $M(6,7)$. The combination $B \cup C^c = M(0,2,3,4,6,7)$, so that its complement $(B \cup C^c)^c = M(1,5)$. This leaves the comon part $A^c(B \cup C^c)^c = M_1$, Hence, $E = M(1,6,7)$. Similarly, $F = A \cup B^cC = M(1,4,5,6,7)$.

A key to establishing the expansion is to note that each minterm is either a subset of the combination or is disjoint from it. The expansion is thus the union of those minterms included in the combination. A general verification using indicator functions is sketched in the last section of this module.

Use of minterm maps

A typical problem seeks the probability of certain Boolean combinations of a class of events when the probabilities of various other combinations is given. We consider several simple examples and illustrate the use of minterm maps in formulation and solution.

Example 2.1.2 Survey on software

Statistical data are taken for a certain student population with personal computers. An individual is selected at random. Let $A =$ the event the person selected has word processing, $B =$ the event he or she has a spread sheet program, and $C =$ the event the person has a data base program. The data imply

- The probability is 0.80 that the person has a *word processing* program: $P(A) = 0.8$
- The probability is 0.65 that the person has a *spread sheet* program: $P(B) = 0.65$
- The probability is 0.30 that the person has a *data base* program: $P(C) = 0.3$
- The probability is 0.10 that the person has *all three*: $P(ABC) = 0.1$
- The probability is 0.05 that the person has *neither word processing nor spread sheet*: $P(A^cB^c = 0.05$
- The probability is 0.65 that the person has *at least two*: $P(AB \cup AC \cup BC) = 0.65$
- The probability of *word processor and data base, but no spread sheet* is twice the probabilty of *spread sheet and data base, but no word processor*: $P(AB^cC) = 2P(A^cBC)$

  a. What is the probability that the person has exactly two of the programs?
  b. What is the probability that the person has only the data base program?

Several questions arise:

- Are these data consistent?
- Are the data sufficient to answer the questions?
- How may the data be utilized to anwer the questions?

Solution

The data, expressed in terms of minterm probabilities, are:

$P(A) = p(4,5,6,7) = 0.80$; hence $P(A^c) = p(0,1,2,3) = 0.20$

$P(B) = p(2,3,6,7) = 0.65$; hence $P(B^c) = p(0,1,4,5) = 0.35$

$P(C) = p(1,3,5,7) = 0.30$; hence $P(C^c) = p(0,2,4,6) = 0.70$

$P(ABC) = p(7) = 0.10$ $P(A^cB^c) = p(0,1) = 0.05$

$P(AB \cup AC \cup BC) = p(3,5,6,7) = 0.65$

$P(AB^cC) = p(5) = 2p(3) = 2P(A^cBC)$

These data are shown on the minterm map in Figure 2.1.3 a. We use the patterns displayed in the minterm map to aid in an algebraic solution for the various minterm probabilities.

$p(2,3) = p(0,1,2,3) - p(0,1) = 0.20 - 0.05 = 0.15$

$p(6,7) = p(2,3,6,7) - p(2,3) = 0.65 - 0.15 = 0.50$

$p(6) = p(6,7) - p(7) = 0.50 - 0.10 = 0.40$

$p(3,5) = p(3,5,6,7) - p(6,7) = 0.65 - 0.50 = 0.15 \Rightarrow p(3) = 0.05$,

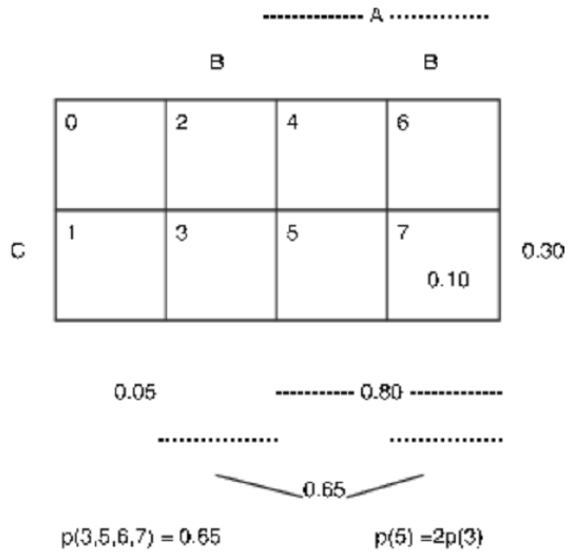
$p(5) = 0.10 \Rightarrow p(2) = 0.10$

$p(1) = p(1,3,5,7) - p(3,5) - p(7) = 0.30 - 0.15 - 0.10 = 0.05 \Rightarrow p(0) = 0$

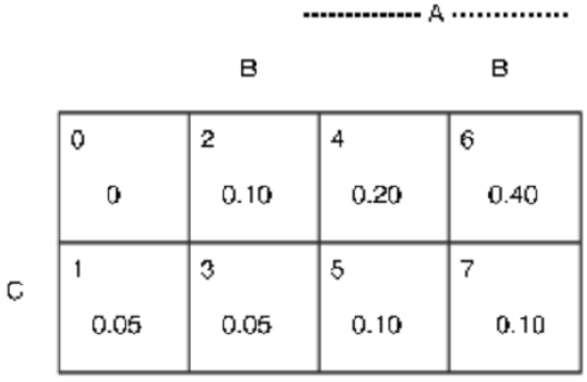
$p(4) = p(4,5,6,7) - p(5) - p(6,7) = 0.80 - 0.10 - 0.50 = 0.20$

Thus, all minterm probabilities are determined. They are displayed in Figure 2.1.3 b. From these we get

$$P(A^cBC \bigvee AB^cC \bigvee ABC^c) = p(3,5,6) = 0.05 + 0.10 + 0.40 = 0.55 \text{ and } P(A^cB^cC) = p(1) = 0.05$$



a. Data for software survey, Example 2.3.1



b. Minterm probabilities for software survey. Example 3.3.1

Figure 2.1.3. Minterm maps for software survey.

Example 2.1.3 Survey on personal computers

A survey of 1000 students shows that 565 have PC compatible desktop computers, 515 have Macintosh desktop computers, and 151 have laptop computers. 51 have all three, 124 have both PC and laptop computers, 212 have at least two of the three, and twice as many own both PC and laptop as those who have both Macintosh desktop and laptop. A person is selected at random from this population. What is the probability he or she has at least one of these types of computer? What is the probability the person selected has only a laptop?

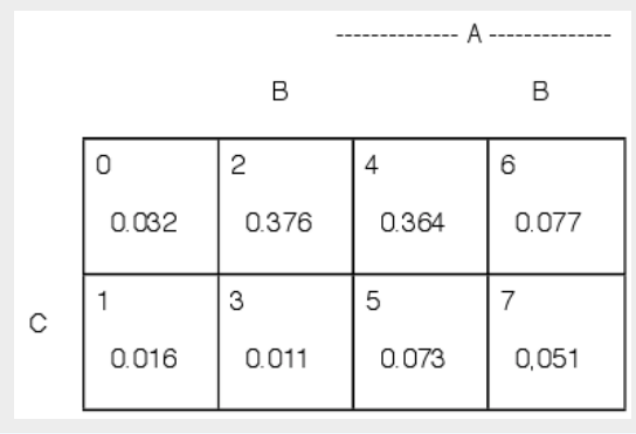


Figure 2.1.4. Minterm probabilities for computer survey. Example 2.1.3

Solution

Let $A =$ the event of owning a PC desktop, $B =$ the event of owning a Macintosh desktop, and $C =$ the event of owning a laptop. We utilize a minterm map for three variables to help determine minterm patterns. For example, the event $AC = M_5 \bigvee M_7$ so that $P(AC) = p(5) + p(7) = p(5, 7)$.

The data, expressed in terms of minterm probabilities, are:

$P(A) = p(4, 5, 6, 7) = 0.565$, hence $P(A^c) = p(0, 1, 2, 3) = 0.435$

$P(B) = p(2, 3, 6, 7) = 0.515$, hence $P(B^c) = p(0, 1, 4, 5) = 0.485$

$P(C) = p(1, 3, 5, 7) = 0.151$, hence $P(C^c) = p(0, 2, 4, 6) = 0.849$

$P(ABC) = p(7) = 0.051$ $P(AC) = p(5, 7) = 0.124$

$P(AB \cup AC \cup BC) = p(3, 5, 6, 7) = 0.212$

$P(AC) = p(5, 7) = 2p(3, 7) = 2P(BC)$

We use the patterns displayed in the minterm map to aid in an algebraic solution for the various minterm probabilities.

$p(5) = p(5, 7) - p(7) = 0.124 - 0.051 = 0.073$

$p(1, 3) = P(A^cC) = 0.151 - 0.124 = 0.027$ $P(AC^c) = p(4, 6) = 0.565 - 0.124 = 0.441$

$p(3, 7) = P(BC) = 0.124/2 = 0.062$

$p(3) = 0.062 - 0.051 = 0.011$

$p(6) = p(3, 4, 6, 7) - p(3) - p(5, 7) = 0.212 - 0.011 - 0.124 = 0.077$

$p(4) = P(A) - p(6) - p(5, 7) = 0.565 - 0.077 - 0.1124 = 0.364$

$p(1) = p(1,3) - p(3) = 0.027 - 0.11 = 0.016$

$p(2) = P(B) - p(3,7) - p(6) = 0.515 - 0.062 - 0.077 = 0.376$

$p(0) = P(C^c) - p(4,6) - p(2) = 0.8490.441 - 0.376 = 0.032$

We have determined the minterm probabilities, which are displayed on the minterm map Figure 2.1.4. We may now compute the probability of any Boolean combination of the generating events A, B, C. Thus,

$$P(A \cup B \cup C) = 1 - P(A^c B^c C^c) - 1 - p(0) = 0.968 \text{ and } P(A^c B^c C) = p(1) = 0.016$$

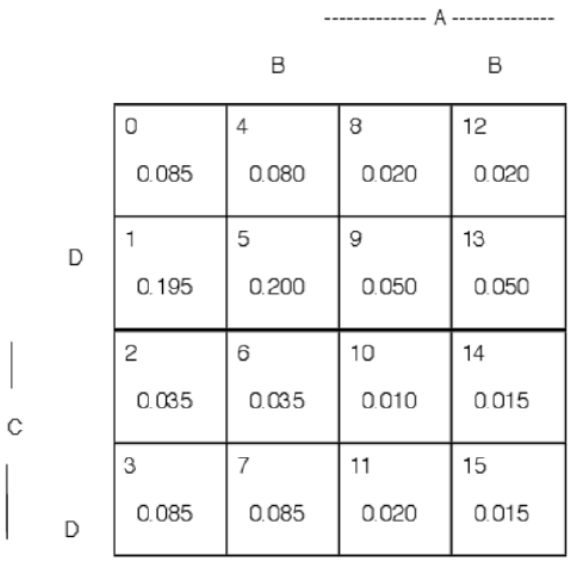


Figure 2.1.5. Minterm probabilities for opinion survey. Example 2.1.4

Example 2.1.4 Opinion survey

A survey of 1000 persons is made to determine their opinions on four propositions. Let A, B, C, D be the events a person selected agrees with the respective propositions. Survey results show the following probabilities for various combinations:

$$P(A) = 0.200,\ P(B) = 0.500,\ P(C) = 0.300,\ P(D) = 0.700,\ P(A(B \cup C^c)D^c) = 0.055$$

$$P(A \cup BC \cup D^c) = 0.520,\ P(A^c BC^c D) = 0.120,\ P(ABCD) = 0.015,\ P(AB^c C) = 0.030$$

$$P(A^c B^c C^c D) = 0.195,\ P(A^c BC) = 0.120,\ P(A^c B^c D^c) = 0.120,\ P(AC^c) = 0.140$$

$$P(ACD^c) = 0.025,\ P(ABC^c D^c) = 0.020$$

Determine the probabilities for each minterm and for each of the following combinations

$A^c(BC^c \cup B^c C)$ - that is, not A and (B or C, but not both)

$A \cup BC^c$ - that is, A or (B and not C)

Solution

At the outset, it is not clear that the data are consistent or sufficient to determine the minterm probabilities. However, an examination of the data shows that there are sixteen items (including the fact that the sum of all minterm probabilities is one). Thus, there is hope, but no assurance, that a solution exists. A step elimination procedure, as in the previous examples,

shows that all minterms can in fact be calculated. The results are displayed on the minterm map in Figure 2.1.5. It would be desirable to be able to analyze the problem systematically. The formulation above suggests a more systematic algebraic formulation which should make possible machine aided solution.

Systematic formulation

Use of a minterm map has the advantage of visualizing the minterm expansion in direct relation to the Boolean combination. The algebraic solutions of the previous problems involved *ad hoc* manipulations of the data minterm probability combinations to find the probability of the desired target combination. We seek a systematic formulation of the data as a set of linear algebraic equations with the minterm probabilities as unknowns, so that standard methods of solution may be employed. Consider again the software survey of Example 2.1.1.

Example 2.1.5 The softerware survey problem reformulated

The data, expressed in terms of minterm probabilities, are:

$P(A) = p(4,5,6,7) = 0.80$

$P(B) = p(2,3,6,7) = 0.65$

$P(C) = p(1,3,5,7) = 0.30$

$P(ABC) = p(7) = 0.10$

$P(A^cB^c) = p(0,1) = 0.05$

$P(AB \cup AC \cup BC) = p(3,5,6,7) = 0.65$

$P(AB^cC) = p(5) = 2p(3) = 2P(A^cBC)$, so that $p(5) - 2p(3) = 0$

We also have in any case

$P(\Omega) = P(A \cup A^c) = p(0,1,2,3,4,5,6,7) = 1$

to complete the eight items of data needed for determining all eight minterm probabilities. The first datum can be expressed as an equation in minterm probabilities:

$$0 \cdot p(0) + 0 \cdot p(1) + 0 \cdot p(2) + 0 \cdot p(3) + 1 \cdot p(4) + 1 \cdot p(5) + 1 \cdot p(6) + 1 \cdot p(7) = 0.80$$

This is an algebraic equation in $p(0), \cdots, p(7)$ with a matrix of coefficients

$$[0\ 0\ 0\ 0\ 1\ 1\ 1\ 1]$$

The others may be written out accordingly, giving eight linear algebraic equations in eight variables $p(0)$ through $p(7)$. Each equation has a matrix or vector of zero-one coefficients indicating which minterms are included. These may be written in matrix form as follows:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & -2 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} p(0) \\ p(1) \\ p(2) \\ p(3) \\ p(4) \\ p(5) \\ p(6) \\ p(7) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.80 \\ 0.65 \\ 0.30 \\ 0.10 \\ 0.05 \\ 0.65 \\ 0 \end{bmatrix} = \begin{bmatrix} P(\Omega) \\ P(A) \\ P(B) \\ P(C) \\ P(ABC) \\ P(A^cB^c) \\ P(AB \cup AC \cup BC) \\ P(AB^cC) - 2P(A^cBC) \end{bmatrix}$$

- The patterns in the coefficient matrix are determined by *logical operations*. We obtained these with the aid of a minterm map.
- The solution utilizes an *algebraic procedure*, which could be carried out in a variety of ways, including several standard computer packages for matrix operations.

We show in the module Minterm Vectors and MATLAB how we may use MATLAB for *both* aspects.

Indicator functions and the minterm expansion

Previous discussion of the indicator function shows that the indicator function for a Boolean combination of sets is a numerical valued function of the indicator functions for the individual sets.

- As an indicator function, it takes on only the values zero and one.
- The value of the indicator function for any Boolean combination must be constant on each minterm. For example, for each ω in the minterm AB^cCD^c, we must have $I_A(\omega) = 1$, $I_B(\omega) = 0$, $I_C(\omega) = 1$, and $I_D(\omega) = 0$. Thus, any function of I_A, I_B, I_C, I_D must be constant over the minterm.
- Consider a Boolean combination E of the generating sets. If ω is in $E \cap M_i$, then $I_E(\omega) = 1$ for all $\omega \in M_i$, so that $M_i \subset E$. Since each $\omega \in M_i$ or some i, E must be the union of those minterms sharing an ω with E.
- Let $\{M_i : i \in J_E\}$ be the subclass of those minterms on which I_E has the value one. Then

$$E = \bigvee_{J_E} M_i$$

which is the minterm expansion of E.

This page titled 2.1: Minterms is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Minterms and MATLAB Calculations

The concepts and procedures in this unit play a significant role in many aspects of the analysis of probability topics and in the use of MATLAB throughout this work.

Minterm vectors and MATLAB

The systematic formulation in the previous module Minterms shows that each Boolean combination, as a union of minterms, can be designated by a vector of zero-one coefficients. A coefficient one in the ith position (numbering from zero) indicates the inclusion of minterm M_i in the union. We formulate this pattern carefully below and show how MATLAB logical operations may be utilized in problem setup and solution.

Suppose E is a Boolean combination of A, B, C. Then, by the minterm expansion,

$$E = \bigvee_{J_E} M_i$$

where M_i is the ith minterm and J_E is the set of indices for those M_i included in E. For example, consider

$$E = A(B \cup C^c) \cup A^c(B \cup C^c)^c = M_1 \bigvee M_4 \bigvee M_6 \bigvee M_7 = M(1, 4, 6, 7)$$

$$F = A^cB^c \cup AC = M_0 \bigvee M_1 \bigvee M_5 \bigvee M_7 = M(0, 1, 5, 7)$$

We may designate each set by a pattern of zeros and ones $(e_0, e_1, \cdots, e_7)$. The ones indicate which minterms are present in the set. In the pattern for set E, minterm M_i is included in E iff $e_i = 1$. This is, in effect, another arrangement of the minterm map. In this form, it is convenient to view the pattern as a *minterm vector*, which may be represented by a row matrix or row vector $[e_0, e_1, \cdots, e_7]$. We find it convenient to use *the same symbol for the name of the event and for the minterm vector or matrix representing it*. Thus, for the examples above,

$$E \sim [0\ 1\ 0\ 0\ 1\ 0\ 1\ 1] \text{ and } F \sim [1\ 1\ 0\ 0\ 0\ 1\ 0\ 1]$$

It should be apparent that this formalization can be extended to sets generated by any finite class.

Minterm vectors for Boolean combinations

If E and F are combinations of n generating sets, then each is represented by a unique minterm vector of length 2^n. In the treatment in the module Minterms, we determine the minterm vector with the aid of a minterm map. We wish to develop a systematic way to determine these vectors.

As a first step, we suppose we have minterm vectors for E and F and want to obtain the minterm vector of Boolean combinations of these.

1. The minterm expansion for $E \cup F$ has all the minterms in either set. This means the jth element of the vector for $E \cup F$ is the *maximum* of the jth elements for the two vectors.
2. The minterm expansion for $E \cap F$ has only those minterms in both sets. This means the jth element of the vector for $E \cap F$ is the *minimum* of the jth elements for the two vectors.
3. The minterm expansion for E^c has only those minterms not in the expansion for E. This means the vector for E^c has zeros and ones interchanged. The jth element of E^c is one iff the corresponding element of E is zero.

We illustrate for the case of the two combinations E and F of three generating sets, considered above

$$E = A(B \cup C^c) \cup A^c(B \cup C^c)^c \cup A^c(B \cup C^c)^c \sim [0\ 1\ 0\ 0\ 1\ 0\ 1\ 1] \text{ and } F = A^cB^c \cup AC \sim [1\ 1\ 0\ 0\ 0\ 1\ 0\ 1]$$

Then

$$E \cup F \sim [1\ 1\ 0\ 0\ 1\ 1\ 1\ 1],\ E \cap F \sim [0\ 1\ 0\ 0\ 0\ 0\ 0\ 1], \text{ and } E^c \sim [1\ 0\ 1\ 1\ 0\ 1\ 0\ 0]$$

MATLAB logical operations

MATLAB logical operations on zero-one matrices provide a convenient way of handling Boolean combinations of minterm vectors represented as matrices. For two *zero-one* matrices E, F of the same size

- $E|F$ is the matrix obtained by taking the maximum element in each place.
- Misplaced '&' is the matrix obtained by taking the minimum element in each place.
- E^c is the matrix obtained by interchanging one and zero in each place in E.

Thus, if E, F are minterm vectors for sets by the same name, then $E|F$ is the minterm vector for $E \cup F$, Misplaced '&' is the minterm vector for $E \cap F$, and $E = 1 - E$ is the minterm vector for E^c.

This suggests a general approach to determining minterm vectors for Boolean combinations.

> Start with minterm vectors for the generating sets.
> Use MATLAB logical operations to obtain the minterm vector for any Boolean combination.

Suppose, for example, the class of generating sets is $\{A, B, C\}$. Then the minterm vectors for A, B, and C, respectively, are

$$A = [0\ 0\ 0\ 0\ 1\ 1\ 1\ 1]\ B = [0\ 0\ 1\ 1\ 0\ 0\ 1\ 1]\ C = [0\ 1\ 0\ 1\ 0\ 1\ 0\ 1]$$

If $E = AB \cup C^c$, then the logical combination Misplaced '&' of the matrices yields $E = [1\ 0\ 1\ 0\ 1\ 0\ 1\ 1]$.

MATLAB implementation

A key step in the procedure just outlined is to obtain the minterm vectors for the generating elements $\{A, B, C\}$. We have an m-function to provide such fundamental vectors. For example to produce the second minterm vector for the family (i.e., the minterm vector for B), the basic zero-one pattern `0 0 1 1` is replicated twice to give

```
01  0     0     1     1     0     0     1     1
```

The function *minterm(n,k)* generates the *k*th minterm vector for a class of *n* generating sets.

minterms for the class {a, b, c}.

```
01  >> A = minterm(3,1)
02  A =  0     0     0     0     1     1     1     1
03  >> B = minterm(3,2)
04  B =  0     0     1     1     0     0     1     1
05  >> C = minterm(3,3)
06  C =  0     1     0     1     0     1     0     1
```

minterm patterns for the boolean combinations

$F = AB \cup B^cC\ G = A \cup A^cC$

```
01  F = (A&B)|(~B&C)
02  F =  0     1     0     0     0     1     1     1
03  >> G = A|(~A&C)
04  G =  0     1     0     1     1     1     1     1
05  >> JF = find(F)-1          % Use of find to determine index set for F
06  JF =   1     5     6     7  % Shows F = M(1, 5, 6, 7)
```

These basic minterm patterns are useful not only for Boolean combinations of events but also for many aspects of the analysis of

those random variables which take on only a finite number of values.

Zero-one arrays in MATLAB

The treatment above hides the fact that a rectangular array of zeros and ones can have two quite different meanings and functions in MATLAB.

> A *numerical matrix* (or vector) subject to the usual operations on matrices..
> A *logical array* whose elements are combined by a. Logical operators to give new logical arrays; b. Array operations (element by element) to give numerical matrices; c. Array operations with numerical matrices to give numerical results.

Some simple examples will illustrate the principal properties.

```
>>> A = minterm(3,1);
>> B = minterm(3,2);
>> C = minterm(3,3);
>> F = (A&B)|(~B&C)
F =  0     1     0     0     0     1     1     1
>> G = A|(~A&C)
G =  0     1     0     1     1     1     1     1
>> islogical(A)       % Test for logical array
ans =    0
>> islogical(F)
ans =   1
>> m = max(A,B)       % A matrix operation
m =   0     0     1     1     1     1     1     1
>> islogical(m)
ans =   0
>> m1 = A|B           % A logical operation
m1 =   0     0     1     1     1     1     1     1
>> islogical(m1)
ans = 1
>> a = logical(A)      % Converts 0-1 matrix into logical array
a =   0     0     0     0     1     1     1     1
>> b = logical(B)
>> m2 = a|b
m2 =   0     0     1     1     1     1     1     1
>> p = dot(A,B)        % Equivalently, p = A*B'
p =  2
>> p1 = total(A.*b)
p1 =  2
>> p3 = total(A.*B)
p3 =  2
>> p4 = a*b'           % Cannot use matrix operations on logical arrays
??? Error using ==> mtimes   % MATLAB error signal
Logical inputs must be scalar.
```

Often it is desirable to have a table of the generating minterm vectors. Use of the function minterm in a simple “for loop” yields the following m-function.

The function *mintable(n)* Generates a table of minterm vectors for n generating sets.

mintable for three variables

```
>> M3 = mintable(3)
M3 = 0     0     0     0     1     1     1     1
```

```
03     0     0     1     1     0     0     1     1
04     0     1     0     1     0     1     0     1
```

As an application of mintable, consider the problem of determining the probability of k of n events. If $\{A_i : 1 \le i \le n\}$ is any finite class of events, the event that exactly k of these events occur on a trial can be characterized simply in terms of the minterm expansion. The event A_{kn} that exactly k occur is given by

$$A_{kn} = \textit{the disjoint union of those minterms with exactly } k \textit{ positions uncomplemented}$$

In the matrix `M = mintable(n)` these are the minterms corresponding to columns with exactly k ones. The event BknBkn that k or more occur is given by

$$B_{kn} = \bigvee_{r=k}^{n} A_{rn}$$

If we have the minterm probabilities, it is easy to pick out the appropriate minterms and combine the probabilities. The following example in the case of three variables illustrates the procedure.

the software survey (continued)

In the software survey problem, the minterm probabilities are

$$pm = [0\ 0.05\ 0.10\ 0.05\ 0.20\ 0.10\ 0.40\ 0.10]$$

where $A =$ event has word processor, $B =$ event has spread sheet, $C =$ event has a data base program. It is desired to get the probability an individual selected has k of these, $k = 0, 1, 2, 3$.

Solution

We form a mintable for three variables. We count the number of "successes" corresponding to each minterm by using the MATLAB function sum, which gives the sum of each column. In this case, it would be easy to determine each distinct value and add the probabilities on the minterms which yield this value. For more complicated cases, we have an m-function called *csort* (for sort and consolidate) to perform this operation.

```
01  >> pm = 0.01*[0 5 10 5 20 10 40 10];
02  >> M = mintable(3)
03  M =
04  0     0     0     0     1     1     1     1
05  0     0     1     1     0     0     1     1
06  0     1     0     1     0     1     0     1
07  >> T = sum(M)                                  % Column sums give number
08  T =  0     1     1     2     1     2     2     3 % of successes on each
09  >> [k,pk] = csort(T,pm);                       % minterm, determines
10                                                 % distinct values in T and
11  >> disp([k;pk]')                               % consolidates probabilities
12  0         0
13  1.0000    0.3500
14  2.0000    0.5500
15  3.0000    0.1000
```

For three variables, it is easy enough to identify the various combinations "by eye" and make the combinations. For a larger number of variables, however, this may become tedious. The approach is much more useful in the case of Independent Events, because of the ease of determining the minterm probabilities.

Minvec procedures

Use of the tilde $\sim$ to indicate the complement of an event is often awkward. It is customary to indicate the complement of an event E by E^c. In MATLAB, we cannot indicate the superscript, so we indicate the complement by E^c instead of $\sim E$. To facilitate writing combinations, we have a family of *minvec* procedures (minvec3, minvec4, ..., minvec10) to expedite expressing Boolean combinations of $n = 3, 4, 5, \cdots, 10$ sets. These generate and name the minterm vector for each generating set and its complement.

boolean combinations using minvec3

We wish to generate a matrix whose rows are the minterm vectors for $\Omega = A \cup A^c, A, AB, ABC, C$, and A^cC^c, respectively.

```
>> minvec3                                % Call for the setup procedure
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired
>> V = [A|Ac; A; A&B; A&B&C; C; Ac&Cc];  % Logical combinations (one per
                                        % row) yield logical vectors
>> disp(V)
1     1     1     1     1     1     1     1   % Mixed logical and
0     0     0     0     1     1     1     1   % numerical vectors
0     0     0     0     0     0     1     1
0     0     0     0     0     0     0     1
0     1     0     1     0     1     0     1
1     0     1     0     0     0     0     0
```

Minterm probabilities and Boolean combination

If we have the probability of every minterm generated by a finite class, we can determine the probability of any Boolean combination of the members of the class. When we know the minterm expansion or, equivalently, the minterm vector, we simply pick out the probabilities corresponding to the minterms in the expansion and add them. In the following example, we do this “by hand” then show how to do it with MATLAB .

Consider $E = A(B \cup C^c) \cup A^c(B \cup C^c)^c$ and $F = A^cB^c \cup AC$ of the example above, and suppose the respective minterm probabilites are

$p_0 = 0.21$, $p_1 = 0.06$, $p_2 = 0.29$, $p_3 = 0.11$, $p_4 = 0.09$, $p_5 = 0.03$, $p_6 = 0.14$, $p_7 = 0.07$

Use of a minterm map shows $E = M(1, 4, 6, 7)$ and $F = M(0, 1, 5, 7)$. so that

$P(E) = p_1 + p_4 + p_6 + p_7 = p(1, 4, 6, 7) = 0.36$ and $P(F) = p(0, 1, 5, 7) = 0.37$

This is easily handled in MATLAB.

- Use minvec3 to set the generating minterm vectors.
- Use *logical* matrix operations

  Misplaced '&' and Misplaced '&'

  to obtain the (logical) minterm vectors for E and F
- If pm is the matrix of minterm probabilities, perform the *algebraic* dot product or scalar product of the pmpm matrix and the minterm vector for the combination. This can be called for by the MATLAB commands PE = E*pm' and PF = F*pm'.

The following is a transcript of the MATLAB operations.

```
>> minvec3                 % Call for the setup procedure
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
>> E  = (A&(B|Cc))|(Ac&~(B|Cc));
```

```
>> F  = (Ac&Bc)|(A&C);
>> pm = 0.01*[21 6 29 11 9 3 14 7];
>> PE = E*pm'              % Picks out and adds the minterm probabilities
PE =  0.3600
>> PF = F*pm'
PF =  0.3700
```

solution of the software survey problem

We set up the matrix equations with the use of MATLAB and solve for the minterm probabilities. From these, we may solve for the desired “target” probabilities.

```
>> minvec3
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Data vector combinations are:
>> DV = [A|Ac; A; B; C; A&B&C; Ac&Bc; (A&B)|(A&C)|(B&C); (A&Bc&C) - 2*(Ac&B&C)]
DV =
1     1     1     1     1     1     1     1    % Data mixed numerical
0     0     0     0     1     1     1     1    % and logical vectors
0     0     1     1     0     0     1     1
0     1     0     1     0     1     0     1
0     0     0     0     0     0     0     1
1     1     0     0     0     0     0     0
0     0     0     1     0     1     1     1
0     0     0    -2     0     1     0     0
>> DP = [1 0.8 0.65 0.3 0.1 0.05 0.65 0];  % Corresponding data probabilities
>> pm = DV\DP'                             % Solution for minterm probabilities
pm =
-0.0000                                      % Roundoff -3.5 x 10-17
0.0500
0.1000
0.0500
0.2000
0.1000
0.4000
0.1000
>> TV = [(A&B&Cc)|(A&Bc&C)|(Ac&B&C); Ac&Bc&C]       % Target combinations
TV =
0     0     0     1     0     1     1     0    % Target vectors
0     1     0     0     0     0     0     0
>> PV = TV*pm                           % Solution for target probabilities
PV =
0.5500                              % Target probabilities
0.0500
```

An alternate approach

The previous procedure first obtained all minterm probabilities, then used these to determine probabilities for the target combinations. The following procedure does not require calculation of the minterm probabilities. Sometimes the data are not sufficient to calculate all minterm probabilities, yet are sufficient to allow determination of the target probabilities.

Suppose the data minterm vectors are linearly independent, and the target minterm vectors are linearly dependent upon the data vectors (i.e., the target vectors can be expressed as linear combinations of the data vectors). Now each target probability is the same linear combination of the data probabilities. To determine the linear combinations, solve the matrix equation

$TV = CT * DV$ *which has the* $MATLAB$ *solution* $CT = TV/DV$

Then the matrix tp of target probabilites is given by $tp = CT * DP'$. Continuing the MATLAB procedures above, we have:

```
01 >> CT = TV/DV;
```

```
01 >> tp = CT*DP'
```

```
01 tp = 0.5500
02          0.0500
```

The procedure mincalc

The procedure *mincalc* performs calculations as in the preceding examples. The *refinements* consist of determining consistency and computability of various individual minterm probabilities and target probilities. The consistency check is principally for negative minterm probabilities. The computability tests are tests for linear independence by means of calculation of ranks of various matrices. The procedure picks out the computable minterm probabilities and the computable target probabilities and calculates them.

To utilize the procedure, the problem must be formulated appropriately and precisely, as follows:

> Use the MATLAB program *minvecq* to set minterm vectors for each of *q* basic events.
> *Data* consist of Boolean combinations of the basic events and the respective probabilities of these combinations. These are organized into two matrices:
>
> - The *data vector matrix* DV has the data Boolean combinations– one on each row. MATLAB translates each row into the minterm vector for the corresponding Boolean combination. The first entry (on the first row) is A | Ac (for $A \bigvee A^c$), which is the whole space. Its minterm vector consists of a row of ones.
> - The *data probability matrix* DP is a row matrix of the data probabilities. The first entry is one, the probability of the whole space.
>
> The *objective* is to determine the probability of various *target* Boolean combinations. These are put into the *target vector matrix* TV, one on each row. MATLAB produces the minterm vector for each corresponding target Boolean combination.

Computational note. In mincalc, it is necessary to turn the arrays DV and TV consisting of zero-one patterns into zero-one matrices. This is accomplished for DV by the operation `DV = ones(size(DV)).*DV`. and similarly for TV. Both the original and the transformed matrices have the same zero-one pattern, but MATLAB interprets them differently.

Usual case

Suppose the data minterm vectors are linearly independent and the target vectors are each linearly dependent on the data minterm vectors. Then each target minterm vector is expressible as a linear combination of data minterm vectors. Thus, there is a matrix CT such that $TV = CT * DV$. MATLAB solves this with the command $CT = TV/DV$ *The target probabilities are the same linear combinations of the data probabilities*. These are obtained by the MATLAB operation $tp = DP * CT'$.

Cautionary notes

The program mincalc depends upon the provision in MATLAB for solving equations when less than full data are available (based on the singular value decomposition). There are several situations which should be dealt with as special cases. It is usually a good idea to check results by hand to determine whether they are consistent with data. The checking by hand is usually much easier than obtaining the solution unaided, so that use of MATLAB is advantageous even in questionable cases.

The Zero Problem. If the total probability of a group of minterms is zero, then it follows that the probability of each minterm in the group is zero. However, if mincalc does not have enough information to calculate the separate minterm probabilities in the case they are not zero, it will not pick up in the zero case the fact that the separate minterm probabilities are zero. It simply considers these minterm probabilities not computable.

Linear dependence. In the case of linear dependence, the operation called for by the command CT = TV/DV may not be able to solve the equations. The matrix may be singular, or it may not be able to decide which of the redundant data equations to use. Should it provide a solution, the result should be checked with the aid of a minterm map.

Consistency check. Since the consistency check is for negative minterms, if there are not enough data to calculate the minterm probabilities, there is no simple check on the consistency. Sometimes the probability of a target vector included in another vector will actually exceed what should be the larger probability. Without considerable checking, it may be difficult to determine consistency.

In a few unusual cases, the command CT = TV/DV does not operate appropriately, even though the data should be adequate for the problem at hand. Apparently the approximation process does not converge.

MATLAB Solutions for examples using mincalc

software survey

```
% file mcalc01   Data for software survey
minvec3;
DV = [A|Ac; A; B; C; A&B&C; Ac&Bc; (A&B)|(A&C)|(B&C); (A&Bc&C)  - 2*(Ac&B&C)];
DP = [1 0.8 0.65 0.3 0.1 0.05 0.65 0];
TV = [(A&B&Cc)|(A&Bc&C)|(Ac&B&C); Ac&Bc&C];
disp('Call for mincalc')
>> mcalc01         % Call for data
Call for mincalc   % Prompt supplied in the data file
>> mincalc
Data vectors are linearly independent
Computable target probabilities
1.0000    0.5500
2.0000    0.0500
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
>> disp(PMA)       % Optional call for minterm probabilities
0         0
1.0000    0.0500
2.0000    0.1000
3.0000    0.0500
4.0000    0.2000
5.0000    0.1000
6.0000    0.4000
7.0000    0.1000
```

computer survey

```
% file mcalc02.m    Data for computer survey
minvec3
DV =   [A|Ac; A; B; C; A&B&C; A&C; (A&B)|(A&C)|(B&C); ...
2*(B&C) - (A&C)];
DP = 0.001*[1000 565 515 151 51 124 212 0];   TV = [A|B|C; Ac&Bc&C];
disp('Call for mincalc')
>> mcalc02
Call for mincalc
```

```
>> mincalc
Data vectors are linearly independent
Computable target probabilities
1.0000    0.9680
2.0000    0.0160
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
>> disp(PMA)
0         0.0320
1.0000    0.0160
2.0000    0.3760
3.0000    0.0110
4.0000    0.3640
5.0000    0.0730
6.0000    0.0770
7.0000    0.0510
```

```
% file mcalc03.m    Data for opinion survey
minvec4
DV = [A|Ac; A; B; C; D; A&(B|Cc)&Dc; A|((B&C)|Dc) ; Ac&B&Cc&D; ...
A&B&C&D; A&Bc&C; Ac&Bc&Cc&D; Ac&B&C; Ac&Bc&Dc; A&Cc; A&C&Dc; A&B&Cc&Dc];
DP =  0.001*[1000 200 500 300 700 55 520 200 15 30 195 120 120 ...
              140 25 20];
TV = [Ac&((B&Cc)|(Bc&C)); A|(B&Cc)];
disp('Call for mincalc')
>> mincalc03
Call for mincalc
>> mincalc
Data vectors are linearly independent
Computable target probabilities
1.0000    0.4000
2.0000    0.4800
The number of minterms is 16
The number of available minterms is 16
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
>> disp(minmap(pma))    % Display arranged as on minterm map
0.0850    0.0800    0.0200    0.0200
0.1950    0.2000    0.0500    0.0500
0.0350    0.0350    0.0100    0.0150
0.0850    0.0850    0.0200    0.0150
```

The procedure mincalct

A useful modification, which we call *mincalct*, computes the available target probabilities, without checking and computing the minterm probabilities. This procedure assumes a data file similar to that for mincalc, except that it does not need the target matrix TV, since it prompts for target Boolean combination inputs. The procedure mincalct may be used after mincalc has performed its operations to calculate probabilities for additional target combinations.

(continued) Additional target datum for the opinion survey

Suppose mincalc has been applied to the data for the opinion survey and that it is desired to determine $P(AD \cup BD^c)$. It is not necessary to recalculate all the other quantities. We may simply use the procedure mincalct and input the desired Boolean combination at the prompt.

```
>> mincalct
Enter matrix of target Boolean combinations  (A&D)|(B&Dc)
Computable target probabilities
    1.0000    0.2850
```

Repeated calls for mcalct may be used to compute other target probabilities.

This page titled 2.2: Minterms and MATLAB Calculations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Problems on Minterm Analysis

Exercise 2.3.1

Consider the class $\{A, B, C, D\}$ of events. Suppose the probability that at least one of the events A or C occurs is 0.75 and the probability that at least one of the four events occurs is 0.90. Determine the probability that neither of the events A or C but at least one of the events B or D occurs.

Answer available at the end of this section.

Exercise 2.3.2

1. Use minterm maps to show which of the following statements are true for any class $\{A, B, C\}$:

   a. $A \cup (BC)^c = A \cup B \cup B^c C^c$
   b. $(A \cup B)^c = A^c C \cup B^c C$
   c. $A \subset AB \cup AC \cup BC$
2. Repeat part (1) using indicator functions (evaluated on minterms).
3. Repeat part (1) using the m-procedure minvec3 and MATLAB logical operations.

Answer available at the end of this section.

Exercise 2.3.3

Use (1) minterm maps, (2) indicator functions (evaluated on minterms), (3) the m-procedure minvec3 and MATLAB logical operations to show that

a. $A(B \cup C^c) \cup A^c BC \subset A(BC \cup C^c) \cup A^c B$
b. $A \cup A^c BC = AB \cup BC \cup AC \cup AB^c C^c$

Answer available at the end of this section.

Exercise 2.3.4

Minterms for the events $\{A, B, C, D\}$, arranged as on a minterm map are

```
01                0.0168    0.0072    0.0252    0.0108
02                0.0392    0.0168    0.0588    0.0252
03                0.0672    0.0288    0.1008    0.0432
04                0.1568    0.0672    0.2352    0.1008
```

What is the probability that three or more of the events occur on a trial? Of exactly two? Of two or fewer?

Answer available at the end of this section.

Exercise 2.3.5

Minterms for the events $\{A, B, C, D, E\}$, arranged as on a minterm map are

```
        0.0216  0.0324  0.0216  0.0324  0.0144  0.0216  0.0144  0.0216
        0.0144  0.0216  0.0144  0.0216  0.0096  0.0144  0.0096  0.0144
        0.0504  0.0756  0.0504  0.0756  0.0336  0.0504  0.0336  0.0504
        0.0336  0.0504  0.0336  0.0504  0.0224  0.0336  0.0224  0.0336
```

What is the probability that three or more of the events occur on a trial? Of exactly four? Of three or fewer? Of either two or four?

Answer available at the end of this section.

Exercise 2.3.6

Suppose $P(A \cup B^cC) = 0.65$, $P(AC) = 0.2$, $P(A^cB) = 0.25$

$P(A^cC^c) = 0.25$, $P(BC) = 0.30$. Determine $P((AC^c \cup A^cC)B^c)$.

Then determine $P((AB^c \cup A^c)C^c)$ and $P(A^c(B \cup C^c))$, if possible.

Answer available at the end of this section.

Exercise 2.3.7

Suppose $P((AB^c \cup A^cB)C) = 0.4$, $P(AB) = 0.2$, $P(A^cC^c) = 0.3$, $P(A) = 0.6$, $P(C) = 0.5$, and $P(AB^cC^c) = 0.1$. Determine $P(A^cC^c \cup AC)$, $P(AB^c \cup A^c)C^c)$, and $P(A^c(B \cup C^c))$, if possible.

Answer available at the end of this section.

Exercise 2.3.8

Suppose $P(A) = 0.6$, $P(C) = 0.4$, $P(AC) = 0.3$, $P(A^cB) = 0.2$ and $P(A^cB^cC^c) = 0.1$.

Determine $P((A \cup B)C^c$, $P(AC^c \cup A^cC)$, and $P(AC^c \cup A^cB)$, if possible.

Answer available at the end of this section.

Exercise 2.3.9

Suppose $P(A) = 0.5$, $P(AB) = P(AC) = 0.3$, and $P(ABC^c) = 0.1$.

Determine $P(A(BC^c)^c$ and $P(AB \cup AC \cup BC)$.

Then repeat with additional data $P(A^cB^cC^c) = 0.1$ and $P(A^cBC) = 0.05$

Answer available at the end of this section.

Exercise 2.3.10

Given $P(A) = 0.6$, $P(A^cB^c) = 0.2$, $P(AC^c) = 0.4$, and $P(ACD^c) = 0.1$.

Determine $P(A^cB \cup A(C^c \cup D))$.

Answer available at the end of this section.

Exercise 2.3.11

A survey of a represenative group of students yields the following information:

- 52 percent are male
- 85 percent live on campus
- 78 percent are male or are active in intramural sports (or both)
- 30 percent live on campus but are not active in sports
- 32 percent are male, live on campus, and are active in sports
- 8 percent are male and live off campus
- 17 percent are male students inactive in sports

a. What is the probability that a randomly chosen student is male and lives on campus?
b. What is the probability of a male, on campus student who is not active in sports?
c. What is the probability of a female student active in sports?

Answer available at the end of this section.

Exercise 2.3.12

A survey of 100 persons of voting age reveals that 60 are male, 30 of whom do not identify with a political party; 50 are members of a political party; 20 nonmembers of a party voted in the last election, 10 of whom are female. How many nonmembers of a political party did not vote? *Suggestion* Express the numbers as a fraction, and treat as probabilities.

Answer available at the end of this section.

Exercise 2.3.13

During a period of unsettled weather, let *A* be the event of rain in Austin, *B* be the event of rain in Houston, and *C* be the event of rain in San Antonio. Suppose:

$$P(AB) = 0.35,\ P(AB^c) = 0.15,\ P(AC) = 0.20,\ P(AB^c \cup A^cB) = 0.45$$

$$P(BC) = 0.30\ P(B^cC) = 0.05\ P(A^cB^cC^c) = 0.15$$

a. What is the probability of rain in all three cities?
b. What is the probability of rain in exactly two of the three cities?
c. What is the probability of rain in exactly one of the cities?

Answer available at the end of this section.

Exercise 2.3.14

One hundred students are questioned about their course of study and plans for graduate study. Let $A =$ the event the student is male; $B =$ the event the student is studying engineering; $C =$ the event the student plans at least one year of foreign language; $D =$ the event the student is planning graduate study (including professional school). The results of the survey are:

There are 55 men students; 23 engineering students, 10 of whom are women; 75 students will take foreign language classes, including all of the women; 26 men and 19 women plan graduate study; 13 male engineering students and 8 women engineering students plan graduate study; 20 engineering students will take a foreign language and plan graduate study; 5 non engineering students plan graduate study but no foreign language courses; 11 non engineering, women students plan foreign language study and graduate study.

a. What is the probability of selecting a student who plans foreign language classes and graduate study?
b. What is the probability of selecting a women engineer who does not plan graduate study?
c. What is the probability of selecting a male student who either studies a foreign language but does not intend graduate study or will not study a foreign language but plans graduate study?

Answer available at the end of this section.

Exercise 2.3.15

A survey of 100 students shows that: 60 are men students; 55 students live on campus, 25 of whom are women; 40 read the student newspaper regularly, 25 of whom are women; 70 consider themselves reasonably active in student affairs—50 of these live on campus; 35 of the reasonably active students read the newspaper regularly; All women who live on campus and 5 who live off campus consider themselves to be active; 10 of the on-campus women readers consider themselves active, as do 5 of the off campus women; 5 men are active, off-campus, non readers of the newspaper.

a. How many active men are either not readers or off campus?
b. How many inactive men are not regular readers?

Answer available at the end of this section.

Exercise 2.3.16

A television station runs a telephone survey to determine how many persons in its primary viewing area have watched three recent special programs, which we call a, b, and c. Of the 1000 persons surveyed, the results are:

221 have seen at least a; 209 have seen at least b; 112 have seen at least c; 197 have seen at least two of the programs; 45 have seen all three; 62 have seen at least a and c; the number having seen at least a and b is twice as large as the number who have seen at least b and c.

- (a) How many have seen at least one special?
- (b) How many have seen only one special program?

Answer available at the end of this section.

Exercise 2.3.17

An automobile safety inspection station found that in 1000 cars tested:

- 100 needed wheel alignment, brake repair, and headlight adjustment
- 325 needed at least two of these three items
- 125 needed headlight and brake work
- 550 needed at wheel alignment

a. How many needed only wheel alignment?
b. How many who do not need wheel alignment need one or none of the other items?

Answer available at the end of this section.

Exercise 2.3.18

Suppose $P(A(B \cup C)) = 0.3$, $P(A^c) = 0.6$, and $P(A^cB^cC^c) = 0.1$.

Determine $P(B \cup C)$, $P((AB \cup A^cB^c)C^c \cup AC)$, and $P(A^c(B \cup C^c))$, if possible.

Repeat the problem with he additional data $P(A^cBC) = 0.2$ and $P(A^cB) = 0.3$.

Answer available at the end of this section.

Exercise 2.3.19

A computer store sells computers, monitors, printers. A customer enters the store. Let *A*, *B*, *C* be the respective events the customer buys a computer, a monitor, a printer. Assume the following probabilities:

- The probability $P(AB)$ of buying both a computer and a monitor is 0.49.
- The probability $P(ABC^c)$ of buying both a computer and a monitor but not a printer is 0.17.
- The probability $P(AC)$ of buying both a computer and a printer is 0.45.
- The probability $P(BC)$ of buying both a monitor and a printer is 0.39
- The probability $P(AC^c \bigvee A^cC)$ of buying a computer or a printer, but not both is 0.50.
- The probability $P(AB^c \bigvee A^cB)$ of buying a computer or a monitor, but not both is 0.43.
- The probability $P(BC^c \bigvee B^cC)$ of buying a monitor or a printer, but not both is 0.43.

a. What is the probability $P(A)$, $P(B)$, or $P(C)$ of buying each?
b. What is the probability of buying exactly two of the three items?
c. What is the probability of buying at least two?
d. What is the probability of buying all three?

Answer available at the end of this section.

Exercise 2.3.20

Data are $P(A) = 0.232$, $P(B) = 0.228$, $P(ABC) = 0.045$, $P(AC) = 0.062$, $P(AB \cup AC \cup BC) = 0.197$ and $P(BC0 = 2P(AC)$.

Determine $P(A \cup B \cup C)$ and $P(A^cB^cC)$, if possible.

Repeat, with the additional data $P(C) = 0.230$.

Answer available at the end of this section.

Exercise 2.3.21

Data are: $P(A) = 0.4$, $P(AB) = 0.3$, $P(ABC) = 0.25$, $P(C) = 0.65$, $P(A^cC^c) = 0.3$. Determine available minterm probabilities and the following,

if computable:

$$P(AC^c \cup A^cC),\ P(A^cB^c),\ P(A \cup B),\ P(AB^c)$$

With only six items of data (including $P(\Omega) = P(A \bigvee A^c) = 1$, not all minterms are available. Try the additional data $P(A^cBC^c) = 0.1$ and $P(A^cB^c) = 0.3$. Are these consistent and linearly independent? Are all minterm probabilities available?

Answer available at the end of this section.

Exercise 2.3.22

Repeat Exercise with $P(AB)$ changed from 0.3 to 0.5. What is the result? Explain the reason for this result.

Answer available at the end of this section.

Exercise 2.3.23

Repeat Exercise with the original data probability matrix, but with AB replaced by AC in the data vector matrix. What is the result? Does mincalc work in this case? Check results on a minterm map.

Answer available at the end of this section.

Answers

Exercise 2.3.1

Use the pattern $P(E \cup F) = P(E) + P(E^cF)$ and $(A \cup C)^c = A^cC^c$.

$P(A \cup C \cup B \cup D) = P(A \cup C) + P(A^cC^c(B \cup D))$, so that $P(A^cC^c(B \cup D)) = 0.90 - 0.75 = 0.15$

Exercise 2.3.2

We use the MATLAB procedure, which displays the essential patterns.

```
minvec3
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
E = A|~(B&C);
F = A|B|(Bc&Cc);
disp([E;F])
     1     1     1     0     1     1     1     1   % Not equal
     1     0     1     1     1     1     1     1
G = ~(A|B);
H = (Ac&C)|(Bc&C);
disp([G;H])
     1     1     0     0     0     0     0     0   % Not equal
     0     1     0     1     0     1     0     0
K = (A&B)|(A&C)|(B&C);
disp([A;K])
     0     0     0     0     1     1     1     1   % A not contained in K
     0     0     0     1     0     1     1     1
```

Exercise 2.3.3

We use the MATLAB procedure, which displays the essential patterns.

```
minvec3
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
E = (A&(B|Cc))|(Ac&B&C);
F = (A&((B&C)|Cc))|(Ac&B);
disp([E;F])
     0     0     0     1     1     0     1     1   % E subset of F
     0     0     1     1     1     0     1     1
G = A|(Ac&B&C);
H = (A&B)|(B&C)|(A&C)|(A&Bc&Cc);
disp([G;H])
     0     0     0     1     1     1     1     1    % G = H
     0     0     0     1     1     1     1     1
```

Exercise 2.3.4

We use mintable(4) and determine positions with correct number(s) of ones (number of occurrences). An alternate is to use minvec4 and express the Boolean combinations which give the correct number(s) of ones.

```
npr02_04
Minterm probabilities are in pm.  Use mintable(4)
a = mintable(4);
s = sum(a);         % Number of ones in each minterm position
P1 = (s>=3)*pm'   % Select and add minterm probabilities
P1 =  0.4716
P2 = (s==2)*pm'
P2 =  0.3728
P3 = (s<=2)*pm'
P3 =  0.5284
```

Exercise 2.3.5

We use mintable(5) and determine positions with correct number(s) of ones (number of occurrences).

```
npr02_05
Minterm probabilities are in pm.  Use mintable(5)
a = mintable(5);
s = sum(a);         % Number of ones in each minterm position
P1 = (s>=3)*pm'   % Select and add minterm probabilities
P1 =  0.5380
P2 = (s==4)*pm'
P2 =  0.1712
P3 = (s<=3)*pm'
P3 =  0.7952
P4 = ((s==2)|(s==4))*pm'
P4 =  0.4784
```

Exercise 2.3.6

```
% file npr02_06.m       % Data file
% Data for Exercise 2.3.6.
minvec3
DV = [A|Ac; A|(Bc&C); A&C; Ac&B; Ac&Cc; B&Cc];
DP = [1      0.65     0.20 0.25  0.25   0.30];
TV = [((A&Cc)|(Ac&C))&Bc; ((A&Bc)|Ac)&Cc; Ac&(B|Cc)];
disp('Call for mincalc')
npr02_06             % Call for data
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent

 Computable target probabilities
    1.0000    0.3000     % The first and third target probability
    3.0000    0.3500     % is calculated. Check with minterm map.
The number of minterms is 8
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.7

```
% file npr02_07.m
% Data for Exercise 2.3.7.
minvec3
DV = [A|Ac; ((A&Bc)|(Ac&B))&C; A&B; Ac&Cc;  A;  C; A&Bc&Cc];
```

```
DP = [ 1        0.4            0.2   0.3   0.6 0.5   0.1];
TV = [(Ac&Cc)|(A&C); ((A&Bc)|Ac)&Cc; Ac&(B|Cc)];
disp('Call for mincalc')
npr02_07             % Call for data
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.7000    % All target probabilities calculable
    2.0000    0.4000    % even though not all minterms are available
    3.0000    0.4000
The number of minterms is 8
The number of available minterms is 6
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.8

```
% file npr02_08.m
% Data for Exercise 2.3.8.
minvec3
DV = [A|Ac; A;  C;  A&C; Ac&B; Ac&Bc&Cc];
DP = [ 1   0.6 0.4  0.3  0.2     0.1];
TV = [(A|B)&Cc; (A&Cc)|(Ac&C); (A&Cc)|(Ac&B)];
disp('Call for mincalc')


npr02_08             % Call for data
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.5000    % All target probabilities calculable
    2.0000    0.4000    % even though not all minterms are available
    3.0000    0.5000
The number of minterms is 8
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.9

```
% file npr02_09.m
% Data for Exercise 2.3.9.
minvec3
DV = [A|Ac;  A; A&B; A&C; A&B&Cc];
DP = [ 1    0.5 0.3  0.3   0.1];
TV = [A&(~(B&Cc)); (A&B)|(A&C)|(B&C)];
disp('Call for mincalc')

% Modification for part 2
% DV = [DV; Ac&Bc&Cc; Ac&B&C];
% DP = [DP 0.1 0.05];
npr02_09             % Call for data
Variables are A, B, C, Ac, Bc, Cc
```

```
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.4000    % Only the first target probability calculable
The number of minterms is 8
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
DV = [DV; Ac&Bc&Cc; Ac&B&C];  % Modification of data
DP = [DP 0.1 0.05];
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.4000             % Both target probabilities calculable
    2.0000    0.4500             % even though not all minterms are available
The number of minterms is 8
The number of available minterms is 6
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.10

```
% file npr02_10.m
% Data for Exercise 2.3.10.
minvec4
DV = [A|Ac;  A;  Ac&Bc; A&Cc; A&C&Dc];
DP = [1     0.6  0.2    0.4    0.1];
TV = [(Ac&B)|(A&(Cc|D))];
disp('Call for mincalc')
npr02_10
Variables are A, B, C, D, Ac, Bc, Cc, Dc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.7000             % Checks with minterm map solution
The number of minterms is 16
The number of available minterms is 0
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.11

```
% file npr02_11.m
% Data for Exercise 2.3.11.
% A = male;  B = on campus;  C = active in sports
minvec3
DV = [A|Ac;  A;   B;  A|C; B&Cc; A&B&C; A&Bc; A&Cc];
DP = [ 1    0.52 0.85 0.78 0.30  0.32   0.08 0.17];
TV = [A&B; A&B&Cc; Ac&C];
disp('Call for mincalc')

npr02_11
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
```

```
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.4400
    2.0000    0.1200
    3.0000    0.2600
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.12

```
% file npr02_12.m
% Data for Exercise 2.3.12.
% A = male;  B = party member; C = voted last election
minvec3
DV = [A|Ac;  A;  A&Bc;  B;  Bc&C; Ac&Bc&C];
DP = [  1   0.60 0.30  0.50 0.20  0.10];
TV = [Bc&Cc];
disp('Call for mincalc')
npr02_12
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.3000
The number of minterms is 8
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.13

```
% file npr02_13.m
% Data for Exercise 2.3.13.
% A = rain in Austin;  B = rain in Houston;
% C = rain in San Antonio
minvec3
DV = [A|Ac; A&B; A&Bc; A&C; (A&Bc)|(Ac&B); B&C; Bc&C; Ac&Bc&Cc];
DP = [  1   0.35 0.15  0.20    0.45        0.30 0.05   0.15];
TV = [A&B&C; (A&B&Cc)|(A&Bc&C)|(Ac&B&C); (A&Bc&Cc)|(Ac&B&Cc)|(Ac&Bc&C)];
disp('Call for mincalc')
npr02_13
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.2000
    2.0000    0.2500
    3.0000    0.4000
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.14

```
% file npr02_14.m
% Data for Exercise 2.3.14.
% A = male;  B = engineering;
% C = foreign language; D = graduate study
minvec4
DV = [A|Ac; A; B; Ac&B; C; Ac&C; A&D; Ac&D; A&B&D; ...
      Ac&B&D; B&C&D; Bc&Cc&D; Ac&Bc&C&D];
DP = [1 0.55 0.23 0.10 0.75 0.45 0.26 0.19 0.13 0.08 0.20 0.05 0.11];
TV = [C&D; Ac&Dc; A&((C&Dc)|(Cc&D))];
disp('Call for mincalc')
npr02_14
Variables are A, B, C, D, Ac, Bc, Cc, Dc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.3900
    2.0000    0.2600          % Third target probability not calculable
The number of minterms is 16
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.15

```
% file npr02_15.m
% Data for Exercise 2.3.15.
% A = men; B = on campus; C = readers; D = active
minvec4
DV = [A|Ac; A;  B;  Ac&B;  C;  Ac&C;  D;  B&D; C&D; ...
     Ac&B&D; Ac&Bc&D; Ac&B&C&D; Ac&Bc&C&D; A&Bc&Cc&D];
DP = [1  0.6 0.55 0.25 0.40 0.25 0.70 0.50 0.35 0.25 0.05 0.10 0.05 0.05];
TV = [A&D&(Cc|Bc); A&Dc&Cc];
disp('Call for mincalc')
npr02_15
Variables are A, B, C, D, Ac, Bc, Cc, Dc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.3000
    2.0000    0.2500
The number of minterms is 16
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.16

```
% file npr02_16.m
% Data for Exercise 2.3.16.
minvec3
DV = [A|Ac; A;    B;    C; (A&B)|(A&C)|(B&C); A&B&C; A&C; (A&B)-2*(B&C)];
DP = [ 1  0.221 0.209 0.112   0.197           0.045  0.062      0];
TV = [A|B|C; (A&Bc&Cc)|(Ac&B&Cc)|(Ac&Bc&C)];
```

```
npr02_16
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.3000
    2.0000    0.1030
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.17

```
% file npr02_17.m
% Data for Exercise 2.3.17.
% A = alignment;  B = brake work;  C = headlight
minvec3
DV = [A|Ac; A&B&C; (A&B)|(A&C)|(B&C); B&C;    A  ];
DP = [ 1    0.100      0.325          0.125 0.550];
TV = [A&Bc&Cc; Ac&(~(B&C))];
disp('Call for mincalc')
npr02_17
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.2500
    2.0000    0.4250
The number of minterms is 8
The number of available minterms is 3
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.18

```
% file npr02_18.m
% Date for Exercise 2.3.18.
minvec3
DV = [A|Ac; A&(B|C); Ac; Ac&Bc&Cc];
DP = [ 1     0.3     0.6    0.1];
TV = [B|C; (((A&B)|(Ac&Bc))&Cc)|(A&C); Ac&(B|Cc)];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&C; Ac&B];
% DP = [DP   0.2     0.3];
npr02_18
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.8000
    2.0000    0.4000
```

```
The number of minterms is 8
The number of available minterms is 2
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
DV = [DV; Ac&B&C; Ac&B];      % Modified data
DP = [DP   0.2     0.3];
mincalc                       % New calculation
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.8000
    2.0000    0.4000
    3.0000    0.4000
The number of minterms is 8
The number of available minterms is 5
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.19

```
% file npr02_19.m
% Data for Exercise 2.3.19.
% A = computer;  B = monitor;  C = printer
minvec3
DV = [A|Ac; A&B; A&B&Cc; A&C; B&C; (A&Cc)|(Ac&C); ...
          (A&Bc)|(Ac&B); (B&Cc)|(Bc&C)];
DP = [1 0.49 0.17 0.45 0.39 0.50 0.43 0.43];
TV = [A; B; C; (A&B&Cc)|(A&Bc&C)|(Ac&B&C); (A&B)|(A&C)|(B&C); A&B&C];
disp('Call for mincalc')
npr02_19
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.8000
    2.0000    0.6100
    3.0000    0.6000
    4.0000    0.3700
    5.0000    0.6900
    6.0000    0.3200
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.20

```
% file npr02_20.m
% Data for Exercise 2.3.20.
minvec3
DV = [A|Ac; A;     B;  A&B&C; A&C; (A&B)|(A&C)|(B&C); B&C - 2*(A&C)];
DP = [  1  0.232 0.228 0.045 0.062      0.197            0];
TV = [A|B|C; Ac&Bc&Cc];
disp('Call for mincalc')
% Modification
% DV = [DV; C];
% DP = [DP  0.230 ];

```

```
npr02_20
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
mincalc
Data vectors are linearly independent
Data probabilities are INCONSISTENT
The number of minterms is 8
The number of available minterms is 6
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
disp(PMA)
    2.0000    0.0480
    3.0000   -0.0450    % Negative minterm probabilities indicate
    4.0000   -0.0100    % inconsistency of data
    5.0000    0.0170
    6.0000    0.1800
    7.0000    0.0450
DV = [DV; C];
DP = [DP 0.230];
mincalc
Data vectors are linearly independent
Data probabilities are INCONSISTENT
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.21

```
% file npr02_21.m
% Data for Exercise 2.3.21.
minvec3
DV = [A|Ac; A;  A&B; A&B&C;  C;  Ac&Cc];
DP = [ 1   0.4  0.3  0.25   0.65  0.3 ];
TV = [(A&Cc)|(Ac&C); Ac&Bc; A|B; A&Bc];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&Cc; Ac&Bc];
% DP = [DP   0.1      0.3 ];
```

```
npr02_21
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
    1.0000    0.3500
    4.0000    0.1000
The number of minterms is 8
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
DV = [DV; Ac&B&Cc; Ac&Bc];
DP = [DP   0.1      0.3 ];
mincalc
Data vectors are linearly independent
 Computable target probabilities
```

```
    1.0000    0.3500
    2.0000    0.3000
    3.0000    0.7000
    4.0000    0.1000
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

Exercise 2.3.22

```
% file npr02_22.m
% Data for Exercise 2.3.22.
minvec3
DV = [A|Ac; A;  A&B; A&B&C;  C;  Ac&Cc];
DP = [ 1   0.4  0.5  0.25   0.65  0.3 ];
TV = [(A&Cc)|(Ac&C); Ac&Bc; A|B; A&Bc];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&Cc; Ac&Bc];
% DP = [DP   0.1      0.3 ];
```

```
npr02_22
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
Data probabilities are INCONSISTENT
The number of minterms is 8
The number of available minterms is 4
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
disp(PMA)
    4.0000   -0.2000
    5.0000    0.1000
    6.0000    0.2500
    7.0000    0.2500
DV = [DV; Ac&B&Cc; Ac&Bc];
DP = [DP   0.1      0.3 ];
mincalc
Data vectors are linearly independent
Data probabilities are INCONSISTENT
The number of minterms is 8
The number of available minterms is 8
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
disp(PMA)
         0    0.2000
    1.0000    0.1000
    2.0000    0.1000
    3.0000    0.2000
    4.0000   -0.2000
    5.0000    0.1000
    6.0000    0.2500
    7.0000    0.2500
```

Exercise 2.3.23

```
% file npr02_23.m
% Data for Exercise 2.3.23.
minvec3
DV = [A|Ac; A;  A&C; A&B&C;  C;  Ac&Cc];
DP = [ 1   0.4  0.3  0.25   0.65  0.3 ];
TV = [(A&Cc)|(Ac&C); Ac&Bc; A|B; A&Bc];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&Cc; Ac&Bc];
% DP = [DP   0.1      0.3 ];
npr02_23
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are NOT linearly independent
Warning: Rank deficient, rank = 5  tol =    5.0243e-15
 Computable target probabilities
    1.0000    0.4500
The number of minterms is 8
The number of available minterms is 2
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
DV = [DV; Ac&B&Cc; Ac&Bc];
DP = [DP   0.1      0.3 ];
mincalc
Data vectors are NOT linearly independent
Warning: Matrix is singular to working precision.
 Computable target probabilities
     1   Inf             % Note that p(4) and p(7) are given in data
     2   Inf
     3   Inf
The number of minterms is 8
The number of available minterms is 6
Available minterm probabilities are in vector pma
To view available minterm probabilities, call for PMA
```

This page titled 2.3: Problems on Minterm Analysis is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

3: CONDITIONAL PROBABILITY

The probability $P(A)$ of an event A is a measure of the likelihood that the event will occur on any trial. Sometimes partial information determines that an event C has occurred. Given this information, it may be necessary to reassign the likelihood for each event A. This leads to the notion of conditional probability. For a fixed conditioning event C, this assignment to all events constitutes a new probability measure which has all the properties of the original probability measure. In addition, because of the way it is derived from the original, the conditional probability measure has a number of special properties which are important in applications.

3.1: Conditional Probability

3.2: Problems on Conditional Probability

This page titled 3: Conditional Probability is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Conditional Probability

The *original* or *prior* probability measure utilizes all available information to make probability assignments $P(A)$, $P(B)$, etc., subject to the defining conditions (P1), (P2), and (P3). The probability $P(A)$ indicates the *likelihood* that event *A* will occur on any trial.

Frequently, *new information* is received which leads to a reassessment of the likelihood of event *A*. For example

- An applicant for a job as a manager of a service department is being interviewed. His résumé shows adequate experience and other qualifications. He conducts himself with ease and is quite articulate in his interview. He is considered a prospect highly likely to succeed. The interview is followed by an extensive background check. His credit rating, because of bad debts, is found to be quite low. With this information, the likelihood that he is a satisfactory candidate changes radically.
- A young woman is seeking to purchase a used car. She finds one that appears to be an excellent buy. It looks "clean," has reasonable mileage, and is a dependable model of a well known make. Before buying, she has a mechanic friend look at it. He finds evidence that the car has been wrecked with possible frame damage that has been repaired. The likelihood the car will be satisfactory is thus reduced considerably.
- A physician is conducting a routine physical examination on a patient in her seventies. She is somewhat overweight. He suspects that she may be prone to heart problems. Then he discovers that she exercises regularly, eats a low fat, high fiber, variagated diet, and comes from a family in which survival well into their nineties is common. On the basis of this new information, he reassesses the likelihood of heart problems.

New, but partial, information determines a *conditioning event* C, which may call for reassessing the likelihood of event A. For one thing, this means that A occurs iff the event AC occurs. Effectively, this makes C a new basic space. The new unit of probability mass is $P(C)$. How should the new probability assignments be made? One possibility is to make the new assignment to A proportional to the probability $P(AC)$. These considerations and experience with the classical case suggests the following procedure for reassignment. Although such a reassignment is not logically necessary, subsequent developments give substantial evidence that this is the appropriate procedure.

Definition

If C is an even having prositive probabilty, the *conditional probability of* A, *given* C is

$$P(A|C) = \frac{P(AC)}{P(C)}$$

For a fixed conditioning event C, we have a new likelihood assignment to the event A. Now

$$P(A|C) \geq 0,\ P(\Omega|C) = 1, \text{ and } P(\bigvee_j A_j|C) = \frac{P(\bigvee_j A_j C}{P(C)} = \sum_j P(A_j C)/P(C) = \sum_j P(A_j|C)$$

Thus, the new function $P(\cdot|C)$ satisfies the three defining properties (P1), (P2), and (P3) for probability, so that for fixed *C*, we have a *new probability measure*, with all the properties of an ordinary probability measure.

Remark. When we write $P(A|C)$ we are evaluating the likelihood of event A when it is known that event C has occurred. This is *not* the probability of a conditional event $A|C$. Conditional events have no meaning in the model we are developing.

Example 3.1.1 Conditional probabilities from joint frequency data

A survey of student opinion on a proposed national health care program included 250 students, of whom 150 were undergraduates and 100 were graduate students. Their responses were categorized Y (affirmative), N (negative), and D (uncertain or no opinion). Results are tabulated below.

	Y	N	D

U	60	40	50
G	70	20	10

Suppose the sample is representative, so the results can be taken as typical of the student body. A student is picked at random. Let *Y* be the event he or she is favorable to the plan, *N* be the event he or she is unfavorable, and *D* is the event of no opinion (or uncertain). Let *U* be the event the student is an undergraduate and *G* be the event he or she is a graduate student. The data may reasonably be interpreted

$$P(G) = 100/250,\ P(U) = 150/250,\ P(Y) = (60 + 70)/250,\ P(YU) = 60/250, \text{ etc.}$$

Then

$$P(Y|U) = \frac{P(YU)}{P(U)} = \frac{60/250}{150/250} = \frac{60}{150}$$

Similarly, we can calculate

$$P(N|U) = 40/150,\ P(D|U) = 50/150,\ P(Y|G) = 70/100,\ P(N|G) = 20/100,\ P(D|G) = 10/100$$

We may also calculate directly

$$P(U|Y) = 60/130,\ P(G|N) = 20/60, \text{ etc.}$$

Conditional probability often provides a natural way to deal with compound trials carried out in several steps.

Example 3.1.2 Jet aircraft with two engines

An aircraft has two jet engines. It will fly with only one engine operating. Let F_1 be the event one engine fails on a long distance flight, and F_2 the event the second fails. Experience indicates that $P(F_1) = 0.0003$. Once the first engine fails, added load is placed on the second, so that $P(F_2|F_1) = 0.001$. Now the second engine can fail only if the other has already failed. Thus $F_2 \subset F_1$ so that

$$P(F_2) = P(F_1F_2) = P(F_1)P(F_2|F_1) = 3 \times 10^{-7}$$

Thus reliability of any one engine may be less than satisfactory, yet the overall reliability may be quite high.

The following example is taken from the UMAP Module 576, by Paul Mullenix, reprinted in UMAP Journal, vol 2, no. 4. More extensive treatment of the problem is given there.

Example 3.1.3 Responses to a sensitive question on a survey

In a survey, if answering "yes" to a question may tend to incriminate or otherwise embarrass the subject, the response given may be incorrect or misleading. Nonetheless, it may be desirable to obtain correct responses for purposes of social analysis. The following device for dealing with this problem is attributed to B. G. Greenberg. By a chance process, each subject is instructed to do one of three things:

1. Respond with an honest answer to the question.
2. Respond "yes" to the question, regardless of the truth in the matter.
3. Respond "no" regardless of the true answer.

Let *A* be the event the subject is told to reply honestly, *B* be the event the subject is instructed to reply "yes," and *C* be the event the answer is to be "no." The probabilities $P(A)$, $P(B)$, and $P(C)$ are determined by a chance mechanism (i.e., a fractio $P(A)$ selected randomly are told to answer honestly, etc.). Let E be the event the reply is "yes." We wish to calculate $P(E|A)$, the probability the answer is "yes" given the response is honest.

Solution

Since $E = EA \bigvee B$, we have

$$P(E) = P(EA) + P(B) = P(E|A)P(A) + P(B)$$

which may be solved algebraically to give

$$P(E|A) = \frac{P(E) - P(B)}{P(A)}$$

Suppose there are 250 subjects. The chance mechanism is such that $P(A) = 0.7$, $P(B) = 0.4$ and $P(C) = 0.16$. There are 62 responses "yes," which we take to mean $P(E) = 62/250$. According to the pattern above

$$P(E|A) = \frac{62/250 - 14/100}{70/100} = \frac{27}{175} \approx 0.154$$

The formulation of conditional probability assumes the conditioning event *C* is well defined. Sometimes there are subtle difficulties. It may not be entirely clear from the problem description what the conditioning event is. This is usually due to some ambiguity or misunderstanding of the information provided.

Example 3.1.4 What is the conditioning event?

Five equally qualified candidates for a job, Jim, Paul, Richard, Barry, and Evan, are identified on the basis of interviews and told that they are finalists. Three of these are to be selected at random, with results to be posted the next day. One of them, Jim, has a friend in the personnel office. Jim asks the friend to tell him the name of one of those selected (other than himself). The friend tells Jim that Richard has been selected. Jim analyzes the problem as follows.

Analysis

Let A_i, $1 \le i \le 5$ be the event the ith of these is hired (A_1 is the event Jim is hired, A_3 is the event Richard is hired, etc.). Now $P(A_i)$ (for each i) is the probability that finalist i is in one of the combinations of three from five. Thus, Jim's probability of being hired, before receiving the information about Richard, is

$$P(A_1) = \frac{1 \times C(4,2)}{C(5,3)} = \frac{6}{10} = P(A_i),\ 1 \le i \le 5$$

The information that Richard is one of those hired is information that the event A_3 has occurred. Also, for any pair $i \ne j$ the number of combinations of three from five including these two is just the number of ways of picking one from the remaining three. Hence,

$$P(A_1A_3) = \frac{C(3,1)}{C(5,3)} = \frac{3}{10} = P(A_iA_j), i \ne j$$

The conditional probability

$$P(A_1|A_3) = \frac{P(A_1A_3)}{P(A_3)} = \frac{3/10}{6/10} = 1/2$$

This is consistent with the fact that if Jim knows that Richard is hired, then there are two to be selected from the four remaining finalists, so that

$$P(A_1|A_3) = \frac{1 \times C(3,1)}{C(4,2)} = \frac{3}{6} = 1/2$$

Discussion

Although this solution seems straightforward, it has been challenged as being incomplete. Many feel that there must be information about *how* the friend chose to name Richard. Many would make an assumption somewhat as follows. The friend took the three names selected: if Jim was one of them, Jim's name was removed and an equally likely choice among the other two was made; otherwise, the friend selected on an equally likely basis one of the three to be hired. Under this assumption, the information assumed is an event B_3 which is *not* the same as A_3. In fact, computation (see Example 5, below) shows

$$P(A_1|B_3) = \frac{6}{10} = P(A_1) \neq P(A_1|A_3)$$

Both results are mathematically correct. The difference is in the conditioning event, which corresponds to the difference in the information given (or assumed).

Some properties

In addition to its properties as a probability measure, conditional probability has *special properties* which are consequences of the way it is related to the original probability measure $P(\cdot)$. The following are easily derived from the definition of conditional probability and basic properties of the prior probability measure, and prove useful in a variety of problem situations.

(CP1) Product rule If $P(ABCD) > 0$, then $P(ABCD) = P(A)P(B|A)P(C|AB)P(D|ABC)$.

Derivation

The defining expression may be written in product form: $P(AB) = P(A)P(B|A)$. Likewise

$$P(ABC) = P(A)\frac{P(AB)}{P(A)} \cdot \frac{P(ABC)}{P(AB)} = P(A)P(B|A)P(C|AB)$$

and

$$P(ABCD) = P(A)\frac{P(AB)}{P(A)} \cdot \frac{P(ABC)}{P(AB)} \cdot \frac{P(ABCD)}{P(ABC)} = P(A)P(B|A)P(C|AB)P(D|ABC)$$

This pattern may be extended to the intersection of any finite number of events. Also, the events may be taken in any order.

— □

Example 3.1.5 Selection of items from a lot

An electronics store has ten items of a given type in stock. One is defective. Four successive customers purchase one of the items. Each time, the selection is on an equally likely basis from those remaining. What is the probability that all four customes get good items?

Solution

Let E_i be the event the ith customer receives a good item. Then the first chooses one of the nine out of ten good ones, the second chooses one of the eight out of nine goood ones, etc., so that

$$P(E_1E_2E_3E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3) = \frac{9}{10} \cdot \frac{8}{9} \cdot \frac{7}{8} \cdot \frac{6}{7} = \frac{6}{10}$$

Note that this result could be determined by a combinatorial argument: under the assumptions, each combination of four of ten is equally likely; the number of combinations of four good ones is the number of combinations of four of the nine. Hence

$$P(E_1E_2E_3E_4) = \frac{C(9,4)}{C(10,4)} = \frac{126}{210} = 3/5$$

Example 3.1.6 A selection problem

Three items are to be selected (on an equally likely basis at each step) from ten, two of which are defective. Determine the probability that the first and third selected are good.

Solution

Let G_i, $1 \le i \le 3$ be the even the ith unit selected is good. Then $G_1G_3 = G_1G_2G_3 \bigvee G_1G_2^cG_3$. By the product rule

$$P(G_1G_3) = P(G_1)P(G_2|G_1)P(G_3|G_1G_2) + P(G_1)P(G_2^c|G_1)P(G_3|G_1G_2^c) = \frac{8}{10}\cdot\frac{7}{9}\cdot\frac{6}{8} + \frac{8}{10}\cdot\frac{2}{9}\cdot\frac{7}{8} = \frac{28}{45} \approx 0.6$$

(CP2) Law of total probability Suppose the class $\{A_i : 1 \le i \le n\}$ of events is mutually exclusive and every outcome in *E* is in one of these events. Thus, $E = A_1E \bigvee A_2E \bigvee \cdots \bigvee A_nE$, a disjoint union. Then

$$P(E) = P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + \cdots + P(E|A_n)P(A_n)$$

Example 3.1.7 a compound experiment

Five cards are numbered one through five. A two-step selection procedure is carried out as follows.

1. Three cards are selected without replacement, on an equally likely basis.
   - If card 1 is drawn, the other two are put in a box
   - If card 1 is not drawn, all three are put in a box
2. One of cards in the box is drawn on an equally likely basis (from either two or three)

Let A_i be the event the ith card is drawn on the first selection and let B_i be the event the card numbered i is drawn on the second selection (from the box). Determine $P(B_5)$, $P(A_1B_5)$, and $P(A_1|B_5)$.

Solution

From Example 3.1.4, we have $P(A_i) = 6/10$ and $P(A_iA_j) = 3/10$. This implies

$$P(A_iA_j^c) = P(A_i) - P(A_iA_j) = 3/10$$

Now we can draw card five on the second selection only if it is selected on the first drawing, so that $B_5 \subset A_5$. Also $A_5 = A_1A_5 \bigvee A_1^cA_5$. We therefore have $B_5 = B_5A_5 = B_5A_1A_5 \bigvee B_5A_1^cA_5$. By the law of total probability (CP2),

$$P(B_5) = P(B_5|A_1A_5)P(A_1A_5) + P(B_5|A_1^cA_5)P(A_1^cA_5) = \frac{1}{2}\cdot\frac{3}{10} + \frac{1}{3}\cdot\frac{3}{10} = \frac{1}{4}$$

Also, since $A_1B_5 = A_1A_5B_5$,

$$P(A_1B_5) = P(A_1A_5B_50 = P(A_1A_5)P(B_5|A_1A_5) = \frac{3}{10}\cdot\frac{1}{2} = \frac{3}{20}$$

We thus have

$$P(A_1|B_5) = \frac{3/20}{5/20} = \frac{6}{10} = P(A_1)$$

Occurrence of event B_1 has no affect on the likelihood of the occurrence of A_1. This condition is examined more thoroughly in the chapter on "Independence of Events".

Often in applications data lead to conditioning with respect to an event but the problem calls for “conditioning in the opposite direction.”

Example 3.1.8 Reversal of conditioning

Students in a freshman mathematics class come from three different high schools. Their mathematical preparation varies. In order to group them appropriately in class sections, they are given a diagnostic test. Let H_i be the event that a student tested is from high school i, $1 \le i \le 3$. Let *F* be the event the student fails the test. Suppose data indicate

$$P(H_1) = 0.2,\ P(H_2) = 0.5,\ P(H_3) = 0.3,\ P(F|H_1) = 0.10,\ P(F|H_2) = 0.02,\ P(F|H_3) = 0.06$$

A student passes the exam. Determine for each i the conditional probability $P(H_i|F^c)$ that the student is from high school i.

Solution

$$P(F^c) = P(F^c|H_1)P(H_1) + P(F^c|H_2)P(H_2) + P(F^c|H_3)P(H_3) = 0.90 \cdot 0.2 + 0.98 \cdot 0.5 + 0.94 \cdot 0.3 = 0.952$$

Then

$$P(H_1|F^c) = \frac{P(F^cH_1)}{P(F^c)} = \frac{P(F^c|H_1)P(H_1)}{P(F^c)} = \frac{180}{952} = 0.1891$$

Similarly,

$$P(H_2|F^c) = \frac{P(F^c|H_2)P(H_2)}{P(F^c)} = \frac{590}{952} = 0.5147 \text{ and } P(H_3|F^c) = \frac{P(F^c|H_3)P(H_3)}{P(F^c)} = \frac{282}{952} = 0.2962$$

The basic pattern utilized in the reversal is the following.

(CP3) Bayes' rule If $E \subset \bigvee_{i=1}^{n} A_i$ (as in the law of total probability), then

$$P(A_i|E) = \frac{P(A_iE)}{P(E)} = \frac{P(E|A_i)P(A_i)}{P(E)} \quad 1 \le i \le n \text{ The law of total probabilty yields } P(E)$$

Such reversals are desirable in a variety of practical situations.

Example 3.1.9 A compound selection and reversal

Begin with items in two lots:

1. Three items, one defective.
2. Four items, one defective.

One item is selected from lot 1 (on an equally likely basis); this item is added to lot 2; a selection is then made from lot 2 (also on an equally likely basis). This second item is good. What is the probability the item selected from lot 1 was good?

Solution

Let G_1 be the event the first item (from lot 1) was good, and G_2 be the event the second item (from the augmented lot 2) is good. We want to determine $P(G_1|G_2)$. Now the data are interpreted as

$$P(G_1) = 2/3,\ P(G_2|G_1) = 4/5,\ P(G_2|G_1^c) = 3/5$$

By the law of total probability (CP2),

$$P(G_2) = P(G_1)P(G_2|G_1) + P(G_1^c)P(G_2|G_1^c) = \frac{2}{3} \cdot \frac{4}{5} + \frac{1}{3} \cdot \frac{3}{5} = \frac{11}{15}$$

By Bayes' rule (CP3),

$$P(G_1|G_2) = \frac{P(G_2|G_1)P(G_1)}{P(G_2)} = \frac{4/5 \times 2/3}{11/15} = \frac{8}{11} \approx 0.73$$

Example 3.1.10 Additional problems requiring reversals

- *Medical tests*. Suppose *D* is the event a patient has a certain disease and *T* is the event a test for the disease is positive. Data are usually of the form: prior probability $P(D)$ (or prior odds $P(D)/P(D^c)$), probability $P(T|D^c)$ of a *false positive*, and probability $P(T^c|D)$ of a false negative. The desired probabilities are $P(D|T)$ and $P(D^c|T^c)$.
- *Safety alarm*. If *D* is the event a dangerous condition exists (say a steam pressure is too high) and *T* is the event the safety alarm operates, then data are usually of the form $P(D)$, $P(T|D^c)$, and $P(T^c|D)$, or equivalently (e.g., $P(T^c|D^c)$ and $P(T|D)$). Again, the desired probabilities are that the safety alarms signals correctly, $P(D|T)$ and $P(D^c|T^c)$.
- *Job success*. If *H* is the event of success on a job, and *E* is the event that an individual interviewed has certain desirable characteristics, the data are usually prior $P(H)$ and reliability of the characteristics as predictors in the form $P(H)$ and $P(E|H^c)$. The desired probability is $P(H|E)$.
- *Presence of oil*. If *H* is the event of the presence of oil at a proposed well site, and *E* is the event of certain geological structure (salt dome or fault), the data are usually $P(H)$ (or the odds), $P(E|H)$, and $P(E|H^c)$. The desired probability is $P(H|E)$.
- *Market condition*. Before launching a new product on the national market, a firm usually examines the condition of a test market as an indicator of the national market. If *H* is the event the national market is favorable and *E* is the event the test market is favorable, data are a prior estimate $P(H)$ of the likelihood the national market is sound, and data $P(E|H)$ and $P(E|H^c)$ indicating the reliability of the test market. What is desired is $P(H|E)$, the likelihood the national market is favorable, given the test market is favorable.

The calculations, as in Example 3.8, are simple but can be tedious. We have an m-procedure called *bayes* to perform the calculations easily. The probabilities $P(A_i)$ are put into a matrix PA and the conditional probabilities $P(E|A_i)$ are put into matrix PEA. The desired probabilities $P(A_i|E)$ and $PA_i|E^c)$ are calculated and displayed

Example 3.1.11 matlab calculations for

```
>> PEA = [0.10 0.02 0.06];
>> PA =  [0.2 0.5 0.3];
>> bayes
Requires input PEA = [P(E|A1) P(E|A2) ... P(E|An)]
and PA = [P(A1) P(A2) ... P(An)]
Determines PAE  = [P(A1|E) P(A2|E) ... P(An|E)]
       and PAEc = [P(A1|Ec) P(A2|Ec) ... P(An|Ec)]
Enter matrix PEA of conditional probabilities  PEA
Enter matrix  PA of probabilities  PA
P(E) = 0.048
P(E|Ai)   P(Ai)     P(Ai|E)   P(Ai|Ec)
0.1000    0.2000    0.4167    0.1891
0.0200    0.5000    0.2083    0.5147
0.0600    0.3000    0.3750    0.2962
Various quantities are in the matrices PEA, PA, PAE, PAEc, named above
```

The procedure displays the results in tabular form, as shown. In addition, the various quantities are in the workspace in the matrices named, so that they may be used in further calculations without recopying.

The following variation of Bayes' rule is applicable in many practical situations.

(CP3*) Ratio form of Bayes' rule $\frac{P(A|C)}{P(B|C)} = \frac{P(AC)}{P(BC)} = \frac{P(C|A)}{P(C|B)} \cdot \frac{P(A)}{P(B)}$

The left hand member is called the *posterior odds*, which is the odds *after* knowledge of the occurrence of the conditioning event.

The second fraction in the right hand member is the *prior odds*, which is the odds *before* knowledge of the occurrence of the conditioning event C. The first fraction in the right hand member is known as the *likelihood ratio*. It is the ratio of the probabilities (or likelihoods) of C for the two different probability measures $P(\cdot|A)$ and $P(\cdot|B)$.

Example 3.1.12 A performance test

As a part of a routine maintenance procedure, a computer is given a performance test. The machine seems to be operating so well that the prior odds it is satisfactory are taken to be ten to one. The test has probability 0.05 of a false positive and 0.01 of a false negative. A test is performed. The result is positive. What are the posterior odds the device is operating properly?

Solution

Let S be the event the computer is operating satisfactorily and let T be the event the test is favorable. The data are $P(S)/P(S^c) = 10$, $P(T|S^c) = 0.05$, and $P(T^c|S) = 0.01$.Then by the ratio form of Bayes' rule

$$\frac{P(S|T)}{P(S^c|T)} = \frac{P(T|S)}{P(T|S^c)} \cdot \frac{P(S)}{P(S^c)} = \frac{0.99}{0.05} \cdot 10 = 198 \text{ so that } P(S|T) = \frac{198}{199} = 0.9950$$

The following property serves to establish in the chapters on "Independence of Events" and "Conditional Independence" a number of important properties for the concept of independence and of conditional independence of events.

(CP4) Some equivalent conditions If $0 < P(A) < 1$ and $0 < P(B) < 1$, then

$$P(A|B) * P(A) \text{ iff } P(B|A) * P(B) \text{ iff } P(AB) * P(A)P(B) \text{ and}$$

$$P(AB) * P(A)P(B) \text{ iff } P(A^cB^c) * P(A^c)P(B^c) \text{ iff } P(AB^c) \diamond P(A)P(B^c)$$

where * is $<, \leq, =, \geq,$ or $>$ and $\diamond$ is $>, \geq, =, \leq,$ or $<$, respectively.

Because of the role of this property in the theory of independence and conditional independence, we examine the derivation of these results.

VERIFICATION of (CP4)

$P(AB) * P(A)P(B)$ iff $P(A|B) * P(A)$ (divide by $P(B)$ - may exchange A and A^c
$P(AB) * P(A)P(B)$ iff $P(B|A) * P(B)$ (divide by $P(A)$ - may exchange B and B^c
$P(AB) * P(A)P(B)$ iff $[P(A) - P(AB^c)] * P(A)[1 - P(B^c)]$ iff $-P(AB^c) * -P(A)P(B^c)$ iff $P(AB^c) \diamond P(A)P(B^c)$
we may use c to get $P(AB) * P(A)P(B)$ iff $P(AB^C) \diamond P(A)P(B^c)$ iff $P(A^cB^c) * P(A^c)P(B^c)$

— □

A number of important and useful propositons may be derived from these.

$P(A|B) + P(A^c|B) = 1$, but, in general, $P(A|B) + P(A|B^c) \neq 1$.
$P(A|B) > P(A)$ iff $P(A|B^c) < P(A)$.
$P(A^c|B) > P(A^c)$ iff $P(A|B) < P(A)$.
$P(A|B) > P(A)$ iff $P(A^c|B^c) > P(A^c)$.

VERIFICATION — Exercises (see problem set)

— □

Repeated conditioning

Suppose conditioning by the event C has occurred. Additional information is then received that event *D* has occurred. We have a new conditioning event CD. There are two possibilities:

Reassign the conditional probabilities. $P_C(A)$ becomes

$$P_C(A|D) = \frac{P_C(AD)}{P_C(D)} = \frac{P(ACD)}{P(CD)}$$

Reassign the total probabilities: $P(A)$ becomes

$$P_{CD}(A) = P(A|CD) = \frac{P(ACD)}{P(CD)}$$

Basic result: $P_C(A|D) = P(A|CD) = P_D(A|C)$. Thus repeated conditioning by two events may be done in any order, or may be done in one step. This result extends easily to repeated conditioning by any finite number of events. This result is important in extending the concept of "Independence of Events" to "Conditional Independence". These conditions are important for many problems of probable inference.

This page titled 3.1: Conditional Probability is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

3.2: Problems on Conditional Probability

Exercise 3.2.1

Given the following data:

$P(A) = 0.55$, $P(AB) = 0.30$, $P(BC) = 0.20$, $P(A^c \cup BC) = 0.55$, $P(A^cBC^c) = 0.15$

Determine, if possible, the conditional probability $P(A^c|B) = P(A^cB)/P(B)$.

Answer available at the end of this section.

Exercise 3.2.2

In Exercise 11 from "Problems on Minterm Analysis," we have the following data: A survey of a represenative group of students yields the following information:

- 52 percent are male
- 85 percent live on campus
- 78 percent are male or are active in intramural sports (or both)
- 30 percent live on campus but are not active in sports
- 32 percent are male, live on campus, and are active in sports
- 8 percent are male and live off campus
- 17 percent are male students inactive in sports

Let A = male, B = on campus, C = active in sports.

a. A student is selected at random. He is male and lives on campus. What is the (conditional) probability that he is active in sports?
b. A student selected is active in sports. What is the(conditional) probability that she is a female who lives on campus?

Answer available at the end of this section.

Exercise 3.2.3

In a certain population, the probability a woman lives to at least seventy years is 0.70 and is 0.55 that she will live to at least eighty years. If a woman is seventy years old, what is the conditional probability she will survive to eighty years? *Note* that if $A \subset B$ then $P(AB) = P(A)$.

Answer available at the end of this section.

Exercise 3.2.4

From 100 cards numbered 00, 01, 02, $\cdots$, 99, one card is drawn. Suppose A_i is the event the sum of the two digits on a card is i, $0 \le i \le 18$, and B_j is the event the product of the two digits is j. Determine $P(A_i|B_0)$ for each possible i.

Answer available at the end of this section.

Exercise 3.2.5

Two fair dice are rolled.

a. What is the (conditional) probability that one turns up two spots, given they show different numbers?
b. What is the (conditional) probability that the first turns up six, given that the sum is k, for each k from two through 12?
c. What is the (conditional) probability that at least one turns up six, given that the sum is k, for each k from two through 12?

Answer

a. There are 6×5 ways to choose all different. There are 2×5 ways that they are different and one turns up two spots. The conditional probability is 2/6.

b. Let A_6 = event first is a six and S_k = event the sum is k. Now $A_6 S_k = \emptyset$ for $k \leq 6$. A table of sums shows $P(A_6 S_k) = 1/36$ and $P(S_k) = 6/36, 5/36, 4/36, 3/36, 2/36, 1/36$ for $k = 7$ through 12, respectively. Hence $P(A_6|S_k) = 1/6, 1/5.1/4, 1/3.1/2, 1$, respectively.

c. If AB_6 is the event at least one is a six, then $AB_6 S_k) = 2/36$ for $k = 7$ through 11 and $P(AB_6 S_1 2) = 1/36$. Thus, the conditional probabilities are 2/6, 2/5, 2/4, 2/3, 1, 1, respectively.

Exercise 3.2.6

Four persons are to be selected from a group of 12 people, 7 of whom are women.

a. What is the probability that the first and third selected are women?
b. What is the probability that three of those selected are women?
c. What is the (conditional) probability that the first and third selected are women, given that three of those selected are women?

Answer

$$P(W_1 W_3) = P(W_1 W_2 W_3) + P(W_1 W_2^c W_3) = \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} + \frac{7}{12} \cdot \frac{5}{11} \cdot \frac{6}{10} = \frac{7}{22}$$

Exercise 3.2.7

Twenty percent of the paintings in a gallery are not originals. A collector buys a painting. He has probability 0.10 of buying a fake for an original but never rejects an original as a fake, What is the (conditional) probability the painting he purchases is an original?

Answer available at the end of this section.

Exercise 3.2.8

Five percent of the units of a certain type of equipment brought in for service have a common defect. Experience shows that 93 percent of the units with this defect exhibit a certain behavioral characteristic, while only two percent of the units which do not have this defect exhibit that characteristic. A unit is examined and found to have the characteristic symptom. What is the conditional probability that the unit has the defect, given this behavior?

Answer available at the end of this section.

Exercise 3.2.9

A shipment of 1000 electronic units is received. There is an equally likely probability that there are 0, 1, 2, or 3 defective units in the lot. If one is selected at random and found to be good, what is the probability of no defective units in the lot?

Answer available at the end of this section.

Exercise 3.2.10

Data on incomes and salary ranges for a certain population are analyzed as follows. S_1= event annual income is less than \$25,000; S_2= event annual income is between \$25,000 and \$100,000; S_3= event annual income is greater than \$100,000. E_1= event did not complete college education; E_2= event of completion of bachelor's degree; E_3= event of completion of graduate or professional degree program. Data may be tabulated as follows: $P(E_1) = 0.65$, $P(E_2) = 0.30$ and $P(E_3) = 0.05$.

$$P(S_i|E_j)$$

	S_1	S_2	S_3
E_1	0.85	0.10	0.05
E_2	0.10	0.80	0.10
E_3	0.05	0.50	0.45
$P(S_i)$	0.50	0.40	0.10

a. Determine $P(E_3S_3)$.
b. Suppose a person has a university education (no graduate study). What is the (conditional) probability that he or she will make \$25,000 or more?
c. Find the total probability that a person's income category is at least as high as his or her educational level.

Answer available at the end of this section.

Exercise 3.2.11

In a survey, 85 percent of the employees say they favor a certain company policy. Previous experience indicates that 20 percent of those who do not favor the policy say that they do, out of fear of reprisal. What is the probability that an employee picked at random really does favor the company policy? It is reasonable to assume that all who favor say so.

Answer available at the end of this section.

Exercise 3.2.12

A quality control group is designing an automatic test procedure for compact disk players coming from a production line. Experience shows that one percent of the units produced are defective. The automatic test procedure has probability 0.05 of giving a false positive indication and probability 0.02 of giving a false negative. That is, if D is the event a unit tested is defective, and T is the event that it tests satisfactory, then $P(T|D) = 0.05$ and $P(T^c|D^c) = 0.02$. Determine the probability $P(D^c|T)$ that a unit which tests good is, in fact, free of defects.

Answer available at the end of this section.

Exercise 3.2.13

Five boxes of random access memory chips have 100 units per box. They have respectively one, two, three, four, and five defective units. A box is selected at random, on an equally likely basis, and a unit is selected at random therefrom. It is defective. What are the (conditional) probabilities the unit was selected from each of the boxes?

Answer available at the end of this section.

Exercise 3.2.14

Two percent of the units received at a warehouse are defective. A nondestructive test procedure gives two percent false positive indications and five percent false negative. Units which fail to pass the inspection are sold to a salvage firm. This firm applies a corrective procedure which does not affect any good unit and which corrects 90 percent of the defective units. A customer buys a unit from the salvage firm. It is good. What is the (conditional) probability the unit was originally defective?

Answer available at the end of this section.

Exercise 3.2.15

At a certain stage in a trial, the judge feels the odds are two to one the defendent is guilty. It is determined that the defendent is left handed. An investigator convinces the judge this is six times more likely if the defendent is guilty than if he were not. What is the likelihood, given this evidence, that the defendent is guilty?

Answer available at the end of this section.

Exercise 3.2.16

Show that if $P(A|C) > P(B|C)$ and $P(A|C^c) > P(B|C^c)$, then $P(A) > P(B)$. Is the converse true? Prove or give a counterexample.

Answer available at the end of this section.

Exercise 3.2.17

Since $P(\cdot|B)$ is a probability measure for a given B, we must have $P(A|B) + P(A^c|B) = 1$. Construct an example to show that in general $P(A|B) + P(A|B^c) \neq 1$.

Answer available at the end of this section.

Exercise 3.2.18

Use property (CP4) to show

a. $P(A|B) > P(A)$ iff $P(A|B^c) < P(A)$

b. $P(A^c|B) > P(A^c)$ iff $P(A|B) < P(A)$

c. $P(A|B) > P(A)$ iff $P(A^c|B^c) > P(A^c)$

Answer available at the end of this section.

Exercise 3.2.19

Show that $P(A|B) \geq (P(A) + P(B) - 1)/P(B)$.

Answer available at the end of this section.

Exercise 3.2.20

Show that $P(A|B) = P(A|BC)P(C|B) + P(A|BC^c)P(C^c|B)$.

Answer available at the end of this section.

Exercise 3.2.21

An individual is to select from among n alternatives in an attempt to obtain a particular one. This might be selection from answers on a multiple choice question, when only one is correct. Let A be the event he makes a correct selection, and B be the event he knows which is correct before making the selection. We suppose $P(B) = p$ and $P(A|B^c) = 1/n$. Determine $P(B|A)$; show that $P(B|A) \geq P(B)$ and $P(B|A)$ increases with n for fixed p.

Answer available at the end of this section.

Exercise 3.2.22

Polya's urn scheme for a contagious disease. An urn contains initially b black balls and r red balls $(r + b = n)$. A ball is drawn on an equally likely basis from among those in the urn, then replaced along with c additional balls of the same color. The process is repeated. There are n balls on the first choice, $n + c$ balls on the second choice, etc. Let B_k be the event of a black ball on the kth draw and R_k be the event of a red ball on the kth draw. Determine

a. $P(B_2|R_1)$
b. $P(B_1B_2)$
c. $P(R_2)$
d. $P(B_1|R_2)$

Answer available at the end of this section.

Answers

Exercise 3.2.1

```
% file npr03_01.m
% Data for Exercise 3.2.1.
minvec3
DV = [A|Ac; A;  A&B; B&C; Ac|(B&C); Ac&B&Cc];
DP = [ 1   0.55 0.30 0.20   0.55     0.15  ];
TV = [Ac&B; B];
disp('Call for mincalc')
npr03_01
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
Call for mincalc
mincalc
Data vectors are linearly independent
 Computable target probabilities
```

```
    1.0000    0.2500
    2.0000    0.5500
The number of minterms is 8
The number of available minterms is 4
- - - - - - - - - - - -
P = 0.25/0.55
P =  0.4545
```

Exercise 3.2.2

```
npr02_11
- - - - - - - - - - - -
mincalc
- - - - - - - - - - - -
mincalct
Enter matrix of target Boolean combinations  [A&B&C; A&B; Ac&B&C; C]
 Computable target probabilities
    1.0000    0.3200
    2.0000    0.4400
    3.0000    0.2300
    4.0000    0.6100
PC_AB = 0.32/0.44
PC_AB =  0.7273
PAcB_C = 0.23/0.61
PAcB_C = 0.3770
```

Exercise 3.2.3

Let $A =$ event she lives to seventy and $B =$ event she lives to eighty. Since $B \subset A$, $P(B|A) = P(AB)/P(A) = P(B)/P(A) = 55/70$.

Exercise 3.2.4

B_0 is the event one of the first ten is draw. A_iB_0 is the event that the card with numbers $0i$ is drawn. $P(a_i|B_0) = (1/100)/(1/10) = 1/10$ for each i, 0 through 9.

Exercise 3.2.7

Let $B =$ the event the collector buys, and $G =$ the event the painting is original. Assume $P(B|G) = 1$ and $P(B|G^c) = 0.1$. If $P(G) = 0.8$, then

$$P(G|B) = \frac{P(GB)}{P(B)} = \frac{P(B|G)P(G)}{P(B|G)P(G) + P(B|G^c)P(G^c)} = \frac{0.8}{0.8 + 0.1 \cdot 0.2} = \frac{40}{41}$$

Exercise 3.2.8

Let $D =$ the event the unit is defective and $C =$ the event it has the characteristic. Then $P(D) = 0.05$, $P(C|D) = 0.93$, and $P(C|D^c) = 0.02$.

$$P(D|C) = \frac{P(C|D)P(D)}{P(C|D)P(D) + P(C|D^c)P(D^c)} = \frac{0.93 \cdot 0.05}{0.93 \cdot 0.05 + 0.02 \cdot 0.95} = \frac{93}{131}$$

Exercise 3.2.9

Let $D_k =$ the event of k defective and G be the event a good one is chosen.

$$P(D_0|G) = \frac{P(G|D_0)P(D_0)}{P(G|D_0)P(D_0) + P(G|D_1)P(D_1) + P(G|D_2)P(D_2) + P(G|D_3)P(D_3)}$$
$$= \frac{1 \cdot 1/4}{(1/4)(1 + 999/1000 + 998/1000 + 997/1000)} = \frac{1000}{3994}$$

Exercise 3.2.10

a. $P(E_3S_3) = P(S_3|E_3)P(E_3) = 0.45 \cdot 0.05 = 0.0225$

b. $P(S_2 \vee S_3|E_2) = 0.80 + 0.10 = 0.90$

c. $p = (0.85 + 0.10 + 0.05) \cdot 0.65 + (0.80 + 0.10) \cdot 0.30 + 0.45 \cdot 0.05 = 0.9425$

Exercise 3.2.11

$P(S) = 0.85$, $P(S|F^c) = 0.20$. Also, reasonable to assume $P(S|F) = 1$.

$$P(S) = P(S|F)P(F) + P(S|F^c)[1 - P(F)] \text{ implies } P(F) = \frac{P(S) - P(S|F^c)}{1 - P(S|F^c)} = \frac{13}{16}$$

Exercise 3.2.12

$$\frac{D^c|T}{P(D|T)} = \frac{P(T|D^c)P(D^c)}{P(T|D)P(D)} = \frac{0.98 \cdot 0.99}{0.05 \cdot 0.01} = \frac{9702}{5}$$
$$P(D^c|T) = \frac{9702}{9707} = 1 - \frac{5}{9707}$$

Exercise 3.2.13

$H_i =$ the event from box i. $P(H_i) = 1/5$ and $P(D|H_i) = i/100$.

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum P(D|H_i)P(H_j)} = i/15,\ 1 \le i \le 5$$

Exercise 3.2.14

Let T = event test indicates defective, D = event initially defective, and G = event unit purchased is good. Data are

$$P(D) = 0.02,\ P(T^c|D) = 0.02,\ P(T|D^c) = 0.05,\ P(GT^c) = 0,$$
$$P(G|DT) = 0.90,\ P(G|D^cT) = 1$$
$$P(D|G) = \frac{P(GD)}{P(G)},\ P(GD) = P(GTD) = P(D)P(T|D)P(G|TD)$$
$$P(G) = P(GT) = P(GDT) + P(GD^cT) = P(D)P(T|D)P(G|TD) + P(D^c)P(T|D^c)P(G|TD^c)$$
$$P(D|G) = \frac{0.02 \cdot 0.98 \cdot 0.90}{0.02 \cdot 0.98 \cdot 0.90 + 0.98 \cdot 0.05 \cdot 1.00} = \frac{441}{1666}$$

Exercise 3.2.15

Let G = event the defendent is guilty, L = the event the defendent is left handed. Prior odds: $P(G)/P(G^c) = 2$. Result of testimony: $P(L|G)/P(L|G^c) = 6$.

$$\frac{P(G|L)}{P(G^c|L)} = \frac{P(G)}{P(G^c)} \cdot \frac{P(L|G)}{P(L|G^c)} = 2 \cdot 6 = 12$$
$$P(G|L) = 12/13$$

Exercise 3.2.16

$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) > P(B|C)P(C) + P(B|C^c)P(C^c) = P(B)$.

The converse is not true. Consider $P(C) = P(C^c) = 0.5$, $P(A|C) = 1/4$.

$P(A|C^c) = 3/4$, $P(B|C) = 1/2$, and $P(B|C^c) = 1/4$. Then

$$1/2 = P(A) = \frac{1}{2}(1/4 + 3/4) > \frac{1}{2}(1/2 + 1/4) = P(B) = 3/8$$

But $P(A|C) < P(B|C)$.

Exercise 3.2.17

Suppose $A \subset B$ with $P(A) < P(B)$. Then $P(A|B) = P(A)/P(B) < 1$ and $P(A|B^c) = 0$ so the sum is less than one.

Exercise 3.2.18

a. $P(A|B) > P(A)$ iff $P(AB) > P(A)P(B)$ iff $P(AB^c) < P(A)P(B^c)$ iff $P(A|B^c) < P(A)$

b. $P(A^c|B) > P(A^c)$ iff $P(A^cB) > P(A^c)P(B)$ iff $P(AB) < P(A)P(B)$ iff $P(A|B) < P(A)$

c. $P(A|B) > P(A)$ iff $P(AB) > P(A)P(B)$ iff $P(A^cB^c) > P(A^c)P(B^c)$ iff $P(A^c|B^c) > P(A^c)$

Exercise 3.2.19

$1 \geq P(A \cup B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A|B)P(B)$. Simple algebra gives the desired result.

Exercise 3.2.20

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(ABC) + P(ABC^c)}{P(B)}$$
$$= \frac{P(A|BC)P(BC) + P(A|BC^c)P(BC^c)}{P(B)} = P(A|BC)P(C|B) + P(A|BC^c)P(C^c|B)$$

Exercise 3.2.21

$P(A|B) = 1$, $P(A|B^c) = 1/n$, $P(B) = p$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{p}{p + \frac{1}{n}(1-p)} = \frac{np}{(n-1)p + 1}$$

$\frac{P(B|A)}{P(B)} = \frac{n}{np + 1 - p}$ increases from 1 to $1/p$ as $n \to \infty$

Exercise 3.2.22

a. $P(B_2|R_1) = \frac{b}{n+c}$

b. $P(B_1B_2) = P(B_2)P(B_2|B_1) = \frac{b}{n} \cdot \frac{b+c}{n+c}$

c. $P(R_2)P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1)$

$$= \frac{r+c}{n+c} \cdot \frac{r}{n} + \frac{r}{n+c} \cdot \frac{b}{n} = \frac{r(r+c+b)}{n(n+c)}$$

d. $P(B_1|R_2) = \frac{P(R_2|B_1)P(B_1)}{P(R_2)}$ with $P(R_2|B_1)P(B_1) = \frac{r}{n+c} \cdot \frac{b}{n}$. Using (c), we have

$$P(B_1|R_2) = \frac{b}{r+b+c} = \frac{b}{n+c}$$

This page titled 3.2: Problems on Conditional Probability is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

4: INDEPENDENCE OF EVENTS

4.1: Independence of Events
4.2: MATLAB and Independent Classes
4.3: Composite Trials
4.4: Problems on Independence of Events

This page titled 4: Independence of Events is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Independence of Events

Historically, the notion of independence has played a prominent role in probability. If events form an independent class, much less information is required to determine probabilities of Boolean combinations and calculations are correspondingly easier. In this unit, we give a precise formulation of the concept of independence in the probability sense. As in the case of all concepts which attempt to incorporate intuitive notions, the consequences must be evaluated for evidence that these ideas have been captured successfully.

Independence as lack of conditioning

There are many situations in which we have an "operational independence."

- Supose a deck of playing cards is shuffled and a card is selected at random then replaced with reshuffling. A second card picked on a repeated try should not be affected by the first choice.
- If customers come into a well stocked shop at different times, each unaware of the choice made by the others, the the item purchased by one should not be affected by the choice made by the other.
- If two students are taking exams in different courses, the grade one makes should not affect the grade made by the other.

The list of examples could be extended indefinitely. In each case, we should expect to model the events as independent in some way. How should we incorporate the concept in our developing model of probability?

We take our clue from the examples above. Pairs of events are considered. The "operational independence" described indicates that knowledge that one of the events has occured does not affect the likelihood that the other will occur. For a pair of events {A, B}, this is the condition

$$P(A|B) = P(A)$$

Occurrence of the event A is not "conditioned by" occurrence of the event B. Our basic interpretation is that $P(A)$ indicates of the likelihood of the occurrence of event A. The development of conditional probability in the module Conditional Probability, leads to the interpretation of $P(A|B)$ as the likelihood that A will occur on a trial, given knowledge that B as occurred. If such knowledge of the occurrence of B does not affect the likelihood of the occurrence of A, we should be inclined to think of the events A and B as being independent in a probability sense.

Independent pairs

We take our clue from the condition $P(A|B) = P(A)$. Property (CP4) for conditional probability (in the case of equality) yields *sixteen equivalent conditions* as follows.

$P(A\|B) = P(A)$	$P(B\|A) = P(B)$	$P(AB) = P(A)P(B)$
$P(A\|B^c) = P(A)$	$P(B^c\|A) = P(B^c)$	$P(AB^c) = P(A)P(B^c)$
$P(A^c\|B) = P(A^c)$	$P(B\|A^c) = P(B)$	$P(A^cB) = P(A^c)P(B)$
$P(A^c\|B^c) = P(A^c)$	$P(B^c\|A^c) = P(B^c)$	$P(A^cB^c) = P(A^c)P(B^c)$

$P(A\|B) = P(A\|B^c)$	$P(A^c\|B) = P(A^c\|B^c)$	$P(B\|A) = P(B\|A^c)$	$P(B^c\|A) = P(B^c\|A^c)$

These conditions are equivalent in the sense that if any one holds, then all hold. We may chose any one of these as the defining condition and consider the others as equivalents for the defining condition. Because of its simplicity and symmetry with respect to the two events, we adopt the *product rule* in the upper right hand corner of the table.

Definition. The pair {A, B} of events is said to be (stochastically) *independent* iff the following *product rule* holds:

$$P(AB) = P(A)P(B)$$

Remark. Although the product rule is adopted as the basis for definition, in many applications the assumptions leading to independence may be formulated more naturally in terms of one or another of the equivalent expressions. We are free to do this, for the effect of assuming any one condition is to assume them all.

The equivalences in the right-hand column of the upper portion of the table may be expressed as a *replacement rule*, which we augment and extend below:

If the pair {A, B} independent, so is any pair obtained by taking the complement of either or both of the events.

We note two relevant facts

- Suppose event N has probability zero (is a *null* event). Then for any event A, we have $0 \leq P(AN) \leq P(N) = 0 = P(A)P(N)$, so that the product rule holds. Thus {N, A} is an independent pair for any event A.
- If event S has probability one (is an *almost sure* event), then its complement S^c is a null event. By the replacement rule and the fact just established, \({S^c\), A} is independent, so {S, A} is independent.

The replacement rule may thus be extended to:

Replacement Rule

If the pair {A, B} independent, so is any pair obtained by replacing either or both of the events by their complements or by a null event or by an almost sure event.

CAUTION

1. Unless at least one of the events has probability one or zero, a pair cannot be both independent and mutually exclusive. Intuitively, if the pair is mutually exclusive, then the occurrence of one requires that the other does not occur. Formally: Suppose $0 < P(A) < 1$ and $0 < P(B) < 1$. {A, B} mutually exclusive implies $P(AB) = P(\emptyset) = 0 \neq P(A)P(B)$. {$A$, B} independent implies $P(AB) = P(A)P(B) > 0 = P(\emptyset)$
2. Independence is not a property of events. Two non mutually exclusive events may be independent under one probability measure, but may not be independent for another. This can be seen by considering various probability distributions on a Venn diagram or minterm map.

Independent classes

Extension of the concept of independence to an arbitrary *class* of events utilizes the product rule.

Definition. A *class* of events is said to be (stochastically) *independent* iff the product rule holds for every finite subclass of two or more events in the class.

A class {A, B, C} is independent iff all four of the following product rules hold

$$P(AB) = P(A)P(B)\ P(AC) = P(A)P(C)\ P(BC) = P(B)P(C)\ P(ABC) = P(A)P(B)P(C)$$

If any *one or more* of these product expressions fail, the class is *not* independent. A similar situation holds for a class of four events: the product rule must hold for every pair, for every triple, and for the whole class. *Note* that we say “not independent” or “nonindependent” rather than dependent. The reason for this becomes clearer in dealing with independent random variables.

We consider some classical exmples of nonindependent classes

SOME NONINDEPENDENT CLASSES

1. Suppose {A_1, A_2, A_3, A_4} is a partition, with each $P(A_i) = 1/4$. Let

   $A = A_1 \bigvee A_2 B = A_1 \bigvee A_3 C = A_1 \bigvee A_4$

Then the class $\{A, B, C\}$ has $P(A) = P(B) = P(C) = 1/2$ and is pairwise independent, but not independent, since $P(AB) = P(A_1) = 1/4 = P(A)P(B)$ and similarly for the other pairs, but $P(ABC) = P(A_1) = 1/4 \neq P(A)P(B)P(C)$

2. Consider the class $\{A, B, C, D\}$ with $AD = BD = \emptyset$, $C = AB \bigvee D$, $P(A) = P(B) = 1/4$, $P(AB) = 1/64$, and $P(D) = 15/64$. Use of a minterm maps shows these assignments are consistent. Elementary calculations show the product rule applies to the class $\{A, B, C\}$ but no two of these three events forms an independent pair.

As noted above, the replacement rule holds for any pair of events. It is easy to show, although somewhat cumbersome to write out, that if the rule holds for any finite number k of events in an independent class, it holds for any $k + 1$ of them. By the principle of mathematical induction, the rule must hold for any finite subclass. We may extend the replacement rule as follows.

General Replacement Rule

If a class is independent, we may replace any of the sets by its complement, by a null event, or by an almost sure event, and the resulting class is also independent. Such replacements may be made for any number of the sets in the class. One immediate and important consequence is the following.

Minterm Probabilities

If $\{A_i : 1 \leq i \leq n\}$ is an independent class and the the class $\{P(A_i) : 1 \leq i \leq n\}$ of individual probabilities is known, then the probability of every minterm may be calculated.

Minterm probabilities for an independent class

Suppose the class $\{A, B, C\}$ is independent with respective probabilities $P(A) = 0.3$, $P(B) = 0.6$, and $P(C) = 0.5$. Then

$\{A^c, B^c, C^c\}$ is independent and $P(M_0) = P(A^c)P(B^c)P(C^c) = 0.14$

$\{A^c, B^c, C\}$ is independent and $P(M_1) = P(A^c)P(B^c)P(C) = 0.14$

Similarly, the probabilities of the other six minterms, in order, are 0.21, 0.21, 0.06, 0.06, 0.09, and 0.09. With these minterm probabilities, the probability of any Boolean combination of A, B, and C may be calculated

In general, eight appropriate probabilities must be specified to determine the minterm probabilities for a class of three events. In the independent case, three appropriate probabilities are sufficient.

Three probabilities yield the minterm probabilities

Suppose $\{A, B, C\}$ is independent with $P(A \cup BC) = 0.51$, $P(AC^c) = 0.15$, and $P(A) = 0.30$. Then $P(C^c) = 0.15/0.3 = 0.5 = P(C)$ and

$$P(A) + P(A^c)P(B)P(C) = 0.51 \text{ so that } P(B) = \frac{0.51 - 0.30}{0.7 \times 0.5} = 0.6$$

With each of the basic probabilities determined, we may calculate the minterm probabilities, hence the probability of any Boolean combination of the events.

MATLAB and the product rule

Frequently we have a large enough independent class $\{E_1, E_2, \cdot\cdot\ cdot, E_n\}$ that it is desirable to use MATLAB (or some other computational aid) to calculate the probabilities of various "and" combinations (intersections) of the events or their complements. Suppose the independent class $\{E_1, E_2, \cdot\cdot\ cdot, E_{10}\}$ has respective probabilities

0.13 0.37 0.12 0.56 0.33 0.71 0.22 0.43 0.57 0.31

It is desired to calculate (a) $P(E_1E_2E_3^cE_4E_5^cE_6^cE_7)$, and (b) $P(E_1^cE_2E_3^cE_4E_5^cE_6^cE_7E_8E_9^cE_{10})$.

We may use the MATLAB function *prod* and the scheme for indexing a matrix.

```
>> p = 0.01*[13 37 12 56 33 71 22 43 57 31];
>> q = 1-p;
>> % First case
>> e = [1 2 4 7];                  % Uncomplemented positions
>> f = [3 5 6];                    % Complemented positions
>> P = prod(p(e))*prod(q(f))       % p(e) probs of uncomplemented factors
P = 0.0010                         % q(f) probs of complemented factors
>> % Case of uncomplemented in even positions; complemented in odd positions
>> g = find(rem(1:10,2) == 0);     % The even positions
>> h = find(rem(1:10,2) ~= 0);     % The odd positions
>> P = prod(p(g))*prod(q(h))
P = 0.0034
```

In the unit on MATLAB and Independent Classes, we extend the use of MATLAB in the calculations for such classes.

This page titled 4.1: Independence of Events is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

4.2: MATLAB and Independent Classes

MATLAB and Independent Classes

In the unit on Minterms, we show how to use minterm probabilities and minterm vectors to calculate probabilities of Boolean combinations of events. In Independence of Events we show that in the independent case, we may calculate all minterm probabilities from the probabilities of the basic events. While these calculations are straightforward, they may be tedious and subject to errors. Fortunately, in this case we have an m-function *minprob* which calculates all minterm probabilities from the probabilities of the basic or generating sets. This function uses the m-function mintable to set up the patterns of p's and q's for the various minterms and then takes the products to obtain the set of minterm probabilities.

Example 4.2.1

```
>> pm = minprob(0.1*[4 7 6])
   pm = 0.0720  0.1080  0.1680  0.2520  0.0480  0.0720  0.1120  0.1680
```

It may be desirable to arrange these as on a minterm map. For this we have an m-function *minmap* which reshapes the row matrix pm, as follows:

```
>> t = minmap(pm)
t = 0.0720    0.1680    0.0480    0.1120
    0.1080    0.2520    0.0720    0.1680
```

Probability of occurrence of k of n independent events

In Example 2, we show how to use the m-functions mintable and csort to obtain the probability of the occurrence of k of n events, when minterm probabilities are available. In the case of an independent class, the minterm probabilities are calculated easily by minprob, It is only necessary to specify the probabilities for the n basic events and the numbers k of events. The size of the class, hence the mintable, is determined, and the minterm probabilities are calculated by minprob. We have two useful m-functions. If P is a matrix of the n individual event probabilities, and k is a matrix of integers less than or equal to n, then

function $y = \text{ikn}(P, k)$ calculates individual probabilities that k of n occur

function $y = \text{ckn}(P, k)$ calculates the probabilities that k or more occur

Example 4.2.2

```
>> p = 0.01*[13 37 12 56 33 71 22 43 57 31];
>> k = [2 5 7];
>> P = ikn(p,k)
P =    0.1401    0.1845    0.0225       % individual probabilities
>> Pc = ckn(p,k)
Pc =   0.9516    0.2921    0.0266       % cumulative probabilities
```

Reliability of systems with independent components

Suppose a system has n components which fail independently. Let E_i be the event the ith component survives the designated time period. Then $R_i = P(E_i)$ is defined to be the *reliability* of that component. The reliability R of the complete system is a function of the component reliabilities. There are three basic configurations. General systems may be decomposed into subsystems of these types. The subsystems become components in the larger configuration. The three fundamental configurations are:

Series. The system operates iff *all n* components *operate*: $R = \prod_{i=1}^{n} R_i$

Parallel. The system operates iff *not all* components *fail*: $R = 1 - \prod_{i=1}^{n}(1 - R_i)$

k of n. The system operates iff k or more components operate. R may be calculated with the m-function ckn. If the component probabilities are all the same, it is more efficient to use the m-function cbinom (see Bernoulli trials and the binomial distribution, below).

MATLAB solution. Put the component reliabilities in matrix $RC = [R_1\ R_2\ \cdots\ R_n]$

Series Configuration

```
>> R = prod(RC)     % prod is a built in MATLAB function
```

Parallel Configuration

```
>> R = parallel(RC) % parallel is a user defined function
```

k of n Configuration

```
>> R = ckn(RC,k)    % ckn is a user defined function (in file ckn.m).
```

Example 4.2.3

There are eight components, numbered 1 through 8. Component 1 is in series with a parallel combination of components 2 and 3, followed by a 3 of 5 combination of components 4 through 8 (see Figure 1 for a schematic representation). Probabilities of the components in order are

0.95 0.90 0.92 0.80 0.83 0.91 0.85 0.85

The second and third probabilities are for the parallel pair, and the last five probabilities are for the 3 of 5 combination.

```
>> RC = 0.01*[95 90 92 80 83 91 85 85];        % Component reliabilities
>> Ra = RC(1)*parallel(RC(2:3))*ckn(RC(4:8),3) % Solution
Ra = 0.9172
```

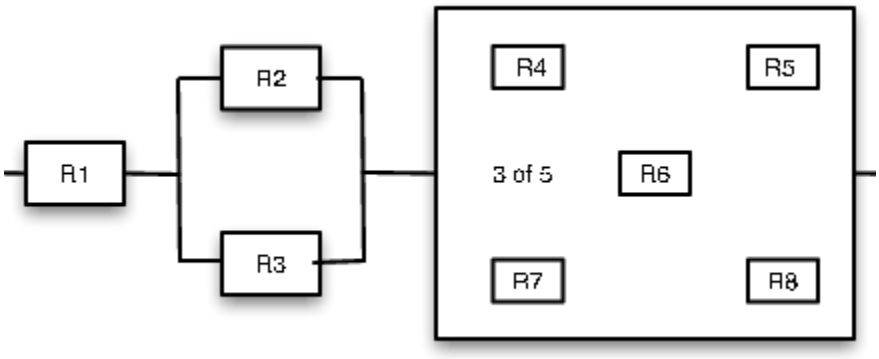


Figure 4.2.1. Schematic representation of the system in Example

Example 4.2.4

```
>> RC = 0.01*[95 90 92 80 83 91 85 85];    % Component reliabilities 1--8
```

```
>> Rb = prod(RC(1:2))*parallel([RC(3),ckn(RC(4:8),3)])     % Solution
Rb = 0.8532
```

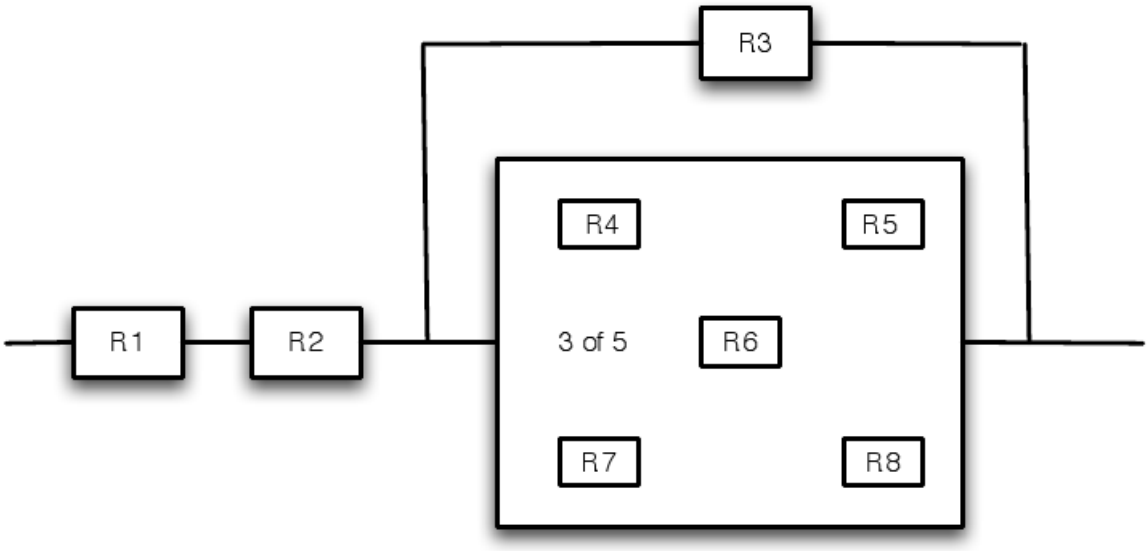


Figure 4.2.2. Schematic representation of the system in Example

A test for independence

It is difficult to look at a list of minterm probabilities and determine whether or not the generating events form an independent class. The m-function *imintest* has as argument a vector of minterm probabilities. It checks for feasible size, determines the number of variables, and performs a check for independence.

Example 4.2.5

```
>> pm = 0.01*[15 5 2 18 25 5 18 12];   % An arbitrary class
>> disp(imintest(pm))
The class is NOT independent
Minterms for which the product rule fails
     1     1     1     0
     1     1     1     0
```

Example 4.2.6

```
>> pm = [0.10 0.15 0.20 0.25 0.30]: %An improper number of probabilities
>> disp(imintest(pm))
The number of minterm probabilities incorrect
```

Example 4.2.7

```
>> pm = minprob([0.5 0.3 0.7]);
>> disp(imintest(pm))
The class is independent
```

Probabilities of Boolean combinations

As in the nonindependent case, we may utilize the minterm expansion and the minterm probabilities to calculate the probabilities of Boolean combinations of events. However, it is frequently more efficient to manipulate the expressions for the Boolean combination to be a disjoint union of intersections.

Example 4.2.8 A simple Boolean combination

Suppose the class $\{A, B, C\}$ is independent, with respective probabilities 0.4, 0.6, 0.8. Determine $P(A \cup BC)$. The minterm expansion is

$$A \cup BC = M(3, 4, 5, 6, 7), \text{ so that } P(A \cup BC) = p(3, 4, 5, 6, 7)$$

It is not difficult to use the product rule and the replacement theorem to calculate the needed minterm probabilities. Thus $p(3) = P(A^c)P(B) = P(C) = 0.6 \cdot 0.6 \cdot 0.8 = 0.2280$. Similarly $p(4) = 0.0320$, $p(5) = 0.1280$, $p(6) = 0.0480$, $p(7) = 0.1920$. The desired probability is the sum of these, 0.6880.

As an alternate approach, we write

$$A \cup BC = A \bigvee A^c BC, \text{ so that } P(A \cup BC) = 0.4 + 0.6 \cdot 0.6 \cdot 0.8 = 0.6880$$

Considerbly fewer arithmetic operations are required in this calculation.

In larger problems, or in situations where probabilities of several Boolean combinations are to be determined, it may be desirable to calculate all minterm probabilities then use the minterm vector techniques introduced earlier to calculate probabilities for various Boolean combinations. As a larger example for which computational aid is highly desirable, consider again the class and the probabilities utilized in Example 4.2.2, above.

Example 4.2.9

Consider again the independent class $\{E_1, E_2, \cdots E_{10}\}$ with respective probabilities [0.13 0.37 0.12 0.56 0.33 0.71 0.22 0.43 0.57 0.31]. We wish to calculate

$$P(F) = P(E_1 \cup E_3(E_4 \cup E_7^c) \cup E_2(E_5^c \cup E_6E_8) \cup E_9E_{10}^c)$$

There are $2^{10} = 1024$ minterm probabilities to be calculated. Each requires the multiplication of ten numbers. The solution with MATLAB is easy, as follows:

```
>> P = 0.01*[13 37 12 56 33 71 22 43 57 31];
>> minvec10
Vectors are A1 thru A10 and A1c thru A10c
They may be renamed, if desired.
>> F = (A1|(A3&(A4|A7c)))|(A2&(A5c|(A6&A8)))|(A9&A10c);
>> pm = minprob(P);
>> PF = F*pm'
PF =  0.6636
```

Writing out the expression for F is tedious and error prone. We could simplify as follows:

```
>> A = A1|(A3&(A4|A7c));
>> B = A2&(A5c|(A6&A8));
>> C = A9&A10c;
>> F = A|B|C;              % This minterm vector is the same as for F above
```

This decomposition of the problem indicates that it may be solved as a series of smaller problems. First, we need some central

facts about independence of Boolean combinations.

Independent Boolean combinations

Suppose we have a Boolean combination of the events in the class $\{A_i : 1 \le i \le n\}$ and a second combination the events in the class $\{B_j : 1 \le j \le m\}$. If the combined class $\{A_i, B_j : 1 \le i \le n, 1 \le j \le m\}$ is independent, we would expect the combinations of the subclasses to be independent. It is important to see that this is in fact a consequence of the product rule, for it is further evidence that the product rule has captured the essence of the intuitive notion of independence. In the following discussion, we exhibit the essential structure which provides the basis for the following general proposition.

Proposition. Consider n distinct subclasses of an independent class of events. If for each i the event $A_i is a Boolean (logical) combination of members of the \$ith subclass, then the class {\(A_1, A_2, \cdot\cdot\cdot, A_n} is an independent class.

Verification of this far reaching result rests on the minterm expansion and two elementary facts about the disjoint subclasses of an independent class. We state these facts and consider in each case an example which exhibits the essential structure. Formulation of the general result, in each case, is simply a matter of careful use of notation.

A class each of whose members is a minterm formed by members of a distinct subclass of an independent class is itself an independent class.

Example 4.2.10

Consider the independent class $\{A_1, A_2, A_3, B_1, B_2, B_3, B_4\}$, with respective probabilities 0.4, 0.7, 0.3, 0.5, 0.8, 0.3, 0.6. Consider M_3, minterm three for the class $\{A_1, A_2, A_3\}$, and N_5, minterm five for the class $\{B_1, B_2, B_3, B_4\}$. Then

$$P(M_3) = P(A_1^c A_2 A_3) = 0.6 \cdot 0.7 \cdot 0.3 = 0.126 \text{ and } P(N_5) = P(B_1^c B_2 B_3^c B_4) = 0.5 \cdot 0.8 \cdot 0.7 \cdot 0.6 = 0.168$$

Also

$$P(M_3 N_5) = P(A_1^c A_2 A_3 B_1^c B_2 B_3^c B_4 0 = 0.6 \cdot 0.7 \cdot 0.3 \cdot 0.5 \cdot 0.8 \cdot 0.7 \cdot 0.6$$

$$= (0.6 \cdot 0.7 \cdot 0.3) \cdot (0.5 \cdot 0.8 \cdot 0.7 \cdot 0.6) = P(M_3)P(N_5) = 0.0212$$

The product rule shows the desired independence.

Again, it should be apparent that the result holds for any number of A_i and B_j; and it can be extended to any number of distinct subclasses of an independent class.

Suppose each member of a class can be expressed as a disjoint union. If each auxiliary class formed by taking one member from each of the disjoint unions is an independent class, then the original class is independent.

Example 4.2.11

Suppose $A = A_1 \bigvee A_2 \bigvee A_3$ and $B = B_1 \bigvee B_2$, with $\{A_i, A_j\}$ independent for each pair i, j. Suppose

$P(A_1) = 0.3$, $P(A_2) = 0.4$, $P(A_3) = 0.1$, $P(B_1) = 0.2$, $P(B_2) = 0.5$

We wish to show that the pair $\{A, B\}$ is independent; i.e., the product rule $P(AB) = P(A)P(B)$ holds.

COMPUTATION

$$P(A) = P(A_1) + P(A_2) + P(A_3) = 0.3 + 0.4 + 0.1 = 0.8 \text{ and } P(B) = P(B_1) + P(B_2) = 0.2 + 0.5 = 0.7$$

Now

$$AB = (A_1 \bigvee A_2 \bigvee A_3)(B_1 \bigvee B_2) = A_1B_1 \bigvee A_1B_2 \bigvee A_2B_1 \bigvee A_2B_2 \bigvee A_3B_1 \bigvee A_3B_2$$

By additivity and pairwise independence, we have

$$P(AB) = P(A_1)P(B_1) + P(A_1)P(B_2) + P(A_2)P(B_1) + P(A_2)P(B_2) + P(A_3)P(B_1) + P(A_3)P(B_2)$$
$$= 0.3 \cdot 0.2 + 0.3 \cdot 0.5 + 0.4 \cdot 0.2 + 0.4 \cdot 0.5 + 0.1 \cdot 0.2 + 0.1 \cdot 0.5 = 0.56 = P(A)P(B)$$

The product rule can also be established algebraically from the expression for $P(AB)$, as follows:

$$P(AB) = P(A_1)[P(B_1) + P(B_2)] + P(A_2)[P(B_1) + P(B_2)] + P(A_3)[P(B_1) + P(B_2)]$$
$$= [P(A_1) + P(A_2) + P(A_3)][P(B_1) + P(B_2)] = P(A)P(B)$$

It should be clear that the pattern just illustrated can be extended to the general case. If

$$A = \bigvee_{i=1}^{n} A_i \text{ and } B = \bigvee_{j=1}^{m} B_j, \text{ with each pair } \{A_i, B_j\} \text{ independent}$$

then the pair $\{A, B\}$ is independent. Also, we may extend this rule to the triple $\{A, B, C\}$

$$A = \bigvee_{i=1}^{n} A_i,\ B = \bigvee_{j=1}^{m} B_j, \text{ and } C = \bigvee_{k=1}^{r} C_k, \text{ with each class } \{A_i, B_j, C_k\} \text{ independent}$$

and similarly for any finite number of such combinations, so that the second proposition holds.

Begin with an independent class $\mathscr{E}$ of n events. Select m distinct subclasses and form Boolean combinations for each of these. Use of the minterm expansion for each of these Boolean combinations and the two propositions just illustrated shows that the class of Boolean combinations is independent

To illustrate, we return to Example 4.2.9, which involves an independent class of ten events.

Example 4.2.12 A hybrid approach

Consider again the independent class $\{E_1, E_2, \cdots, E_{10}\}$ with respective probabilities {0.13 0.37 0.12 0.56 0.33 0.71 0.22 0.43 0.57 0.31}. We wish to calculate

$$P(F) = P(E_1 \cup E_3(E_4 \cup E_7^c) \cup E_2(E_5^c \cup E_6E_8) \cup E_9E_10^c)$$

In the previous solution, we use minprob to calculate the $2^{10} = 1024$ minterms for all ten of the E_i and determine the minterm vector for F. As we note in the alternate expansion of F,

$$F = A \cup B \cup C, \text{ when } A = E_1 \cup E_3(E_4 \cup E_7^c)\ B = E_2(E_5^c \cup E_6E_8)\ C = E_9E_{10}^c$$

We may calculate directly $P(C) = 0.57 \cdot 0.69 = 0.3933$. Now A is a Boolean combination of $\{E_1, E_3, E_4, E_7\}$ and *B* is a combination of $\{E_2, E_5, E_6E_8\}$. By the result on independence of Boolean combinations, the class $\{A, B, C\}$ is independent. We use the m-procedures to calculate $P(A)$ and $P(B)$. Then we deal with the independent class $\{A, B, C\}$ to obtain the probability of F.

```
>> p  = 0.01*[13 37 12 56 33 71 22 43 57 31];
>> pa = p([1 3 4 7]);     % Selection of probabilities for A
>> pb = p([2 5 6 8]);     % Selection of probabilities for B
>> pma = minprob(pa);     % Minterm probabilities for calculating P(A)
>> pmb = minprob(pb);     % Minterm probabilities for calculating P(B)
>> minvec4;
>> a = A|(B&(C|Dc));      % A corresponds to E1, B to E3, C to E4, D to E7
>> PA = a*pma'
PA = 0.2243
>> b = A&(Bc|(C&D));      % A corresponds to E2, B to E5, C to E6, D to E8
>> PB = b*pmb'
PB = 0.2852
>> PC = p(9)*(1 - p(10))
PC = 0.3933
>> pm = minprob([PA PB PC]);
>> minvec3                % The problem becomes a three variable problem
```

```
>> F = A|B|C;              % with {A,B,C} an independent class
>> PF = F*pm'
PF = 0.6636                % Agrees with the result of Example 4.2.7
```

This page titled 4.2: MATLAB and Independent Classes is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Composite Trials

Composite trials and component events

Often a trial is a composite one. That is, the fundamental trial is completed by performing several steps. In some cases, the steps are carried out sequentially in time. In other situations, the order of performance plays no significant role. Some of the examples in the unit on Conditional Probability involve such multistep trials. We examine more systematically how to model composite trials in terms of events determined by the components of the trials. In the subsequent section, we illustrate this approach in the important special case of Bernoulli trials, in which each outcome results in a success or failure to achieve a specified condition.

We call the individual steps in the composite trial *component trials*. For example, in the experiment of flipping a coin ten times, we refer the ith toss as the ith component trial. In many cases, the component trials will be performed sequentially in time. But we may have an experiment in which ten coins are flipped simultaneously. For purposes of analysis, we impose an ordering— usually by assigning indices. The question is how to model these repetitions. Should they be considered as ten trials of a single simple experiment? It turns out that this is not a useful formulation. We need to consider the composite trial as a single outcome— i.e., represented by a single point in the basic space ω.

Some authors give considerable attention the the nature of the basic space, describing it as a Cartesian product space, with each coordinate corresponding to one of the component outcomes. We find that unnecessary, and often confusing, in setting up the basic model. We simply suppose the basic space has enough elements to consider each possible outcome. For the experiment of flipping a coin ten times, there must be at least $2^{10} = 1024$ elements, one for each possible sequence of heads and tails.

Of more importance is describing the various events associated with the experiment. We begin by identifying the appropriate *component events*. A component event is determined by propositions about the outcomes of the corresponding component trial.

Example 4.3.1 Component events

- In the coin flipping experiment, consider the event H_3 that the third toss results in a head. Each outcome ω of the experiment may be represented by a sequence of H's and T's, representing heads and tails. The event H_3 consists of those outcomes represented by sequences with H in the third position. Suppose A is the event of a head on the third toss and a tail on the ninth toss. This consists of those outcomes corresponding to sequences with H in the third position and T in the ninth. Note that this event is the intersection $H_3H_9^c$.
- A somewhat more complex example is as follows. Suppose there are two boxes, each containing some red and some blue balls. The experiment consists of selecting at random a ball from the first box, placing it in the second box, then making a random selection from the modified contents of the second box. The composite trial is made up of two component selections. We may let R_1 be the event of selecting a red ball on the first component trial (from the first box), and R_2 be the event of selecting a red ball on the second component trial. Clearly R_1 and R_2 are component events.

In the first example, it is reasonable to assume that the class $\{H_i : 1 \leq i \leq 10\}$ is independent, and each component probability is usually taken to be 0.5. In the second case, the assignment of probabilities is somewhat more involved. For one thing, it is necessary to know the numbers of red and blue balls in each box before the composite trial begins. When these are known, the usual assumptions and the properties of conditional probability suffice to assign probabilities. This approach of utilizing component events is used tacitly in some of the examples in the unit on Conditional Probability.

When appropriate component events are determined, various Boolean combinations of these can be expressed as minterm expansions.

Example 4.3.2

Four persons take one shot each at a target. Let E_i be the event the ith shooter hits the target center. Let A_3 be the event exacty three hit the target. Then A_3 is the union of those minterms generated by the E_i which have three places uncomplemented.

$$A_3 = E_1E_2E_3E_4^c \bigvee E_1E_2E_3^cE_4 \bigvee E_1E_2^cE_3E_4 \bigvee E_1^cE_2E_3E_4^c$$

Usually we would be able to assume the E_i form an independent class. If each $P(E_i)$ is known, then all minterm probabilities can be calculated easily.

The following is a somewhat more complicated example of this type.

Example 4.3.3

Ten race cars are involved in time trials to determine pole positions for an upcoming race. To qualify, they must post an average speed of 125 mph or more on a trial run. Let E_i be the event the ith car makes qualifying speed. It seems reasonable to suppose the class $\{E_i : 1 \le i \le 10\}$ is independent. If the respective probabilities for success are 0.90, 0.88, 0.93, 0.77, 0.85, 0.96, 0.72, 0.83, 0.91, 0.84, what is the probability that k or more will qualify ($k = 6, 7, 8, 9, 10$)?

Solution

Let A_k be the event exactly k qualify. The class $\{E_i : 1 \le i \le 10\}$ generates $2^{10} = 1024$ minterms. The event A_k is the union of those minterms which have exactly k places uncomplemented. The event B_k that k or more qualify is given by

$$B_k = \bigvee_{r=k}^{n} A_r$$

The task of computing and adding the minterm probabilities by hand would be tedious, to say the least. However, we may use the function ckn, introduced in the unit on MATLAB and Independent Classes and illustrated in Example 4.4.2, to determine the desired probabilities quickly and easily.

```
>> P = [0.90, 0.88, 0.93, 0.77, 0.85, 0.96,0.72, 0.83, 0.91, 0.84];
>> k = 6:10;
>> PB = ckn(P,k)
PB =   0.9938    0.9628    0.8472    0.5756    0.2114
```

An alternate approach is considered in the treatment of random variables.

Bernoulli trials and the binomial distribution

Many composite trials may be described as a sequence of *success-failure* trials. For each component trial in the sequence, the outcome is one of two kinds. One we designate a *success* and the other a *failure*. Examples abound: heads or tails in a sequence of coin flips, favor or disapprove of a proposition in a survey sample, and items from a production line meet or fail to meet specifications in a sequence of quality control checks. To represent the situation, we let E_i be the event of a success on the ith component trial in the sequence. The event of a failure on the ith component trial is thus E_i^c.

In many cases, we model the sequence as a *Bernoulli sequence*, in which the results on the successive component trials are independent and have the same probabilities. Thus, formally, a sequence of success-failure trials is Bernoulli iff

The class $\{E_i : 1 \le i\}$ is independent.
The probability $P(E_i) = p$, invariant with i.

Simulation of Bernoulli trials

It is frequently desirable to simulate Bernoulli trials. By flipping coins, rolling a die with various numbers of sides (as used in certain games), or using spinners, it is relatively easy to carry this out physically. However, if the number of trials is large—say several hundred—the process may be time consuming. Also, there are limitations on the values of p, the probability of success. We have a convenient two-part m-procedure for simulating Bernoulli sequences. The first part, called *btdata*, sets the parameters. The second, called bt, uses the random number generator in MATLAB to produce a sequence of zeros and ones (for failures and successes). Repeated calls for bt produce new sequences.

Example 4.3.4

```
>> btdata
Enter n, the number of trials  10
Enter p, the probability of success on each trial  0.37
 Call for bt
 >> bt
n = 10   p = 0.37     % n is kept small to save printout space
Frequency = 0.4
To view the sequence, call for SEQ
>> disp(SEQ)          % optional call for the sequence
     1     1
     2     1
     3     0
     4     0
     5     0
     6     0
     7     0
     8     0
     9     1
    10     1
```

Repeated calls for bt yield new sequences with the same parameters.

To illustrate the power of the program, it was used to take a run of 100,000 component trials, with probability p of success 0.37, as above. Successive runs gave relative frequencies 0.37001 and 0.36999. Unless the random number generator is "seeded" to make the same starting point each time, successive runs will give different sequences and usually different relative frequencies.

The binomial distribution

A basic problem in Bernoulli sequences is to determine the probability of k successes in n component trials. We let S_n be the number of successes in n trials. This is a special case of a simple random variable, which we study in more detail in the chapter on "Random Variables and Probabilities".

Let us characterize the events $A_{kn} = \{S_n = k\}$, $0 \le k \le n$. As noted above, the event A_{kn} of exactly k successes is the union of the minterms generated by $\{E_i : 1 \le i\}$ in which there are k successes (represented by k uncomplemented E_i) and $n - k$ failures (represented by $n - k$ complemented E_i^c). Simple combinatorics show there are $C(n, k)$ ways to choose the k places to be uncomplemented. Hence, among the 2^n minterms, there are $C(n, k) = \dfrac{n!}{k!(n-k)!}$ which have k places uncomplemented. Each such minterm has probability $p^k(1-p)^{n-k}$. Since the minterms are mutually exclusive, their probabilities add. We conclude that

$$P(S_n = k) = C(n, k)p^k(1-p)^{n-k} = C(n, k)p^k q^{n-k} \text{ where } q = 1 - p \text{ for } 0 \le k \le n$$

These probabilities and the corresponding values form the *distribution* for S_n. This distribution is known as the *binomial distribution*, with parameters (n, p). We shorten this to binomial (n, p), and often writ S_n ~ binomial (n, p). A related set of probabilities is $P(S_n \ge k) = P(B_{kn})$, $0 \le k \le n$. If the number n of component trials is small, direct computation of the probabilities is easy with hand calculators.

Example 4.3.5 A reliability problem

A remote device has five similar components which fail independently, with equal probabilities. The system remains operable if three or more of the components are operative. Suppose each unit remains active for one year with probability 0.8. What is the probability the system will remain operative for that long?

Solution

$$P = C(5,3)0.8^3 \cdot 0.2^2 + C(5,4)0.8^4 \cdot 0.2 + C(5,5)0.8^5 = 10 \cdot 0.8^3 \cdot 0.2^2 + 5 \cdot 0.8^4 \cdot 0.2 + 0.8^5 = 0.9421$$

Because Bernoulli sequences are used in so many practical situations as models for success-failure trials, the probabilities $P(S_n = k)$ and $P(S_n \geq k)$ have been calculated and tabulated for a variety of combinations of the parameters (n, p). Such tables are found in most mathematical handbooks. Tables of $P(S_n = k)$ are usually given a title such as *binomial distribution, individual terms*. Tables of $P(S_n \geq k)$ have a designation such as *binomial distribution, cumulative terms*. Note, however, some tables for cumulative terms give $P(S_n \leq k)$. Care should be taken to note which convention is used.

Example 4.3.6 A reliability problem

Consider again the system of Example 5, above. Suppose we attempt to enter a table of Cumulative Terms, Binomial Distribution at $n = 5$, $k = 3$, and $p = 0.8$. Most tables will *not* have probabilities greater than 0.5. In this case, we may work with *failures*. We just interchange the role of E_i and E_i^c. Thus, the number of failures has the binomial (n, p) distribution. Now there are three or more successes iff there are *not* three or more failures. We go the the table of cumulative terms at $n = 5$, $k = 3$, and $p = 0.2$. The probability entry is 0.0579. The desired probability is 1 - 0.0579 = 0.9421.

In general, there are k or more successes in n trials iff there are not $n - k + 1$ or more failures.

m-functions for binomial probabilities

Although tables are convenient for calculation, they impose serious limitations on the available parameter values, and when the values are found in a table, they must still be entered into the problem. Fortunately, we have convenient m-functions for these distributions. When MATLAB is available, it is much easier to generate the needed probabilities than to look them up in a table, and the numbers are entered directly into the MATLAB workspace. And we have great freedom in selection of parameter values. For example we may use n of a thousand or more, while tables are usually limited to n of 20, or at most 30. The two m-functions for calculating $P(A_{kn})$ and $P(B_{kn})$ are

$P(A_{kn})$ is calculated by `y = ibinom(n,p,k)`, where k is a row or column vector of integers between 0 and n. The result y is a row vector of the same size as k.
$P(B_{kn})$ is calculated by `y = cbinom(n,p,k)`, where k is a row or column vector of integers between 0 and n. The result y is a row vector of the same size as k.

Example 4.3.7 Use of m-functions ibinom and cbinom

If $n = 10$ and $p = 0.39$, determine $P(A_{kn})$ and $P(B_{kn})$ for $k = 3, 5, 6, 8$.

```
>> p = 0.39;
>> k = [3 5 6 8];
>> Pi = ibinom(10,p,k)  % individual probabilities
Pi = 0.2237    0.1920    0.1023    0.0090
>> Pc = cbinom(10,p,k)  % cumulative probabilities
Pc = 0.8160    0.3420    0.1500    0.0103
```

Note that we have used probability $p = 0.39$. It is quite unlikely that a table will have this probability. Although we use only $n = 10$, frequently it is desirable to use values of several hundred. The m-functions work well for n up to 1000 (and even higher for small values of *p* or for values very near to one). Hence, there is great freedom from the limitations of tables. If a table with a specific range of values is desired, an m-procedure called *binomial* produces such a table. The use of large n raises the question of cumulation of errors in sums or products. The level of precision in MATLAB calculations is sufficient that such roundoff errors are well below pratical concerns.

Example 4.3.8

```
>> binomial                              % call for procedure
Enter n, the number of trials  13
Enter p, the probability of success  0.413
Enter row vector k of success numbers  0:4
    n            p
   13.0000    0.4130
       k      P(X=k)    P(X>=k)
          0    0.0010    1.0000
    1.0000    0.0090    0.9990
    2.0000    0.0379    0.9900
    3.0000    0.0979    0.9521
    4.0000    0.1721    0.8542
```

Remark. While the m-procedure binomial is useful for constructing a table, it is usually not as convenient for problems as the m-functions ibinom or cbinom. The latter calculate the desired values and *put them directly into the* MATLAB *workspace*.

Joint Bernoulli trials

Bernoulli trials may be used to model a variety of practical problems. One such is to compare the results of two sequences of Bernoulli trials carried out independently. The following simple example illustrates the use of MATLAB for this.

Example 4.3.9 A joint Bernoulli trial

Bill and Mary take ten basketball free throws each. We assume the two seqences of trials are independent of each other, and each is a Bernoulli sequence.

Mary: Has probability 0.80 of success on each trial.

Bill: Has probability 0.85 of success on each trial.

What is the probability Mary makes more free throws than Bill?

Solution

We have two Bernoulli sequences, operating independently.

Mary: $n = 10$, $p = 0.80$

Bill: $n = 10$, $p = 0.85$

Let

M be the event Mary wins

M_k be the event Mary makes k or more freethrows.

B_j be the event Bill makes exactly j reethrows

Then Mary wins if Bill makes none and Mary makes one or more, or Bill makes one and Mary makes two or more, etc. Thus

$$M = B_0M_1 \bigvee B_1M_2 \bigvee \cdots \bigvee B_9M_{10}$$

and

$$P(M) = P(B_0)P(M_1) + P(B_1)P(M_2) + \cdots + P(B_9)P(M_{10})$$

We use cbinom to calculate the cumulative probabilities for Mary and ibinom to obtain the individual probabilities for Bill.

```
>> pm = cbinom(10,0.8,1:10);     % cumulative probabilities for Mary
>> pb = ibinom(10,0.85,0:9);     % individual probabilities for Bill
>> D = [pm; pb]'                 % display: pm in the first column
   D =                           % pb in the second column
    1.0000    0.0000
    1.0000    0.0000
    0.9999    0.0000
    0.9991    0.0001
    0.9936    0.0012
    0.9672    0.0085
    0.8791    0.0401
    0.6778    0.1298
    0.3758    0.2759
    0.1074    0.3474
```

To find the probability $P(M)$ that Mary wins, we need to multiply each of these pairs together, then sum. This is just the dot or scalar product, which MATLAB calculates with the command $pm * pb'$. We may combine the generation of the probabilities and the multiplication in one command:

```
>> P = cbinom(10,0.8,1:10)*ibinom(10,0.85,0:9)'
   P = 0.273
```

The ease and simplicity of calculation with MATLAB make it feasible to consider the effect of different values of *n*. Is there an optimum number of throws for Mary? Why should there be an optimum?

An alternate treatment of this problem in the unit on Independent Random Variables utilizes techniques for independent simple random variables.

Alternate MATLAB implementations

Alternate implementations of the functions for probability calculations are found in the *Statistical Package* available as a supplementary package. We have utilized our formulation, so that only the basic MATLAB package is needed.

This page titled 4.3: Composite Trials is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Problems on Independence of Events

Exercise 4.4.1

The minterms generated by the class $\{A, B, C\}$ have minterm probabilities

$$pm = [0.15\ 0.05\ 0.02\ 0.18\ 0.25\ 0.05\ 0.18\ 0.12]$$

Show that the product rule holds for all three, but the class is not independent.

Answer available at the end of this section.

Exercise 4.4.2

The class $\{A, B, C, D\}$is independent, with respective probabilities 0.65, 0.37, 0.48, 0.63. Use the m-function minprob to obtain the minterm probabilities. Use the m-function minmap to put them in a 4 by 4 table corresponding to the minterm map convention we use.

Answer available at the end of this section.

Exercise 4.4.3

The minterm probabilities for the software survey in Example 2 from "Minterms" are

$$pm = [0\ 0.05\ 0.10\ 0.05\ 0.20\ 0.10\ 0.40\ 0.10]$$

Show whether or not the class $\{A, B, C\}$ is independent: (1) by hand calculation, and (2) by use of the m-function imintest.

Answer available at the end of this section.

Exercise 4.4.4

The minterm probabilities for the computer survey in Example 3 from "Minterms" are

$$pm = [0.032\ 0.016\ 0.376\ 0.011\ 0.364\ 0.073\ 0.077\ 0.051]$$

Show whether or not the class $\{A, B, C\}$ is independent: (1) by hand calculation, and (2) by use of the m-function imintest.

Answer available at the end of this section.

Exercise 4.4.5

Minterm probabilities $p(0)$ through $p(15)$ for the class $\{A, B, C, D\}$ are, in order,

$$pm = [0.084\ 0.196\ 0.036\ 0.084\ 0.085\ 0.196\ 0.035\ 0.084\ 0.021\ 0.049\ 0.009\ 0.021\ 0.020\ 0.049\ 0.010\ 0.021]$$

Use the m-function imintest to show whether or not the class $\{A, B, C, D\}$ is independent.

Answer available at the end of this section.

Exercise 4.4.6

Minterm probabilities $p(0)$ through $p(15)$ for the opinion survey in Example 4 from "Minterms" are

$$pm = [0.085\ 0.195\ 0.035\ 0.085\ 0.080\ 0.200\ 0.035\ 0.085\ 0.020\ 0.050\ 0.010\ 0.020\ 0.020\ 0.050\ 0.015\ 0.015]$$

show whether or not the class $\{A, B, C, D\}$ is independent.

Answer available at the end of this section.

Exercise 4.4.7

The class $\{A, B, C\}$ is independent, with $P(A) = 0.30$, $P(B^cC) = 0.32$, and $P(AC) = 0.12$. Determine the minterm probabilities.

Answer available at the end of this section.

Exercise 4.4.8

The class $\{A, B, C\}$ is independent, with $P(A \cup B) = 0.6$, $P(A \cup C) = 0.7$, and $P(C) = 0.4$. Determine the probability of each minterm.

Answer available at the end of this section.

Exercise 4.4.9

A pair of dice is rolled five times. What is the probability the first two results are “sevens” and the others are not?

Answer available at the end of this section.

Exercise 4.4.10

David, Mary, Joan, Hal, Sharon, and Wayne take an exam in their probability course. Their probabilities of making 90 percent or more are

0.72 0.83 0.75 0.92 0.65 0.79

respectively. Assume these are independent events. What is the probability three or more, four or more, five or more make grades of at least 90 percent?

Answer available at the end of this section.

Exercise 4.4.11

Two independent random numbers between 0 and 1 are selected (say by a random number generator on a calculator). What is the probability the first is no greater than 0.33 and the other is at least 57?

Answer available at the end of this section.

Exercise 4.4.12

Helen is wondering how to plan for the weekend. She will get a letter from home (with money) with probability 0.05. There is a probability of 0.85 that she will get a call from Jim at SMU in Dallas. There is also a probability of 0.5 that William will ask for a date. What is the probability she will get money and Jim will not call or that both Jim will call and William will ask for a date?

Answer available at the end of this section.

Exercise 4.4.13

A basketball player takes ten free throws in a contest. On her first shot she is nervous and has probability 0.3 of making the shot. She begins to settle down and probabilities on the next seven shots are 0.5, 0.6 0.7 0.8 0.8, 0.8 and 0.85, respectively. Then she realizes her opponent is doing well, and becomes tense as she takes the last two shots, with probabilities reduced to 0.75, 0.65. Assuming independence between the shots, what is the probability she will make k or more for $k = 2, 3, \cdots 10$?

Answer available at the end of this section.

Exercise 4.4.14

In a group there are M men and W women; *m* of the men and w of the women are college graduates. An individual is picked at random. Let A be the event the individual is a woman and B be the event he or she is a college graduate. Under what condition is the pair $\{A, B\}$ independent?

Answer available at the end of this section.

Exercise 4.4.15

Consider the pair $\{A, B\}$ of events. Let $P(A) = p$, $P(A^c) = q = 1 - p$, $P(B|A) = p_1$, and $P(B|A^c) = p_2$. Under what condition is the pair $\{A, B\}$ independent?

Answer available at the end of this section.

Exercise 4.4.16

Show that if event A is independent of itself, then $P(A) = 0$ or $P(A) = 1$. (This fact is key to an important "zero-one law".)

Answer available at the end of this section.

Exercise 4.4.17

Does $\{A, B\}$ independent and $\{B, C\}$ independent imply $\{A, C\}$ is independent? Justify your answer.

Answer available at the end of this section.

Exercise 4.4.18

Suppose event A implies B (i.e. $A \subset B$). Show that if the pair $\{A, B\}$ is independent, then either $P(A) = 0$ or $P(B) = 1$.

Answer available at the end of this section.

Exercise 4.4.19

A company has three task forces trying to meet a deadline for a new device. The groups work independently, with respective probabilities 0.8, 0.9, 0.75 of completing on time. What is the probability at least one group completes on time? (Think. Then solve “by hand.”)

Answer available at the end of this section.

Exercise 4.4.20

Two salesmen work differently. Roland spends more time with his customers than does Betty, hence tends to see fewer customers. On a given day Roland sees five customers and Betty sees six. The customers make decisions independently. If the probabilities for success on Roland's customers are 0.7, 0.8, 0.8, 0.6, 0.7 and for Betty's customers are 0.6, 0.5, 0.4, 0.6, 0.6, 0.4, what is the probability Roland makes more sales than Betty? What is the probability that Roland will make three or more sales? What is the probability that Betty will make three or more sales?

Answer available at the end of this section.

Exercise 4.4.21

Two teams of students take a probability exam. The entire group performs individually and independently. Team 1 has five members and Team 2 has six members. They have the following indivudal probabilities of making an `”A” on the exam.

Team 1: 0.83 0.87 0.92 0.77 0.86 Team 2: 0.68 0.91 0.74 0.68 0.73 0.83

a. What is the probability team 1 will make *at least as many* A's as team 2?
b. What is the probability team 1 will make *more* A's than team 2?

Answer

```
P1 = 0.01*[83 87 92 77 86];
P2 = 0.01*[68 91 74 68 73 83];
P1geq = ikn(P2,0:5)*ckn(P1,0:5)'
P1geq =  0.5527
P1g = ikn(P2,0:4)*ckn(P1,1:5)'
P1g =    0.2561
```

Exercise 4.4.22

A system has five components which fail independently. Their respective reliabilities are 0.93, 0.91, 0.78, 0.88, 0.92. Units 1 and 2 operate as a “series” combination. Units 3, 4, 5 operate as a two of three subsytem. The two subsystems operate as a parallel combination to make the complete system. What is reliability of the complete system?

Answer available at the end of this section.

Exercise 4.4.23

A system has eight components with respective probabilities

0.96 0.90 0.93 0.82 0.85 0.97 0.88 0.80

Units 1 and 2 form a parallel subsytem in series with unit 3 and a three of five combination of units 4 through 8. What is the reliability of the complete system?

Answer

```
R = 0.01*[96 90 93 82 85 97 88 80];
Ra = parallel(R(1:2))
Ra =  0.9960
Rb = ckn(R(4:8),3)
Rb =  0.9821
Rs = prod([Ra R(3) Rb])
Rs =  0.9097
```

Exercise 4.4.24

How would the reliability of the system in Exercise 4.4.23. change if units 1, 2, and 3 formed a parallel combination in series with the three of five combination?

Answer available at the end of this section.

Exercise 4.4.25

How would the reliability of the system in Exercise 4.4.23. change if the reliability of unit 3 were changed from 0.93 to 0.96? What change if the reliability of unit 2 were changed from 0.90 to 0.95 (with unit 3 unchanged)?

Answer available at the end of this section.

Exercise 4.4.26

Three fair dice are rolled. What is the probability at least one will show a six?

Answer available at the end of this section.

Exercise 4.4.27

A hobby shop finds that 35 percent of its customers buy an electronic game. If customers buy independently, what is the probability that at least one of the next five customers will buy an electronic game?

Answer available at the end of this section.

Exercise 4.4.28

Under extreme noise conditions, the probability that a certain message will be transmitted correctly is 0.1. Successive messages are acted upon independently by the noise. Suppose the message is transmitted ten times. What is the probability it is transmitted correctly at least once?

Answer available at the end of this section.

Exercise 4.4.29

Suppose the class $\{A_i : 1 \leq i \leq n\}$ is independent, with $P(A_i) = p_i$, $1 \leq i \leq n$. What is the probability that at least one of the events occurs? What is the probability that none occurs?

Answer available at the end of this section.

Exercise 4.4.30

In one hundred random digits, 0 through 9, with each possible digit equally likely on each choice, what is the probility 8 or more are sevens?

Answer available at the end of this section.

Exercise 4.4.31

Ten customers come into a store. If the probability is 0.15 that each customer will buy a television set, what is the probability the store will sell three or more?

Answer available at the end of this section.

Exercise 4.4.32

Seven similar units are put into service at time $t = 0$. The units fail independently. The probability of failure of any unit in the first 400 hours is 0.18. What is the probability that three or more units are still in operation at the end of 400 hours?

Answer available at the end of this section.

Exercise 4.4.33

A computer system has ten similar modules. The circuit has redundancy which ensures the system operates if any eight or more of the units are operative. Units fail independently, and the probability is 0.93 that any unit will survive between maintenance periods. What is the probability of no system failure due to these units?

Answer available at the end of this section.

Exercise 4.4.34

Only thirty percent of the items from a production line meet stringent requirements for a special job. Units from the line are tested in succession. Under the usual assumptions for Bernoulli trials, what is the probability that three satisfactory units will be found in eight or fewer trials?

Answer available at the end of this section.

Exercise 4.4.35

The probability is 0.02 that a virus will survive application of a certain vaccine. What is the probability that in a batch of 500 viruses, fifteen or more will survive treatment?

Answer available at the end of this section.

Exercise 4.4.36

In a shipment of 20,000 items, 400 are defective. These are scattered randomly throughout the entire lot. Assume the probability of a defective is the same on each choice. What is the probability that

1. Two or more will appear in a random sample of 35?
2. At most five will appear in a random sample of 50?

Answer

$P1$ = **cbinom**(35, 0.02, 2) = 0.1547.

$P2$ = 1 – **cbinom**(35, 0.02, 6) = 0.9999

Exercise 4.4.37

A device has probability p of operating successfully on any trial in a sequence. What probability p is necessary to ensure the probability of successes on all of the first four trials is 0.85? With that value of p, what is the probability of four or more successes in five trials?

Answer available at the end of this section.

Exercise 4.4.38

A survey form is sent to 100 persons. If they decide independently whether or not to reply, and each has probability 1/4 of replying, what is the probability of k or more replies, where $k = 15, 20, 25, 30, 35, 40$?

Answer available at the end of this section.

Exercise 4.4.39

Ten numbers are produced by a random number generator. What is the probability four or more are less than or equal to 0.63?

Answer available at the end of this section.

Exercise 4.4.40

A player rolls a pair of dice five times. She scores a "hit" on any throw if she gets a 6 or 7. She wins iff she scores an *odd* number of hits in the five throws. What is the probability a player wins on any sequence of five throws? Suppose she plays the game 20 successive times. What is the probability she wins at least 10 times? What is the probability she wins more than half the time?

Answer available at the end of this section.

Exercise 4.4.41

Erica and John spin a wheel which turns up the integers 0 through 9 with equal probability. Results on various trials are independent. Each spins the wheel 10 times. What is the probability Erica turns up a seven more times than does John?

Answer available at the end of this section.

Exercise 4.4.42

Erica and John play a different game with the wheel, above. Erica scores a point each time she gets an integer 0, 2, 4, 6, or 8. John scores a point each time he turns up a 1, 2, 5, or 7. If Erica spins eight times; John spins 10 times. What is the probability John makes *more* points than Erica?

Answer available at the end of this section.

Exercise 4.4.43

A box contains 100 balls; 30 are red, 40 are blue, and 30 are green. Martha and Alex select at random, with replacement and mixing after each selection. Alex has a success if he selects a red ball; Martha has a success if she selects a blue ball. Alex selects seven times and Martha selects five times. What is the probability Martha has more successes than Alex?

Answer available at the end of this section.

Exercise 4.4.44

Two players roll a fair die 30 times each. What is the probability that each rolls the same number of sixes?

Answer available at the end of this section.

Exercise 4.4.45

A warehouse has a stock of n items of a certain kind, r of which are defective. Two of the items are chosen at random, without replacement. What is the probability that at least one is defective? Show that for large n the number is very close to that for selection with replacement, which corresponds to two Bernoulli trials with pobability $p = r/n$ of success on any trial.

Answer available at the end of this section.

Exercise 4.4.46

A coin is flipped repeatedly, until a head appears. Show that with probability one the game will terminate.

tip:

The probability of not terminating in n trials is q^n.

Answer available at the end of this section.

Exercise 4.4.47

Two persons play a game consecutively until one of them is successful or there are ten unsuccesful plays. Let E_i be the event of a success on the ith play of the game. Suppose $\{E_i : 1 \le i\}$ is an independent class with $P(E_i) = p_1$ for *i* odd and $P(E_i) = p_2$ for i even. Let A be the event the first player wins, B be the event the second player wins, and C be the event that neither wins.

a. Express A, B, and C in terms of the E_i.
b. Determine $P(A)$, $P(B)$, and $P(C)$ in terms of p_1, p_2, $q_1 = 1 - p_1$, and $q_2 = 1 - p_2$. Obtain numerical values for the case $p_1 = 1/4$ and $p_2 = 1/3$.
c. Use appropriate facts about the geometric series to show that $P(A) = P(B)$ iff $p_1 = p_2/(1 + p_2)$.
d. Suppose $p_2 = 0.5$. Use the result of part (c) to find the value of p_1 to make $P(A) = P(B)$ and then determine $P(A)$, $P(B)$, and $P(C)$.

Answer

a. $C = \bigcap_{i=1}^{10} E_i^c$.

$$A = E_1 \bigvee E_1^c E_2^c E_3 \bigvee E_1^c E_2^c E_3^c E_4^c E_5 \bigvee E_1^c E_2^c E_3^c E_4^c E_5^c E_6^c E_7 \bigvee E_1^c E_2^c E_3^c E_4^c E_5^c E_6^c E_7^c E_8^c E_9$$

$$B = E_1^c E_2 \bigvee E_1^c E_2^c E_3^c E_4 \bigvee E_1^c E_2^c E_3^c E_4^c E_5^c E_6 \bigvee E_1^c E_2^c E_3^c E_4^c E_5^c E_6^c E_7^c E_8 \bigvee E_1^c E_2^c E_3^c E_4^c E_5^c E_6^c E_7^c E_8^c E_9^c E_{10}$$

$$P(A) = p_1[1 + q_1q_2 + (q_1q_2)^2 + (q_1q_2)^3 + (q_1q_2)^4] = p1\frac{1 - (q_1q_2)^5}{1 - q_1q_2}$$

$$P(B) = q_1p_2\frac{1 - (q_1q_2)^5}{1 - q_1q_2} P(C) = (q_1q_2)^5$$

For $p_1 = 1/4$, $p_2 = 1/3$, we have $q_1q_2 = 1/2$ and $q_1p_2 = 1/4$. In this case

$$P(A) = \frac{1}{4} \cdot \frac{31}{16} = 31/64 = 0.4844 = P(B), P(C) = 1/32$$

Note that $P(A) + P(B) + P(C) = 1$.

c. $P(A) = P(B)$ iff $p_1 = q_1p_2 = (1 - p_1)p_2$ iff $p_1 = p_2/(1 + p_2)$.

d. $p_1 = 0.5/1.5 = 1/3$

Exercise 4.4.48

Three persons play a game consecutively until one achieves his objective. Let E_i be the event of a success on the ith trial, and suppose $\{E_i : 1 \le i\}$ is an independent class, with $P(E_i) = p_1$ for $i = 1, 4, 7, \cdots, P(E_i) = p_2$ for $i = 2, 5, 8, \cdots$, and $P(E_i) = p_3$ for $i = 3, 6, 9, \cdots$. Let A, B, C be the respective events the first, second, and third player wins.

a. Express A, B, and C in terms of the E_i.

b. Determine the probabilities in terms of p_1, p_2, p_3, then obtain numerical values in the case $p_1 = 1/4$, $p_2 = 1/3$, and $p_3 = 1/2$.

Answer available at the end of this section.

Exercise 4.4.49

What is the probability of a success on the ith trial in a Bernoulli sequence of n component trials, given there are r successes?

Answer available at the end of this section.

Exercise 4.4.50

A device has N similar components which may fail independently, with probability p of failure of any component. The device fails if one or more of the components fails. In the event of failure of the device, the components are tested sequentially.

a. What is the probability the first defective unit tested is the nth, given one or more components have failed?
b. What is the probability the defective unit is the nth, given that exactly one has failed?
c. What is the probability that more than one unit has failed, given that the first defective unit is the nth?

Answer

Let A_1 = event one failure, B_1 = event of one or more failures, B_2 = event of two or more failures, and F_n = the event the first defective unit found is the nth.

a. $F_n \subset B_1$ implies $P(F_n|B_1) = P(F_n)/P(B_1) = \dfrac{q^{n-1}p}{1-q^N}$

$$P(F_n|A_1) = \frac{P(F_nA_1}{P(A_1)} = \frac{q^{n-1}pq^{N-n}}{Npq^{N-1}} = \frac{1}{N}$$

(see Exercise)

b. Since probability not all from nth are good is $1 - q^{N-n}$.

$$P(B_2|F_n) = \frac{P(B_2F_n}{P(F_n)} = \frac{q^{n-1}p(1-Q^{N-1}}{q^{n-1}p} = 1 - q^{N-n}$$

Answers

Exercise 4.4.1

```
pm = [0.15 0.05 0.02 0.18 0.25 0.05 0.18 0.12];
y = imintest(pm)
The class is NOT independent
Minterms for which the product rule fails
y =
     1     1     1     0
     1     1     1     0   % The product rule hold for M7 = ABC
```

Exercise 4.4.2

```
P = [0.65 0.37 0.48 0.63];
p = minmap(minprob(P))
p =
    0.0424    0.0249    0.0788    0.0463
    0.0722    0.0424    0.1342    0.0788
    0.0392    0.0230    0.0727    0.0427
    0.0667    0.0392    0.1238    0.0727
```

Exercise 4.4.3

```
pm = [0 0.05 0.10 0.05 0.20 0.10 0.40 0.10];
y = imintest(pm)
The class is NOT independent
Minterms for which the product rule fails
y =
     1     1     1     1    % By hand check product rule for any minterm
     1     1     1     1
```

Exercise 4.4.4

```
npr04_04
Minterm probabilities for Exercise 4.4.4. are in pm
y = imintest(pm)
The class is NOT independent
Minterms for which the product rule fails
y =
     1     1     1     1
     1     1     1     1
```

Exercise 4.4.5

```
npr04_05
Minterm probabilities for Exercise 4.4.5. are in pm
imintest(pm)
The class is NOT independent
Minterms for which the product rule fails
ans =
     0     1     0     1
     0     0     0     0
     0     1     0     1
     0     0     0     0
```

Exercise 4.4.6

```
npr04_06
Minterm probabilities for Exercise 4.4.6. are in pm
y = imintest(pm)
The class is NOT independent
Minterms for which the product rule fails
y =
     1     1     1     1
     1     1     1     1
     1     1     1     1
     1     1     1     1
```

Exercise 4.4.7

$P(C) = P(AC)/P(A) = 0.40$ AND $P(B) = 1 - P(B^cC)/P(C) = 0.20$.

```
pm = minprob([0.3 0.2 0.4])
pm =  0.3360  0.2240  0.0840  0.0560  0.1440  0.0960  0.0360  0.0240
```

Exercise 4.4.8

$P(A^cC^c) = P(A^c)P(C^c) = 0.3$ implies $P(A^c) = 0.3/0.6 = 0.5 = P(A)$.
$P(A^cB^c) = P(A^c)P(B^c) = 0.4$ implies $P(B^c) = 0.4/0.5 = 0.8$ implies $P(B) = 0.2$

```
P = [0.5 0.2 0.4];
pm = minprob(P)
pm =  0.2400  0.1600  0.0600  0.0400  0.2400  0.1600  0.0600  0.0400
```

Exercise 4.4.9

$P = (1/6)^2(5/6)^3 = 0.0161$.

Exercise 4.4.10

```
P = 0.01*[72 83 75 92 65 79];
y = ckn(P,[3 4 5])
y =   0.9780    0.8756    0.5967
```

Exercise 4.4.11

$P = 0.33 \cdot (1 - 0.57) = 0.1419$

Exercise 4.4.12

A ~ letter with money, B ~ call from Jim, C ~ William ask for date

```
P = 0.01*[5 85 50];
minvec3
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
pm = minprob(P);
p = ((A&Bc)|(B&C))*pm'
p =  0.4325
```

Exercise 4.4.13

```
P = 0.01*[30 50 60 70 80 80 80 85 75 65];
k = 2:10;
p = ckn(P,k)
p =
  Columns 1 through 7
    0.9999    0.9984    0.9882    0.9441    0.8192    0.5859    0.3043
  Columns 8 through 9
    0.0966    0.0134
```

Exercise 4.4.14

$P(A|B) = w/(m+w) = W/(W+M) = P(A)$

Exercise 4.4.15

$p_1 = P(B|A) = P(B|A^c) = p_2$ (see table of equivalent conditions).

Exercise 4.4.16

$P(A) = P(A \cap A) = P(A)P(A)$. $x^2 = x$ iff $x = 0$ or $x = 1$.

Exercise 4.4.17

```
% No. Consider for example the following minterm probabilities:
pm = [0.2 0.05 0.125 0.125 0.05 0.2 0.125 0.125];
minvec3
Variables are A, B, C, Ac, Bc, Cc
They may be renamed, if desired.
PA = A*pm'
PA =  0.5000
PB = B*pm'
PB =  0.5000
PC = C*pm'
PC =  0.5000
PAB = (A&B)*pm'  % Product rule holds
PAB =  0.2500
PBC = (B&C)*pm' % Product rule holds
PBC =  0.2500
PAC = (A&C)*pm'  % Product rule fails
PAC =  0.3250
```

Exercise 4.4.18

$A \subset B$ implies $P(AB) = P(A)$; independence implies $P(AB) = P(A)P(B)$. $P(A) = P(A)P(B)$ only if $P(B) = 1$ or $P(A) = 0$.

Exercise 4.4.19

At least one completes iff not all fail. $P = 1 - 0.2 \cdot 0.1 \cdot 0.25 = 0.9950$

Exercise 4.4.20

```
PR = 0.1*[7 8 8 6 7];
PB = 0.1*[6 5 4 6 6 4];
PR3 = ckn(PR,3)
PR3 =  0.8662
PB3 = ckn(PB,3)
PB3 =  0.6906
PRgB = ikn(PB,0:4)*ckn(PR,1:5)'
```

```
PRgB = 0.5065
```

Exercise 4.4.22

```
R = 0.01*[93 91 78 88 92];
Ra = prod(R(1:2))
Ra =  0.8463
Rb = ckn(R(3:5),2)
Rb =  0.9506
Rs = parallel([Ra Rb])
Rs =  0.9924
```

Exercise 4.4.24

```
Rc = parallel(R(1:3))
Rc =  0.9997
Rss = prod([Rb Rc])
Rss = 0.9818
```

Exercise 4.4.25

```
R1 = R;
R1(3) =0.96;
Ra = parallel(R1(1:2))
Ra =  0.9960
Rb = ckn(R1(4:8),3)
Rb =  0.9821
Rs3 = prod([Ra R1(3) Rb])
Rs3 = 0.9390
R2 = R;
R2(2) = 0.95;
Ra = parallel(R2(1:2))
Ra =  0.9980
Rb = ckn(R2(4:8),3)
Rb =  0.9821
Rs4 = prod([Ra R2(3) Rb])
Rs4 = 0.9115
```

Exercise 4.4.26

$P = 1 - (5/6)^3 = 0.4213$

Exercise 4.4.27

$P = 1 - 0.65^5 = 0.8840$

Exercise 4.4.28

$P = 1 - 0.9^{10} = 0.6513$

Exercise 4.4.29

$P1 = 1 - P0$, $P0 = \prod_{i=1}^{n}(1 - p_i)$

Exercise 4.4.30

P = **cbinom**(100, 0.1, 8) = 0.7939

Exercise 4.4.31

P = **cbinom**(10, 0.15, 3) = 0.1798

Exercise 4.4.32

P = **cbinom**(7, 0.82, 3) = 0.9971

Exercise 4.4.33

P = **cbinom**(10,0.93,8) = 0.9717

Exercise 4.4.34

P = **cbinom**(8, 0.3, 3) = 0.4482

Exercise 4.4.35

P = **cbinom**(500, 0.02, 15) = 0.0814

Exercise 4.4.37

$p = 0.85^{1/4}\backslash 0, \backslash (P$ **cbinom**(5, p, 4) = 0.9854.

Exercise 4.4.38

```
P = cbinom(100,1/4,15:5:40)
P =  0.9946    0.9005    0.5383    0.1495    0.0164    0.0007
```

Exercise 4.4.39

$P1$ = **cbinom**(10, 0.63, 4) = 0.9644

Exercise 4.4.40

Each roll yields a hit with probability $p = \frac{6}{36} + \frac{5}{36} = \frac{11}{36}$.

```
PW = sum(ibinom(5,11/36,[1 3 5]))
PW =  0.4956
P2 = cbinom(20,PW,10)
P2 =  0.5724
P3 = cbinom(20,PW,11)
P3 =  0.3963
```

Exercise 4.4.41

P = **ibinom**(10, 0.1, 0:9) * **cbinom**(10, 0.1, 1:10)' = 0.3437

Exercise 4.4.42

P = **ibinom**(8, 0.5, 0:8) * **cbinom**(10, 0.4, 1:9)' = 0.4030

Exercise 4.4.43

P = **ibinom**(7, 0.3, 0:4) * **cbinom**(5, 0.4, 1:5)' = 0.3613

Exercise 4.4.44

P = sum(**ibinom**(30, 1/6, 0:30).^2) = 0.1386

Exercise 4.4.45

$$P1 = \frac{r}{n} \cdot \frac{r-1}{n-1} + \frac{r}{n} \cdot \frac{n-r}{n-1} + \frac{n-r}{n} \cdot \frac{r}{n-1} = \frac{(2n-1)r - r^2}{n(n-1)}$$

$$P2 = 1 - (\frac{r}{n})^2 = \frac{2nr - r^2}{n^2}$$

Exercise 4.4.46

Let N = event never terminates and N_k = event does not terminate in k plays. Then $N \subset N_k$ for all k implies $0 \le P(N) \le P(N_k) = 1/2^k$ for all k, we conclude $P(N) = 0$.

Exercise 4.4.48

a. $A = E_1 \bigvee \bigvee_{k=1}^{\infty} \bigcap_{i=1}^{3k} E_i^c E_{3k+1}$

$B = E_1^c E_2 \bigvee \bigvee_{k=1}^{\infty} \bigcap_{i=1}^{3k+1} E_i^c E_{3k+2}$

$C = E_1^c E_2^c E_3 \bigvee \bigvee_{k=1}^{\infty} \bigcap_{i=1}^{3k+2} E_i^c E_{3k+3}$

b. $P(A) = p_1 \sum_{k=0}^{\infty} (q_1 q_2 q_3)^k = \frac{p_1}{1 - q_1 q_2 q_3}$

$P(B) = \frac{q_1 p_2}{1 - q_1 q_2 q_3}$

$P(C) = \frac{q_1 q_2 p_3}{1 - q_1 q_2 q_3}$

For $p_1 = 1/4$, $p_2 = 1/3$. $p_3 = 1/2$, $P(A) = P(B) = P(C) = 1/3$.

Exercise 4.4.49

$P(A_{rn} E_i = pC(n-1, r-1) p^{r-1} q^{n-r}$ and $P(A_{rn}) = C(n, r) p^r q^{n-r}$.

Hence $P(E_i | A_A rn) = C(n-1, r-1)/C(n, r) = r/n$.

This page titled 4.4: Problems on Independence of Events is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

5: CONDITIONAL INDEPENDENCE

5.1: Conditional Independence
5.2: Patterns of Probable Inference
5.3: Problems on Conditional Independence

This page titled 5: Conditional Independence is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Conditional Independence

5.1. Conditional Independence*

The idea of stochastic (probabilistic) independence is explored in the unit Independence of Events. The concept is approached as lack of conditioning: $P(A|B) = P(A)$. This is equivalent to the product rule $P(AB) = P(A)P(B)$. We consider an extension to conditional independence.

The concept

Examination of the independence concept reveals two important mathematical facts:

- Independence of a class of non mutually exclusive events depends upon the probability measure, and not on the relationship between the events. Independence cannot be displayed on a Venn diagram, unless probabilities are indicated. For one probability measure a pair may be independent while for another probability measure the pair may not be independent.
- Conditional probability is a probability measure, since it has the three defining properties and all those properties derived therefrom.

This raises the question: is there a useful conditional independence—i.e., independence with respect to a conditional probability measure? In this chapter we explore that question in a fruitful way.

Among the simple examples of “operational independence" in the unit on independence of events, which lead naturally to an assumption of “probabilistic independence” are the following:

- If customers come into a well stocked shop at different times, each unaware of the choice made by the other, the the item purchased by one should not be affected by the choice made by the other.
- If two students are taking exams in different courses, the grade one makes should not affect the grade made by the other.

Example 5.1.1 Buying umbrellas and the weather

A department store has a nice stock of umbrellas. Two customers come into the store “independently.” Let *A* be the event the first buys an umbrella and *B* the event the second buys an umbrella. Normally, we should think the events $\{A, B\}$ form an independent pair. But consider the effect of weather on the purchases. Let *C* be the event the weather is rainy (i.e., is raining or threatening to rain). Now we should think $P(A|C) > P(A|C^c)$ and $P(B|C) > P(B|C^c)$. The weather has a decided effect on the likelihood of buying an umbrella. But given the fact the weather is rainy (event *C* has occurred), it would seem reasonable that purchase of an umbrella by one should not affect the likelihood of such a purchase by the other. Thus, it may be reasonable to suppose

$$P(A|C) = P(A|BC) \text{ or, in another notation, } P_C(A) = P_C(A|B)$$

An examination of the sixteen equivalent conditions for independence, with probability measure P replaced by probability measure P_C, shows that we have independence of the pair $\{A, B\}$ with respect to the conditional probability measure $P_C(\cdot) = P(\cdot|C)$. Thus, $P(A|C^c) = P(A|BC^c)$. For this example, we should also expect that $P(A|C^c = P(A|BC^c)$, so that there is independence with respect to the conditional probability measure $P(\cdot|C^c)$. Does this make the pair $\{A, B\}$ independent (with respect to the prior probability measure P)? Some numerical examples make it plain that only in the most unusual cases would the pair be independent. Without calculations, we can see why this should be so. If the first customer buys an umbrella, this indicates a higher than normal likelihood that the weather is rainy, in which case the second customer is likely to buy. The condition leads to $P(B|A) > P(B)$. Consider the following numerical case. Suppose $P(AB|C) = P(A|C)P(B|C)$ and $P(AB|C^c) = P(A|C^c)P(B|C^c)$ and

$$P(A|C) = 0.60,\ P(A|C^c) = 0.20,\ P(B|C) = 0.50,\ P(B|C^c) = 0.15, \text{ with } P(C) = 0.30.$$

Then

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) = 0.3200 \; P(B) = P(B|C)P(C) + P(B|C^c)P(C^c) = 0.2550$$

$$P(AB) = P(AB|C)P(C) + P(AB|C^c)P(C^c) = P(A|C)P(B|C)P(C) + P(A|C^c)P(C^c) = 0.1110$$

As a result,

$$P(A)P(B) = 0.0816 \neq 0.1110 = P(AB)$$

The product rule fails, so that the pair is *not* independent. An examination of the pattern of computation shows that independence would require very special probabilities which are not likely to be encountered.

Example 5.1.2 Students and exams

Two students take exams in different courses, Under normal circumstances, one would suppose their performances form an independent pair. Let *A* be the event the first student makes grade 80 or better and *B* be the event the second has a grade of 80 or better. The exam is given on Monday morning. It is the fall semester. There is a probability 0.30 that there was a football game on Saturday, and both students are enthusiastic fans. Let *C* be the event of a game on the previous Saturday. Now it is reasonable to suppose

$$P(A|C) = P(A|BC) \text{ and } P(A|C^c) = P(A|BC^c)$$

If we know that there was a Saturday game, additional knowledge that *B* has occurred does not affect the lielihood that *A* occurs. Again, use of equivalent conditions shows that the situation may be expressed

$$P(AB|C) = P(A|C)P(B|C) \text{ and } P(AB|C^c) = P(A|C^c)P(B|C^c)$$

Under these conditions, we should suppose that $P(A|C) < P(A|C^c)$ and $P(B|C) < P(B|C^c)$. If we knew that one did poorly on the exam, this would increase the likelihoood there was a Saturday game and hence increase the likelihood that the other did poorly. The failure to be independent arises from a common chance factor that affects both. Although their performances are "operationally" independent, they are not independent in the probability sense. As a numerical example, suppose

$$P(A|C) = 0.7 \; P(A|C^c) = 0.9 \; P(B|C) = 0.6 \; P(B|C^c) = 0.8 \; P(C) = 0.3$$

Straightforward calculations show $P(A) = 0.8400$, $P(B) = 0.7400$, $P(AB) = 0.6300$. Note that $P(A|B) = 0.8514 > P(A)$ as would be expected.

Sixteen equivalent conditions

Using the facts on repeated conditioning and the equivalent conditions for independence, we may produce a similar table of equivalent conditions for conditional independence. In the hybrid notation we use for repeated conditioning, we write

$$P_C(A|B) = P_C(A) \text{ or } P_C(AB) = P_C(A)P_C(B)$$

This translates into

$$P(A|BC) = P(A|C) \text{ or } P(AB|C) = P(A|C)P(B|C)$$

If it is known that C has occurred, then additional knowledge of the occurrence of B does not change the likelihood of A.

If we write the sixteen equivalent conditions for independence in terms of the conditional probability measure $P_C(\cdot)$, then translate as above, we have the following equivalent conditions.

Table 5.1. Sixteen equivalent conditions

$P(A\|BC) = P(A\|C)$	$P(B\|AC) = P(B\|C)$	$P(AB\|C) = P(A\|C)P(B\|C)$
$P(A\|B^cC) = P(A\|C)$	$P(B^c\|AC) = P(B^c\|C)$	$P(AB^c\|C) = P(A\|C)P(B^c\|C)$
$P(A^c\|BC) = P(A^c\|C)$	$P(B\|A^cC) = P(B\|C)$	$P(A^cB\|C) = P(A^c\|C)P(B\|C)$

$P(A^c\|B^cC) = P(a^c\|C)$	$P(B^c\|A^cC) = P(B^c\|C)$	$P(A^cB^c\|C) = P(A^c\|C)P(B^c\|C)$

Table 5.2.

$P(A\|BC) = P(A\|B^cC)$	$P(A^c\|B^cC) = P(A^c\|B^cC)$	$P(B\|AC) = P(B\|A^cC)$	$P(B^c\|AC) = P(B^c\|A^cC)$

The patterns of conditioning in the examples above belong to this set. In a given problem, one or the other of these conditions may seem a reasonable assumption. As soon as *one* of these patterns is recognized, then *all* are equally valid assumptions. Because of its simplicity and symmetry, we take as the defining condition the *product rule* $P(AB|C) = P(A|C) = P(B|C)$.

Definition

A pair of events $\{A, B\}$ is said to be *conditionally independent, given C*, designated $\{A, B\}$ iff the following product rule holds: $P(AB|C) = P(A|C)P(B|C)$.

The equivalence of the four entries in the right hand column of the upper part of the table, establish

The replacement rule

If any of the pairs $\{A, B\}$, $\{A, B^c\}$, $\{A^c, B\}$ or $\{A^c, B^c\}$ is conditionally independent, given *C*, then so are the others.

— □

This may be expressed by saying that if a pair is conditionally independent, we may replace either or both by their complements and still have a conditionally independent pair.

To illustrate further the usefulness of this concept, we note some other common examples in which similar conditions hold: there is operational independence, but some chance factor which affects both.

- Two contractors work quite independently on jobs in the same city. The operational independence suggests probabilistic independence. However, both jobs are outside and subject to delays due to bad weather. Suppose *A* is the event the first contracter completes his job on time and *B* is the event the second completes on time. If *C* is the event of "good" weather, then arguments similar to those in Examples 1 and 2 make it seem reasonable to suppose $\{A, B\}$ ci $|C$ and $\{A, B\}$ ci $|C^c$. *Remark*. In formal probability theory, an event must be sharply defined: on any trial it occurs or it does not. The event of "good weather" is not so clearly defined. Did a trace of rain or thunder in the area constitute bad weather? Did rain delay on one day in a month long project constitute bad weather? Even with this ambiguity, the pattern of probabilistic analysis may be useful.
- A patient goes to a doctor. A preliminary examination leads the doctor to think there is a thirty percent chance the patient has a certain disease. The doctor orders two independent tests for conditions that indicate the disease. Are results of these tests really independent? There is certainly operational independence—the tests may be done by different laboratories, neither aware of the testing by the others. Yet, if the tests are meaningful, they must both be affected by the actual condition of the patient. Suppose *D* is the event the patient has the disease, *A* is the event the first test is positive (indicates the conditions associated with the disease) and *B* is the event the second test is positive. Then it would seem reasonable to suppose $\{A, B\}$ ci $|D$ and $\{A, B\}$ ci $|D^c$.

In the examples considered so far, it has been reasonable to assume conditional independence, given an event *C*, and conditional independence, given the complementary event. But there are cases in which the effect of the conditioning event is asymmetric. We consider several examples.

- Two students are working on a term paper. They work quite separately. They both need to borrow a certain book from the library. Let *C* be the event the library has two copies available. If *A* is the event the first completes on time and *B* the event the second is successful, then it seems reasonable to assume $\{A, B\}$ ci $|C$. However, if only one book is available, then the two conditions would *not* be conditionally independent. In general $P(B|AC^c) < P(B|C^c)$, since if the first student completes on time, then he or she must have been successful in getting the book, to the detriment of the second.
- If the two contractors of the example above both need material which may be in scarce supply, then successful completion would be conditionally independent, give an adequate supply, whereas they would not be conditionally independent, given a short supply.

- Two students in the same course take an exam. If they prepared separately, the event of both getting good grades should be conditionally independent. If they study together, then the likelihoods of good grades would not be independent. With neither cheating or collaborating on the test itself, if one does well, the other should also.

Since conditional independence is ordinary independence with respect to a conditional probability measure, it should be clear how to extend the concept to larger classes of sets.

Definition

A class $\{A_i : i \in J\}$, where J is an arbitrary index set, is *conditionally independent, given* event C, denoted $\{A_i : i \in J\}$ ci $|C$, iff the product rule holds for every finite subclass of two or more.

As in the case of simple independence, the replacement rule extends.

The replacement rule

If the class $\{A_i : i \in J\}$ ci $|C$, then any or all of the events A_i may be replaced by their complements and still have a conditionally independent class.

The use of independence techniques

Since conditional independence is independence, we may use independence techniques in the solution of problems. We consider two types of problems: an inference problem and a conditional Bernoulli sequence.

Example 5.1.3 Use of independence techniques

Sharon is investigating a business venture which she thinks has probability 0.7 of being successful. She checks with five "independent" advisers. If the prospects are sound, the probabilities are 0.8, 0.75, 0.6, 0.9, and 0.8 that the advisers will advise her to proceed; if the venture is not sound, the respective probabilities are 0.75, 0.85, 0.7, 0.9, and 0.7 that the advice will be negative. Given the quality of the project, the advisers are independent of one another in the sense that no one is affected by the others. Of course, they are not independent, for they are all related to the soundness of the venture. We may reasonably assume conditional independence of the advice, given that the venture is sound and also given that the venture is not sound. If Sharon goes with the majority of advisers, what is the probability she will make the right decision?

Solution

If the project is sound, Sharon makes the right choice if three or more of the five advisors are positive. If the venture is unsound, she makes the right choice if three or more of the five advisers are negative. Let $H =$ the event the project is sound, $F =$ the event three or more advisers are positive, $G = F^c =$ the event three or more are negative, and $E =$ the event of the correct decision. Then

$$P(E) = P(FH) + P(GH^c) = P(F|H)P(H) + P(G|H^c)P(H^c)$$

Let E_i be the event the ith adviser is positive. Then $P(F|H) =$ the sum of probabilities of the form $P(M_k|H)$, where M_k are minterms generated by the class $\{E_i : 1 \le i \le 5\}$. Because of the assumed conditional independence,

$$P(E_1E_2^cE_3^cE_4E_5|H) = P(E_1|H)P(E_2^c|H)P(E_3^c|H)P(E_4|H)P(E_5|H)$$

with similar expressions for each $P(M_k|H)$ and $P(M_k|H^c)$. This means that if we want the probability of three or more successes, given H, we can use ckn with the matrix of conditional probabilities. The following MATLAB solution of the investment problem is indicated.

```
P1 = 0.01*[80 75 60 90 80];
P2 = 0.01*[75 85 70 90 70];
PH = 0.7;
```

```
PE = ckn(P1,3)*PH + ckn(P2,3)*(1 - PH)
PE =    0.9255
```

Often a Bernoulli sequence is related to some conditioning event *H*. In this case it is reasonable to assume the sequence $\{E_i : 1 \le i \le n\}$ ci $|H$ and ci $|H^c$. We consider a simple example.

Example 5.1.4 Test of a claim

A race track regular claims he can pick the winning horse in any race 90 percent of the time. In order to test his claim, he picks a horse to win in each of ten races. There are five horses in each race. If he is simply guessing, the probability of success on each race is 0.2. Consider the trials to constitute a Bernoulli sequence. Let H be the event he is correct in his claim. If S is the number of successes in picking the winners in the ten races, determine $P(H|S = k)$ for various numbers k of correct picks. Suppose it is equally likely that his claim is valid or that he is merely guessing. We assume two conditional Bernoulli trials:

claim is valid: Ten trials, probability $p = P(E_i|H) = 0.9$.

Guessing at random: Ten trials, probability $p = P(E_i|H^c) = 0.2$.

Let $S =$ number of correct picks in ten trials. Then

$$\frac{P(H|S = k}{P(H^c|S = k)} = \frac{P(H)}{P(H^c)} \cdot \frac{P(S = k|H)}{P(S = k|H^c)}, 0 \le k \le 10$$

Giving him the benefit of the doubt, we suppose $P(H)/P(H^c) = 1$ and calculate the conditional odds.

```
k = 0:10;
Pk1 = ibinom(10,0.9,k);    % Probability of k successes, given H
Pk2 = ibinom(10,0.2,k);    % Probability of k successes, given H^c
OH  = Pk1./Pk2;            % Conditional odds-- Assumes P(H)/P(H^c) = 1
e   = OH > 1;              % Selects favorable odds
disp(round([k(e);OH(e)']))
           6           2      % Needs at least six to have creditability
           7          73      % Seven would be creditable,
           8        2627      % even if P(H)/P(H^c) = 0.1
           9       94585
          10     3405063
```

Under these assumptions, he would have to pick at least seven correctly to give reasonable validation of his claim.

This page titled 5.1: Conditional Independence is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

5.2: Patterns of Probable Inference

Some Patterns of Probable Inference

We are concerned with the likelihood of some hypothesized condition. In general, we have evidence for the condition which can never be absolutely certain. We are forced to assess probabilities (likelihoods) on the basis of the evidence. Some typical examples:

Table 5.3.

HYPOTHESIS	EVIDENCE
Job success	Personal traits
Presence of oil	Geological structures
Operation of a device	Physical condition
Market condition	Test market condition
Presence of a disease	Tests for symptoms

If H is the event the hypothetical condition exists and E is the event the evidence occurs, the probabilities available are usually $P(H)$ (or an odds value), $P(E|H)$, and . What is desired is $P(H|E)$ or, equivalently, the odds $P(H|E)/P(H^c|E)$. We simply use Bayes' rule to reverse the direction of conditioning.

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(E|H)}{P(E|H^c)} \cdot \frac{P(H)}{P(H^c)}$$

No conditional independence is involved in this case.

Independent evidence for the hypothesized condition

Suppose there are two “independent” bits of evidence. Now obtaining this evidence may be “operationally” independent, but if the items both relate to the hypothesized condition, then they cannot be really independent. The condition assumed is usually of the form $P(E_1|H) = P(E_1|HE_2)$ —if H occurs, then knowledge of E_2 does not affect the likelihood of E_1. Similarly, we usually have $P(E_1|H^c) = P(E_1|H^cE_2)$. Thus $\{E_1, E_2\}$ ci $|H$ and $\{E_1, E_2\}$ ci $|H^c$.

Example 5.2.1 Independent medical tests

Suppose a doctor thinks the odds are 2/1 that a patient has a certain disease. She orders two independent tests. Let H be the event the patient has the disease and E_1 and E_2 be the events the tests are positive. Suppose the first test has probability 0.1 of a false positive and probability 0.05 of a false negative. The second test has probabilities 0.05 and 0.08 of false positive and false negative, respectively. If both tests are positive, what is the posterior probability the patient has the disease?

Solution

Assuming $\{E_1, E_2\}$ ci $|H$ and ci $|H^c$, we work first in terms of the odds, then convert to probability.

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{P(H)}{P(H^c)} \cdot \frac{P(E_1E_2|H)}{P(E_1E_2|H^c)} = \frac{P(H)}{P(H^c)} \cdot \frac{P(E_1|H)P(E_2|H)}{P(E_1|H^c)P(E_2|H^c)}$$

The data are

$$P(H)/P(H^c) = 2,\ P(E_1|H) = 0.95,\ P(E_1|H^c) = 0.1,\ P(E_2|H) = 0.92,\ P(E_2|H^c) = 0.05$$

Substituting values, we get

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2} = 2 \cdot \frac{0.95 \cdot 0.92}{0.10 \cdot 0.05} = \frac{1748}{5} \text{ so that } P(H|E_1E_2) = \frac{1748}{1753} = 1 - \frac{5}{1753} = 1 - 0.0029$$

Evidence for a symptom

Sometimes the evidence dealt with is not evidence for the hypothesized condition, but for some condition which is stochastically related. For purposes of exposition, we refer to this intermediary condition as a *symptom*. Consider again the examples above.

Table 5.4.

HYPOTHESIS	SYMPTOM	EVIDENCE
Job success	Personal traits	Diagnostic test results
Presence of oil	Geological structures	Geophysical survey results
Operation of a device	Physical condition	Monitoring report
Market condition	Test market condition	Market survey result
Presence of a disease	Physical symptom	Test for symptom

We let S be the event the symptom is present. The usual case is that the evidence is directly related to the symptom and not the hypothesized condition. The diagnostic test results can say something about an applicant's personal traits, but cannot deal directly with the hypothesized condition. The test results would be the same whether or not the candidate is successful in the job (he or she does not have the job yet). A geophysical survey deals with certain structural features beneath the surface. If a fault or a salt dome is present, the geophysical results are the same whether or not there is oil present. The physical monitoring report deals with certain physical characteristics. Its reading is the same whether or not the device will fail. A market survey treats only the condition in the test market. The results depend upon the test market, not the national market. A blood test may be for certain physical conditions which frequently are related (at least statistically) to the disease. But the result of the blood test for the physical condition is not directly affected by the presence or absence of the disease.

Under conditions of this type, we may assume

$$P(E|SH) = P(E|SH^c) \text{ and } P(E|S^cH) = P(E|S^cH^c)$$

These imply $\{E, H\}$ ci $|S$ and ci $|S^c$. Now

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(HE)}{P(H^cE)} = \frac{P(HES) + P(HES^c)}{P(H^cES) + P(H^cES^c)} = \frac{P(HS)P(E|HS) + P(HS^c)P(E|HS^c)}{P(H^cS)P(E|H^cS) + P(H^cS^c)P(E|H^cS^c)}$$

$$= \frac{P(HS)P(E|S)P(HS^c)P(E|S^c)}{P(H^cS)P(E|S) + P(H^cS^c)P(E|S^c)}$$

It is worth noting that each term in the denominator differs from the corresponding term in the numerator by having H^c in place of H. Before completing the analysis, it is necessary to consider how H and S are related stochastically in the data. Four cases may be considered.

Data are $P(S|H)$, $P(S|H^c)$, and $P(H)$.
Data are $P(S|H)$, $P(S|H^c)$, and $P(S)$.
Data are $P(H|S)$, $P(H|S^c)$, and $P(S)$.
Data are $P(H|S)$, $P(H|S^c)$, and $P(H)$.

Case a:

\dfrac{P(H|S)}{P(H^c|S)} = \dfrac{P(H) P(S|H) P(E|S) + P(H) P(S^c|H) P(E|S^c)}{P(H^c) P(S|H^c) P(E|S) + P(H^c) P(S^c|H^c) P(E|S^c)}\)

Example 5.2.2 Geophysical survey

Let H be the event of a successful oil well, S be the event there is a geophysical structure favorable to the presence of oil, and E be the event the geophysical survey indicates a favorable structure. We suppose $\{H, E\}$ ci $|S$ and ci $|S^c$. Data are

$$P(H)/P(H^c) = 3,\ P(S|H) = 0.92,\ P(S|H^c) = 0.20,\ P(E|S) = 0.95,\ P(E|S^c) = 0.15$$

Then

$$\frac{P(H|E)}{P(H^c|E)} = 3 \cdot \frac{0.92 \cdot 0.95 + 0.08 \cdot 0.15}{0.20 \cdot 0.95 + 0.80 \cdot 0.15} = \frac{1329}{155} = 8.5742$$

$$\text{so that } P(H|E) = 1 - \frac{155}{1484} = 0.8956$$

The geophysical result moved the prior odds of 3/1 to posterior odds of 8.6/1, with a corresponding change of probabilities from 0.75 to 0.90.

Case b: Data are $P(S)P(S|H)$, $P(S|H^c)$, $P(E|S)$, and $P(E|S^c)$. If we can determine $P(H)$, we can proceed as in case a. Now by the law of total probability

$$P(S) = P(S|H)P(H) + P(S|H^c)[1 - P(H)]$$

which may be solved algebraically to give

$$P(H) = \frac{P(S) - P(S|H^c)}{P(S|H) - P(S|H^c)}$$

Example 5.2.3 Geophysical survey revisited

In many cases a better estimate of $P(S)$ or the odds $P(S)/P(S^c)$ can be made on the basis of previous geophysical data. Suppose the prior odds for S are 3/1, so that $P(S) = 0.75$. Using the other data in Example, we have

$$P(H) = \frac{P(S) - P(S|H^c)}{P(S|H) - P(S|H^c)} = \frac{0.75 - 0.20}{0.92 - 0.20} = 55/72, \text{ so that } \frac{\(P(H)}{P(H^c)} = 55/17$$

Using the pattern of case a, we have

$$\frac{P(H|E)}{P(H^c|E)} = \frac{55}{17} \cdot \frac{0.92 \cdot 0.95 + 0.08 \cdot 0.15}{0.20 \cdot 0.95 + 0.80 \cdot 0.15} = \frac{4873}{527} = 9.2467$$

$$\text{so that } P(H|E) = 1 - \frac{527}{5400} = 0.9024$$

Usually data relating test results to symptom are of the form $P(E|S)$ and $P(E|S^c)$, or equivalent. Data relating the symptom and the hypothesized condition may go either way. In cases a and b, the data are in the form $P(S|H)$ and $P(S|H^c)$, or equivalent, derived from data showing the fraction of times the symptom is noted when the hypothesized condition is identified. But these data may go in the opposite direction, yielding $P(H|S)$ and $P(H|S^c)$, or equivalent. This is the situation in cases c and d.

Data c: Data are $P(E|S)$, $P(E|S^c)$, $P(H|S)$, $P(H|S^c)$ and $P(S)$.

Example 5.2.4 Evidence for a disease symptom with prior P(S)

When a certain blood syndrome is observed, a given disease is indicated 93 percent of the time. The disease is found without this syndrome only three percent of the time. A test for the syndrome has probability 0.03 of a false positive and 0.05 of a false negative. A preliminary examination indicates a probability 0.30 that a patient has the syndrome. A test is performed; the result is negative. What is the probability the patient has the disease?

Solution

In terms of the notation above, the data are

$$P(S) = 0.30,\ P(E|S^c) = 0.03,\ P(E^c|S) = 0.05$$

$$P(H|S) = 0.93, \text{ and } P(H|S^c) = 0.03$$

We suppose $\{H, E\}$ ci $|S$ and ci $|S^c$.

$$\frac{P(H|E^c)}{P(H^c|E^c)} = \frac{P(S)P(H|S)P(E^c|S) + P(S^c)P(H|S^c)P(E^c|S^c)}{P(S)P(H^c|S)P(E^c|s) + P(S^c)P(H^c|S^c)P(E^c|S^c)}$$

$$= \frac{0.30 \cdot 0.93 \cdot 0.05 + 0.07 \cdot 0.03 \cdot 0.97}{0.30 \cdot 0.07 \cdot 0.05 + 0.70 \cdot 0.97 \cdot 0.97} = \frac{429}{8246}$$

which implies $P(H|E^c) = 429/8675 \approx 0.05$

Case d: This differs from case c only in the fact that a prior probability for H is assumed. In this case, we determine the corresponding probability for S by

$$P(S) = \frac{P(H) - P(H|S^c)}{P(H|S) - P(H|S^c)}$$

and use the pattern of case c.

Example 5.2.5 Evidence for a disease symptom with prior P(h)

Suppose for the patient in Example the physician estimates the odds favoring the presence of the disease are 1/3, so that $P(H) = 0.25$. Again, the test result is negative. Determine the posterior odds, given E^c.

Solution

First we determine

$$P(S) = \frac{P(H) - P(H|S^c)}{P(H|S) - P(H|S^c)} = \frac{0.25 - 0.03}{0.93 - 0.03} = 11/45$$

Then

$$\frac{P(H|E^c)}{P(H^c|E^c)} = \frac{(11/45) \cdot 0.93 \cdot 0.05 + (34/45) \cdot 0.03 \cdot 0.97}{(11/45) \cdot 0.07 \cdot 0.05 + (34/45) \cdot 0.97 \cdot 0.97} = \frac{15009}{320291} = 0.047$$

The result of the test drops the prior odds of 1/3 to approximately 1/21.

Independent evidence for a symptom

In the previous cases, we consider only a single item of evidence for a symptom. But it may be desirable to have a "second opinion." We suppose the tests are for the symptom and are not directly related to the hypothetical condition. If the tests are operationally independent, we could reasonably assume

$$P(E_1|SE_2) = P(E_1|SE_2^c)\ \{E_1, E_2\} \text{ ci } |S$$
$$P(E_1|SH) = P(E_1|SH^c)\ \{E_1, H\} \text{ ci } |S$$
$$P(E_2|SH) = P(E_2|SH^c)\ \{E_2, H\} \text{ ci } |S$$
$$P(E_1E_2|SH) = P(E_1E_2|SH^c)\ \{E_1, E_2, H\} \text{ ci } |S$$

This implies $\{E_1, E_2, H\}$ ci $|S$. A similar condition holds for S^c. As for a single test, there are four cases, depending on the tie between S and H. We consider a "case a" example.

Example 5.2.6 A market survey problem

A food company is planning to market nationally a new breakfast cereal. Its executives feel confident that the odds are at least 3 to 1 the product would be successful. Before launching the new product, the company decides to investigate a test market. Previous experience indicates that the reliability of the test market is such that if the national market is favorable, there is probability 0.9 that the test market is also. On the other hand, if the national market is unfavorable, there is a probability of only 0.2 that the test market will be favorable. These facts lead to the following analysis. Let

H be the event the national market is favorable (hypothesis)

S be the event the test market is favorable (symptom)

The *initial data* are the following probabilities, based on past experience:

- (a) Prior odds: $P(H)/P(H^c) = 3$
- (b) Reliability of the test market: $P(S|H) = 0.9\ P(S|H^c) = 0.2$

If it were known that the test market is favorable, we should have

$$\frac{P(H|S)}{P(H^c|S)} = \frac{P(S|H)P(H)}{P(S|H^c)P(H^c)} = \frac{0.9}{0.2} \cdot 3 = 13.5$$

Unfortunately, it is not feasible to know with certainty the state of the test market. The company decision makers engage two market survey companies to make independent surveys of the test market. The reliability of the companies may be expressed as follows. Let

E_1 be the event the first company reports a favorable test market.
E_2 be the event the second company reports a favorable test market.

On the basis of previous experience, the reliability of the evidence about the test market (the symptom) is expressed in the following conditional probabilities.

$$P(E_1|S) = 0.9\ P(E_1|S^c) = 0.3\ P(E_2|S) = 0.8\ B(E_2|S^c) = 0.2$$

Both survey companies report that the test market is favorable. What is the probability the national market is favorable, given this result?

Solution

The two survey firms work in an “operationally independent” manner. The report of either company is unaffected by the work of the other. Also, each report is affected only by the condition of the test market— regardless of what the national market may be. According to the discussion above, we should be able to assume

$$\{E_1, E_2, H\} \text{ ci } |S \text{ and } \{E_1, E_2, H\} \text{ ci } S^c$$

We may use a pattern similar to that in Example 2, as follows:

$$\backslash\text{dfrasc}P(H|E_1E_2)P(H^c|E_1E_2) = \frac{P(H)}{P(H^c)} \cdot \frac{P(S|H)P(E_1|S)P(E_2|S) + P(S^c|H)P(E_1|S^c)P(E_2|S^2)}{P(S|H^c)P(E_1|S)P(E_2|S) + P(S^c|H^c)P(E_1|S^c)P(E_2|S^c)}$$

$$= 3 \cdot \frac{0.9 \cdot 0.9 \cdot 0.8 + 0.1 \cdot 0.3 \cdot 0.2}{0.2 \cdot 0.9 \cdot 0.8 + 0.8 \cdot 0.3 \cdot 0.2} = \frac{327}{32} \approx 10.22$$

in terms of the posterior probability, we have

$$P(H|E_1E_2) = \frac{327/32}{1 + 327/32} = \frac{327}{359} = 1 - \frac{32}{359} \approx 0.91$$

We note that the odds favoring H, given positive indications from both survey companies, is 10.2 as compared with the odds favoring *H*, given a favorable test market, of 13.5. The difference reflects the residual uncertainty about the test market after the market surveys. Nevertheless, the results of the market surveys increase the odds favoring a satisfactory market from the prior 3 to 1 to a posterior 10.2 to 1. In terms of probabilities, the market surveys increase the likelihood of a favorable market from the original $P(H) = 0.75$ to the posterior $P(H|E_1E_2)$. The conditional independence of the results of the survey makes possible direct use of the data.

A classification problem

A population consists of members of two subgroups. It is desired to formulate a battery of questions to aid in identifying the subclass membership of randomly selected individuals in the population. The questions are designed so that for each individual the answers are independent, in the sense that the answers to any subset of these questions are not affected by and do not affect the answers to any other subset of the questions. The answers are, however, affected by the subgroup membership. Thus, our treatment of conditional idependence suggests that it is reasonable to supose the answers are conditionally independent, given the subgroup membership. Consider the following numerical example.

Example 5.2.7 A classification problem

A sample of 125 subjects is taken from a population which has two subgroups. The subgroup membership of each subject in the sample is known. Each individual is asked a battery of ten questions designed to be independent, in the sense that the answer to any one is not affected by the answer to any other. The subjects answer independently. Data on the results are summarized in the following table:

Table 5.5.

GROUP 1 (69 members)				GROUP 2 (56 members)		
Q	Yes	No	Unc.	Yes	No	Unc.
1	42	22	5	20	31	5
2	34	27	8	16	37	3
3	15	45	9	33	19	4
4	19	44	6	31	18	7
5	22	43	4	23	28	5
6	41	13	15	14	37	5
7	9	52	8	31	17	8
8	40	26	3	13	38	5
9	48	12	9	27	24	5
10	20	37	12	35	16	5

Assume the data represent the general population consisting of these two groups, so that the data may be used to calculate probabilities and conditional probabilities.

Several persons are interviewed. The result of each interview is a “profile” of answers to the questions. The goal is to classify the person in one of the two subgroups on the basis of the profile of answers.

The following profiles were taken.

- Y, N, Y, N, Y, U, N, U, Y. U
- N, N, U, N, Y, Y, U, N, N, Y
- Y, Y, N, Y, U, U, N, N, Y, Y

Classify each individual in one of the subgroups.

Solution

Let $G_1 =$ the event the person selected is from group 1, and $G_2 = G_1^c =$ the event the person selected is from group 2. Let

A_i = the event the answer to the ith question is “Yes”

B_i = the event the answer to the ith question is “No”

C_i = the event the answer to the ith question is “Uncertain”

The data are taken to mean $P(A_1|G_1) = 42/69$, $P(B_3|G_2) = 19/56$, etc. The profile

Y, N, Y, N, Y, U, N, U, Y. U corresponds to the event $E = A_1B_2A_3B_4A_5C_6B_7C_8A_9C_{10}$

We utilize the ratio form of Bayes' rule to calculate the posterior odds

$$\frac{P(G_1|E)}{P(G_2|E)} = \frac{P(E|G_1)}{P(E|G_2)} \cdot \frac{P(G_1)}{P(G_2)}$$

If the ratio is greater than one, classify in group 1; otherwise classify in group 2 (we assume that a ratio exactly one is so unlikely that we can neglect it). Because of conditional independence, we are able to determine the conditional probabilities

$$P(E|G_1) = \frac{42 \cdot 27 \cdot 15 \cdot 44 \cdot 22 \cdot 15 \cdot 52 \cdot 3 \cdot 48 \cdot 12}{69^{10}} \text{ and}$$

$$P(E|G_2) = \frac{29 \cdot 37 \cdot 33 \cdot 18 \cdot 23 \cdot 5 \cdot 17 \cdot 5 \cdot 24 \cdot 5}{56^{10}}$$

The odds $P(G_2)/P(G_2) = 69/56$. We find the posterior odds to be

$$\frac{P(G_1|E)}{P(G_2|E)} = \frac{42 \cdot 27 \cdot 15 \cdot 44 \cdot 22 \cdot 15 \cdot 52 \cdot 3 \cdot 48 \cdot 12}{29 \cdot 37 \cdot 33 \cdot 18 \cdot 23 \cdot 5 \cdot 17 \cdot 5 \cdot 24 \cdot 5} \cdot \frac{56^9}{69^9} = 5.85$$

The factor $56^9/69^9$ comes from multiplying $56^{10}/69^{10}$ by the odds $P(G_1)/P(G_2) = 69/56$. Since the resulting posterior odds favoring Group 1 is greater than one, we classify the respondent in group 1.

While the calculations are simple and straightforward, they are tedious and error prone. To make possible rapid and easy solution, say in a situation where successive interviews are underway, we have several m-procedures for performing the calculations. Answers to the questions would normally be designated by some such designation as Y for yes, N for no, and U for uncertain. In order for the m-procedure to work, these answers must be represented by numbers indicating the appropriate columns in matrices *A* and *B*. Thus, in the example under consideration, each Y must be translated into a 1, each N into a 2, and each U into a 3. The task is not particularly difficult, but it is much easier to have MATLAB make the translation as well as do the calculations. The following two-stage approach for solving the problem works well.

The first m-procedure *oddsdf* sets up the frequency information. The next m-procedure *odds* calculates the odds for a given profile. The advantage of splitting into two m-procedures is that we can set up the data once, then call repeatedly for the calculations for different profiles. As always, it is necessary to have the data in an appropriate form. The following is an example in which the data are entered in terms of actual frequencies of response.

```
01 % file oddsf4.m
02 % Frequency data for classification
```

```
A = [42 22 5; 34 27 8; 15 45 9; 19 44 6; 22 43 4;
     41 13 15; 9 52 8; 40 26 3; 48 12 9; 20 37 12];
B = [20 31 5; 16 37 3; 33 19 4; 31 18 7; 23 28 5;
     14 37 5; 31 17 8; 13 38 5; 27 24 5; 35 16 5];
disp('Call for oddsdf')
```

Example 5.2.8 Classification using frequency data

```
oddsf4              % Call for data in file oddsf4.m
Call for oddsdf     % Prompt built into data file
oddsdf              % Call for m-procedure oddsdf
Enter matrix A of frequencies for calibration group 1  A
Enter matrix B of frequencies for calibration group 2  B
Number of questions = 10
Answers per question = 3
 Enter code for answers and call for procedure "odds"
y = 1;              % Use of lower case for easier writing
n = 2;
u = 3;
odds                % Call for calculating procedure
Enter profile matrix E  [y n y n y u n u y u]   % First profile
Odds favoring Group 1:   5.845
Classify in Group 1
odds                % Second call for calculating procedure
Enter profile matrix E  [n n u n y y u n n y]   % Second profile
Odds favoring Group 1:   0.2383
Classify in Group 2
odds                % Third call for calculating procedure
Enter profile matrix E  [y y n y u u n n y y]   % Third profile
Odds favoring Group 1:   5.05
Classify in Group 1
```

The principal feature of the m-procedure odds is the scheme for selecting the numbers from the A and B matrices. If E = [$yynyuunnyy$] , then the coding translates this into the actual numerical matrix

[1 1 2 1 3 3 2 2 1 1] used internally. Then $A(:, E)$ is a matrix with columns corresponding to elements of E. Thus

```
e = A(:,E)
e =   42    42    22    42     5     5    22    22    42    42
      34    34    27    34     8     8    27    27    34    34
      15    15    45    15     9     9    45    45    15    15
      19    19    44    19     6     6    44    44    19    19
      22    22    43    22     4     4    43    43    22    22
      41    41    13    41    15    15    13    13    41    41
       9     9    52     9     8     8    52    52     9     9
      40    40    26    40     3     3    26    26    40    40
      48    48    12    48     9     9    12    12    48    48
      20    20    37    20    12    12    37    37    20    20
```

The ith entry on the ith column is the count corresponding to the answer to the ith question. For example, the answer to the third question is N (no), and the corresponding count is the third entry in the N (second) column of A. The element on the diagonal in the third column of $A(:, E)$ is the third element in that column, and hence the desired third entry of the N column. By picking out the elements on the diagonal by the command diag(A(:,E)), we have the desired set of counts corresponding to the profile. The same is true for diag(B(:,E)).

Sometimes the data are given in terms of conditional probabilities and probabilities. A slight modification of the procedure handles this case. For purposes of comparison, we convert the problem above to this form by converting the counts in matrices A and B to conditional probabilities. We do this by dividing by the total count in each group (69 and 56 in this case). Also, $P(G_1) = 69/125 = 0.552$ and $P(G_2) = 56/125 = 0.448$.

Table 5.6.

GROUP 1 $P(G_1) = 69/125$				**GROUP 2** $P(G_2) = 56/125$		
Q	Yes	No	Unc.	Yes	No	Unc.
1	0.6087	0.3188	0.0725	0.3571	0.5536	0.0893
2	0.4928	0.3913	0.1159	0.2857	0.6607	0.0536
3	0.2174	0.6522	0.1304	0.5893	0.3393	0.0714
4	0.2754	0.6376	0.0870	0.5536	0.3214	0.1250
5	0.3188	0.6232	0.0580	0.4107	0.5000	0.0893
6	0.5942	0.1884	0.2174	0.2500	0.6607	0.0893
7	0.1304	0.7536	0.1160	0.5536	0.3036	0.1428
8	0.5797	0.3768	0.0435	0.2321	0.6786	0.0893
9	0.6957	0.1739	0.1304	0.4821	0.4286	0.0893
10	0.2899	0.5362	0.1739	0.6250	0.2857	0.0893

These data are in an m-file *oddsp4.m*. The modified setup m-procedure *oddsdp* uses the conditional probabilities, then calls for the m-procedure odds.

Example 5.2.9 Calculation using conditional probability data

```
oddsp4                % Call for converted data (probabilities)
oddsdp                % Setup m-procedure for probabilities
Enter conditional probabilities for Group 1  A
Enter conditional probabilities for Group 2  B
Probability p1 individual is from Group 1  0.552
 Number of questions = 10
 Answers per question = 3
 Enter code for answers and call for procedure "odds"
y = 1;
n = 2;
u = 3;
odds
Enter profile matrix E  [y n y n y u n u y u]
Odds favoring Group 1:  5.845
Classify in Group 1
```

The slight discrepancy in the odds favoring Group 1 (5.8454 compared with 5.8452) can be attributed to rounding of the conditional probabilities to four places. The presentation above rounds the results to 5.845 in each case, so the discrepancy is not apparent. This is quite acceptable, since the discrepancy has no effect on the results.

This page titled 5.2: Patterns of Probable Inference is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

5.3: Problems on Conditional Independence

Exercise 5.3.1

Suppose $\{A., B\}$ ci $|C$ and $\{A, B\}$ ci $|C^c$, $P(C) = 0.7$, and

$$P(A|C) = 0.4,\ P(B|C) = 0.6,\ P(A|C^c) = 0.3,\ P(B|C^c) = 0.2$$

Show whether or not the pair $\{A., B\}$ is independent.

Answer available at the end of this section.

Exercise 5.3.2

Suppose $\{A_1, A_2, A_3\}$ ci $|C$ and ci $|C^c$, with $P(C) = 0.4$, and

$$P(A_i|C) = 0.90, 0.85, 0.80\ P(A_i|C^c) = 0.20, 0.15, 0.20 \text{ for } i = 1, 2, 3, \text{ respectively}$$

Determine the posterior odds $P(C|A_1A_2^cA_3)/P(C^c|A_1A_2^cA_3)$.

Answer available at the end of this section.

Exercise 5.3.3

Five world class sprinters are entered in a 200 meter dash. Each has a good chance to break the current track record. There is a thirty percent chance a late cold front will move in, bringing conditions that adversely affect the runners. Otherwise, conditions are expected to be favorable for an outstanding race. Their respective probabilities of breaking the record are:

- Good weather (no front): 0.75, 0.80, 0.65, 0.70, 0.85
- Poor weather (front in): 0.60, 0.65, 0.50, 0.55, 0.70

The performances are (conditionally) independent, given good weather, and also, given poor weather. What is the probability that three or more will break the track record?

Hint. If B_3 is the event of three or more, $P(B_3) = P(B_3|W)P(W) + P(B_3|W^c)P(W^c)$.

Answer available at the end of this section.

Exercise 5.3.4

A device has five sensors connected to an alarm system. The alarm is given if three or more of the sensors trigger a switch. If a dangerous condition is present, each of the switches has high (but not unit) probability of activating; if the dangerous condition does not exist, each of the switches has low (but not zero) probability of activating (falsely). Suppose $D =$ the event of the dangerous condition and $A =$ the event the alarm is activated. Proper operation consists of $AD \bigvee A^cD^c$. Suppose $E_i =$ the event the ith unit is activated. Since the switches operate independently, we suppose

$$\{E_1, E_2, E_3, E_4, E_5\} \text{ ci } |D \text{ and ci } |D^c$$

Assume the conditional probabilities of the E_1, given D, are 0.91, 0.93, 0.96, 0.87, 0.97, and given D^c, are 0.03, 0.02, 0.07, 0.04, 0.01, respectively. If $P(D) = 0.02$, what is the probability the alarm system acts properly? *Suggestion*. Use the conditional independence and the procedure ckn.

Answer

```
P1 = 0.01*[91 93 96 87 97];
P2 = 0.01*[3 2 7 4 1];
P  = ckn(P1,3)*0.02 + (1 - ckn(P2,3))*0.98
P =  0.9997
```

Exercise 5.3.5

Seven students plan to complete a term paper over the Thanksgiving recess. They work independently; however, the likelihood of completion depends upon the weather. If the weather is very pleasant, they are more likely to engage in outdoor activities and put off work on the paper. Let E_i be the event the ith student completes his or her paper, A_k be the event that k or more complete during the recess, and *W* be the event the weather is highly conducive to outdoor activity. It is reasonable to suppose $\{E_i : 1 \le i \le 7\}$ and ci $|W^c$. Suppose

$$P(E_i|W) = 0.4, 0.5, 0.3, 0.7, 0.5, 0.6, 0.2$$

$$P(E_i|W^c) = 0.7, 0.8, 0.5, 0.9, 0.7, 0.8, 0.5$$

respectively, and $P(W) = 0.8$. Determine the probability $P(A_4)$ that four our more complete their papers and $P(A_5)$ that five or more finish.

Answer available at the end of this section.

Exercise 5.3.6

A manufacturer claims to have improved the reliability of his product. Formerly, the product had probability 0.65 of operating 1000 hours without failure. The manufacturer claims this probability is now 0.80. A sample of size 20 is tested. Determine the odds favoring the new probability for various numbers of surviving units under the assumption the prior odds are 1 to 1. How many survivors would be required to make the claim creditable?

Answer available at the end of this section.

Exercise 5.3.7

A real estate agent in a neighborhood heavily populated by affluent professional persons is working with a customer. The agent is trying to assess the likelihood the customer will actually buy. His experience indicates the following: if *H* is the event the customer buys, S is the event the customer is a professional with good income, and *E* is the event the customer drives a prestigious car, then

$$P(S) = 0.7 \; P(S|H) = 0.90 \; P(S|H^c) = 0.2 \; P(E|S) = 0.95 \; P(E|S^c) = 0.25$$

Since buying a house and owning a prestigious car are not related for a given owner, it seems reasonable to suppose $P(E|HS) = P(E|H^cS)$ and $P(E|HS^c) = P(E|H^cS^c)$. The customer drives a Cadillac. What are the odds he will buy a house?

Answer

Assumptions amount to $\{H, E\}$ ci $|S$ and ci $|S^c$.

$$\frac{P(H|S)}{P(H^c|S)} = \frac{P(H)P(S|H)}{P(H^c)P(S|H^c)}$$

$P(S) = P(H)P(S|H) + [1 - P(H)]P(S|H^c)$ which implies

$$P(H) = \frac{P(S) - P(S|H^c)}{P(S|H) - P(S|H^c)} = 5/7 \text{ so that } \frac{P(H|S)}{P(H^c|S)} = \frac{5}{2} \cdot \frac{0.9}{0.2} = \frac{45}{4}$$

Exercise 5.3.8

In deciding whether or not to drill an oil well in a certain location, a company undertakes a geophysical survey. On the basis of past experience, the decision makers feel the odds are about four to one favoring success. Various other probabilities can be assigned on the basis of past experience. Let

- H be the event that a well would be successful
- S be the event the geological conditions are favorable
- E be the event the results of the geophysical survey are positive

The initial, or prior, odds are $P(H)/P(H^c) = 4$. Previous experience indicates

$$P(S|H) = 0.9 \; P(S|H^c) = 0.20 \; P(E|S) = 0.95 \; P(E|S^c) = 0.10$$

Make reasonable assumptions based on the fact that the result of the geophysical survey depends upon the geological formations and not on the presence or absence of oil. The result of the survey is favorable. Determine the posterior odds $P(H|E)/P(H^c|E)$.

Answer

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \cdot \frac{P(S|H)P(E|S) + P(S^c|H)P(E|S^c)}{P(S|H^c)P(E|S) + P(S^c|H^c)P(E|S^c)}$$

$$= 4 \cdot \frac{0.90 \cdot 0.95 + 0.10 \cdot 0.10}{0.20 \cdot 0.95 + 0.80 \cdot 0.10} = 12.8148$$

Exercise 5.3.9

A software firm is planning to deliver a custom package. Past experience indicates the odds are at least four to one that it will pass customer acceptance tests. As a check, the program is subjected to two different benchmark runs. Both are successful. Given the following data, what are the odds favoring successful operation in practice? Let

- H be the event the performance is satisfactory
- S be the event the system satisfies customer acceptance tests
- E_1 be the event the first benchmark tests are satisfactory.
- E_2 be the event the second benchmark test is ok.

Under the usual conditions, we may assume $\{H, E_1, E_2\}$ ci $|S$ and ci $|S^c$. Reliability data show

$$P(H|S) = 0.95, \; P(H|S^c) = 0.45$$

$$P(E_1|S) = 0.90 \; P(E_1|S^c) = 0.25 \; P(E_2|S) = 0.95 \; P(E_2|S^c) = 0.20$$

Determine the posterior odds $P(H|E_1E_2)/P(H^c|E_1E_2)$.

Answer available at the end of this section.

Exercise 5.3.10

A research group is contemplating purchase of a new software package to perform some specialized calculations. The systems manager decides to do two sets of diagnostic tests for significant bugs that might hamper operation in the intended application. The tests are carried out in an operationally independent manner. The following analysis of the results is made.

- H = the event the program is satisfactory for the intended application

- S = the event the program is free of significant bugs
- E_1 = the event the first diagnostic tests are satisfactory
- E_2 = the event the second diagnostic tests are satisfactory

Since the tests are for the presence of bugs, and are operationally independent, it seems reasonable to assume $\{H, E_1, E_2\}$ ci $|S$ and $\{H, E_1, E_2\}$ ci $|S^c$. Because of the reliability of the software company, the manager thinks $P(S) = 0.85$. Also, experience suggests

$P(H\|S) = 0.95$	$P(E_1\|S) = 0.90$	$P(E_2\|S) = 0.95$
$P(H\|S^c) = 0.30$	$P(E_1\|S^c) = 0.20$	$P(E_2\|S^c) = 0.25$

Determine the posterior odds favoring H if results of both diagnostic tests are satisfactory.

Answer available at the end of this section.

Exercise 5.3.11

A company is considering a new product now undergoing field testing. Let

- H be the event the product is introduced and successful
- S be the event the R&D group produces a product with the desired characteristics.
- E be the event the testing program indicates the product is satisfactory

The company assumes $P(S) = 0.9$ and the conditional probabilities

$$P(H|S) = 0.90 \; P(H|S^c) = 0.10 \; P(E|S) = 0.95 \; P(E|S^c) = 0.15$$

Since the testing of the merchandise is not affected by market success or failure, it seems reasonable to suppose $\{H, E\}$ ci $|S$ and ci $|S^c$. The field tests are favorable. Determine $P(H|E)/P(H^c|E)$.

Answer

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(S)P(H|S)P(E|S) + P(S^c)P(H|S^c)P(E|S^c)}{P(S)P(H^c|S)P(E|S) + P(S^c)P(H^c|S^c)P(E|S^c)}$$

$$= \frac{0.90 \cdot 0.90 \cdot 0.95 + 0.10 \cdot 0.10 \cdot 0.15}{0.90 \cdot 0.10 \cdot 0.95 + 0.10 \cdot 0.90 \cdot 0.15} = 7.7879$$

Exercise 5.3.12

Martha is wondering if she will get a five percent annual raise at the end of the fiscal year. She understands this is more likely if the company's net profits increase by ten percent or more. These will be influenced by company sales volume. Let

- H = the event she will get the raise
- S = the event company profits increase by ten percent or more
- E = the event sales volume is up by fifteen percent or more

Since the prospect of a raise depends upon profits, not directly on sales, she supposes $\{H, E\}$ ci $|S$ and $\{H, E\}$ ci $|S^c$. She thinks the prior odds favoring suitable profit increase is about three to one. Also, it seems reasonable to suppose

$$P(H|S) = 0.80 \; P(H|S^c) = 0.10 \; P(E|S) = 0.95 \; P(E|S^c) = 0.10$$

End of the year records show that sales increased by eighteen percent. What is the probability Martha will get her raise?

Answer

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(S)P(H|S)P(E|S) + P(S^c)P(H|S^c)P(E|S^c)}{P(S)P(H^c|S)P(E|S) + P(S^c)P(H^c|S^c)P(E|S^c)}$$

$$= \frac{0.75 \cdot 0.80 \cdot 0.95 + 0.25 \cdot 0.10 \cdot 0.10}{0.75 \cdot 0.20 \cdot 0.95 + 0.25 \cdot 0.90 \cdot 0.10} = 3.4697$$

Exercise 5.3.13

A physician thinks the odds are about 2 to 1 that a patient has a certain disease. He seeks the "independent" advice of three specialists. Let H be the event the disease is present, and A, B, Cbe the events the respective consultants agree this is the case. The physician decides to go with the majority. Since the advisers act in an operationally independent manner, it seems reasonable to suppose $\{A, B, C\}$ ci $|H$ and ci $|H^c$. Experience indicates

$$P(A|H) = 0.8,\ P(B|H) = 0.7,\ P(C|H) - 0.75$$

$$P(A^c|H^c) = 0.85,\ P(B^c|H^c) = 0.8,\ P(C^c|H^c) = 0.7$$

What is the probability of the right decision (i.e., he treats the disease if two or more think it is present, and does not if two or more think the disease is not present)?

Answer available at the end of this section.

Exercise 5.3.14

A software company has developed a new computer game designed to appeal to teenagers and young adults. It is felt that there is good probability it will appeal to college students, and that if it appeals to college students it will appeal to a general youth market. To check the likelihood of appeal to college students, it is decided to test first by a sales campaign at Rice and University of Texas, Austin. The following analysis of the situation is made.

- H = the event the sales to the general market will be good
- s = the event the game appeals to college students
- E_1= the event the sales are good at Rice
- E_2 = the event the sales are good at UT, Austin

Since the tests are for the reception are at two separate universities and are operationally independent, it seems reasonable to assume $\{H, E_1, E_2\}$ ci $|S$ and $\{H, E_1, E_2\}$ ci $|S^c$. Because of its previous experience in game sales, the managers think $P(S) = 0.80$. Also, experience suggests

$P(H\|S) = 0.95$	$P(E_1\|S) = 0.90$	$P(E_2\|S) = 0.95$
$P(H\|S^c) = 0.30$	$P(E_1\|S^c) = 0.20$	$P(E_2\|S^c) = 0.25$

Determine the posterior odds favoring H if sales results are satisfactory at both schools.

Answer available at the end of this section.

Exercise 5.3.15

In a region in the Gulf Coast area, oil deposits are highly likely to be associated with underground salt domes. If H is the event that an oil deposit is present in an area, and S is the event of a salt dome in the area, experience indicates $P(S|H) = 0.9$ and $P(S|H^c) = 1$. Company executives believe the odds favoring oil in the area is at least 1 in 10. It decides to conduct two independent geophysical surveys for the presence of a salt dome. Let $E - 1, E_2$ be the events the surveys indicate a salt

dome. Because the surveys are tests for the geological structure, not the presence of oil, and the tests are carried out in an operationally independent manner, it seems reasonable to assume $\{H, E_1, E_2\}$ ci $|S$ and ci $|S^c$. Data on the reliability of the surveys yield the following probabilities

$$P(E_1|S) = 0.95 \; P(E_1|S^c) = 0.05 \; P(E_2|S) = 0.90 \; P(E_2|S^c) = 0.10$$

Determine the posterior odds $\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)}$. Should the well be drilled?

Answer

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{P(HE_1E_2S) + P(HE_1E_2S^c)}{P(H^cE_1E_2S) + P(H^cE_1E_2S^c)}$$

$$P(HE_1E_2S) = P(H)P(S|H)P(E_1|SH)P(E_2|SHE_1) = P(H)P(S|H)P(E_1|S)P(E_2|S)$$

with similar expressions for the other terms.

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{1}{10} \cdot \frac{0.9 \cdot 0.95 \cdot 0.90 + 0.10 \cdot 0.05 \cdot 0.10}{0.1 \cdot 0.95 \cdot 0.90 + 0.90 \cdot 0.05 \cdot 0.10} = 0.8556$$

Exercise 5.3.16

A sample of 150 subjects is taken from a population which has two subgroups. The subgroup membership of each subject in the sample is known. Each individual is asked a battery of ten questions designed to be independent, in the sense that the answer to any one is not affected by the answer to any other. The subjects answer independently. Data on the results are summarized in the following table:

GROUP 1 (84 members)				GROUP 2 (66 members)		
Q	Yes	No	Unc	Yes	No	Unc
1	51	26	7	27	34	5
2	42	32	10	19	43	4
3	19	54	11	39	22	5
4	24	53	7	38	19	9
5	27	52	5	28	33	5
6	49	19	16	19	41	6
7	16	59	9	37	21	8
8	47	32	5	19	42	5
9	55	17	12	27	33	6
10	24	53	7	39	21	6

Assume the data represent the general population consisting of these two groups, so that the data may be used to calculate probabilities and conditional probabilities.

Several persons are interviewed. The result of each interview is a "profile" of answers to the questions. The goal is to classify the person in one of the two subgroups

For the following profiles, classify each individual in one of the subgroups

i. y, n, y, n, y, u, n, u, y. u
ii. n, n, u, n, y, y, u, n, n, y

iii. y, y, n, y, u, u, n, n, y, y

Answer

```
% file npr05_16.m
% Data for Exercise 5.3.16.
A = [51 26  7; 42 32 10; 19 54 11; 24 53  7; 27 52  5;
     49 19 16; 16 59  9; 47 32  5; 55 17 12; 24 53  7];
B = [27 34  5; 19 43  4; 39 22  5; 38 19  9; 28 33  5;
     19 41  6; 37 21  8; 19 42  5; 27 33  6; 39 21  6];
disp('Call for oddsdf')
npr05_16
Call for oddsdf
oddsdf
Enter matrix A of frequencies for calibration group 1  A
Enter matrix B of frequencies for calibration group 2  B
Number of questions = 10
Answers per question = 3
 Enter code for answers and call for procedure "odds"
y = 1;
n = 2;
u = 3;
odds
Enter profile matrix E  [y n y n y u n u y u]
Odds favoring Group 1:   3.743
Classify in Group 1
odds
Enter profile matrix E  [n n u n y y u n n y]
Odds favoring Group 1:   0.2693
Classify in Group 2
odds
Enter profile matrix E  [y y n y u u n n y y]
Odds favoring Group 1:   5.286
Classify in Group 1
```

Exercise 5.3.17

The data of Exercise 5.3.16., above, are converted to conditional probabilities and probabilities, as follows (probabilities are rounded to two decimal places).

GROUP 1 $P(G_1) = 0.56$				GROUP 2 $P(G_2) = 0.44$		
Q	Yes	No	Unc	Yes	No	Unc
1	0.61	0.31	0.08	0.41	0.51	0.08
2	0.50	0.38	0.12	0.29	0.65	0.06
3	0.23	0.64	0.13	0.59	0.33	0.08
4	0.29	0.63	0.08	0.57	0.29	0.14
5	0.32	0.62	0.06	0.42	0.50	0.08
6	0.58	0.23	0.19	0.29	0.62	0.09
7	0.19	0.70	0.11	0.56	0.32	0.12
8	0.56	0.38	0.06	0.29	0.63	0.08

9	0.65	0.20	0.15	0.41	0.50	0.09
10	0.29	0.63	0.08	0.59	0.32	0.09

For the following profiles classify each individual in one of the subgroups.

i. y, n, y, n, y, u, n, u, y, u
ii. n, n, u, n, y, y, u, n, n, y
iii. y, y, n, y, u, u, n, n, y, y

Answer

```
npr05_17
% file npr05_17.m
% Data for Exercise 5.3.17.
PG1 = 84/150;
PG2 = 66/125;
A = [0.61 0.31 0.08
     0.50 0.38 0.12
     0.23 0.64 0.13
     0.29 0.63 0.08
     0.32 0.62 0.06
     0.58 0.23 0.19
     0.19 0.70 0.11
     0.56 0.38 0.06
     0.65 0.20 0.15
     0.29 0.63 0.08];

B = [0.41 0.51 0.08
     0.29 0.65 0.06
     0.59 0.33 0.08
     0.57 0.29 0.14
     0.42 0.50 0.08
     0.29 0.62 0.09
     0.56 0.32 0.12
     0.29 0.64 0.08
     0.41 0.50 0.09
     0.59 0.32 0.09];
disp('Call for oddsdp')
Call for oddsdp
oddsdp
Enter matrix A of conditional probabilities for Group 1  A
Enter matrix B of conditional probabilities for Group 2  B
Probability p1 an individual is from Group 1  PG1
Number of questions = 10
Answers per question = 3
 Enter code for answers and call for procedure "odds"
y = 1;
n = 2;
u = 3;
odds
Enter profile matrix E  [y n y n y u n u y u]
Odds favoring Group 1:   3.486
Classify in Group 1
odds
Enter profile matrix E  [n n u n y y u n n y]
Odds favoring Group 1:   0.2603
```

```
Classify in Group 2
odds
Enter profile matrix E  [y y n y u u n n y y]
Odds favoring Group 1:   5.162
Classify in Group 1
```

Answers

Exercise 5.3.1

$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c)$, $P(B) = P(B|C)P(C)$ + P(B|C^c) P(C^c)\), and
$P(AB) = P(A|C)P(B|C)P(C) + P(A|C^c)P(B|C^c)P(B|C^c)P(C^c)$

```
PA = 0.4*0.7 + 0.3*0.3
PA =  0.3700
PB = 0.6*0.7 + 0.2*0.3
PB =  0.4800
PA*PB
ans = 0.1776
PAB = 0.4*0.6*0.7 + 0.3*0.2*0.3
PAB = 0.1860       % PAB not equal PA*PB;  not independent
```

Exercise 5.3.2

$$\frac{P(C|A_1A_2^cA_3)}{P(C^c|A_1A_2^cA_3)} = \frac{P(C)}{P(C^c)} \cdot \frac{P(A_1C)P(A_2^c|C)P(A_3|C)}{P(A_1|C^c)P(A_2^c|C^c)P(A_3|C^c)}$$
$$= \frac{0.4}{0.6} \cdot \frac{0.9 \cdot 0.15 \cdot 0.80}{0.20 \cdot 0.85 \cdot 0.20} = \frac{108}{51} = 2.12$$

Exercise 5.3.3

```
PW = 0.01*[75 80 65 70 85];
PWc = 0.01*[60 65 50 55 70];
P = ckn(PW,3)*0.7 + ckn(PWc,3)*0.3
P =  0.8353
```

Exercise 5.3.5

```
PW = 0.1*[4 5 3 7 5 6 2];
PWc = 0.1*[7 8 5 9 7 8 5];
PA4 = ckn(PW,4)*0.8 + ckn(PWc,4)*0.2
PA4 =  0.4993
PA5 = ckn(PW,5)*0.8 + ckn(PWc,5)*0.2
PA5 =  0.2482
```

Exercise 5.3.6

Let E_1 be the event the probability is 0.80 and E_2 be the event the probability is 0.65. Assume $P(E_1)/P(E_2) = 1$.

$$\frac{P(E_1|S_n = k)}{P(E_2|S_n = k)} = \frac{P(E_1)}{P(E_2)} \cdot \frac{P(S_n = k|E_1)}{P(S_n = k|E_2)}$$

```
k = 1:20;
odds = ibinom(20,0.80,k)./ibinom(20,0.65,k);
disp([k;odds]')
- - - - - - - - - - - - -
   13.0000    0.2958
   14.0000    0.6372
   15.0000    1.3723   % Need at least 15 or 16 successes
   16.0000    2.9558
   17.0000    6.3663
   18.0000   13.7121
   19.0000   29.5337
   20.0000   63.6111
```

Exercise 5.3.9

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{P(HE_1E_2S) + P(HE_1E_2S^c)}{P(H^cE_1E_2S) + P(H^cE_1E_2S^c)}$$
$$= \frac{P(S)P(H|S)P(E_1|S)P(E_2|S) + P(S^c)P(H|S^c)P(E_1|S^c)P(E_2|S^c)}{P(S)P(H^c|S)P(E_1|S)P(E_2|S) + P(S^c)P(H^c|S^c)P(E_1|S^c)P(E_2|S^c)}$$
$$= \frac{0.80 \cdot 0.95 \cdot 0.90 \cdot 0.95 + 0.20 \cdot 0.45 \cdot 0.25 \cdot 0.20}{0.80 \cdot 0.05 \cdot 0.90 \cdot 0.95 + 0.20 \cdot 0.55 \cdot 0.25 \cdot 0.20} = 16.64811$$

Exercise 5.3.10

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{P(HE_1E_2S) + P(HE_1E_2S^c)}{P(H^cE_1E_2S) + P(H^cE_1E_2S^c)}$$
$$P(HE_1E_2S) = P(S)P(H|S)P(E_1|SH)P(E_2|SHE_1) = P(S)P(H|S)P(E_1|S)P(E_2|S)$$

with similar expressions for the other terms.

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{0.85 \cdot 0.95 \cdot 0.90 \cdot 0.95 + 0.15 \cdot 0.30 \cdot 0.25 \cdot 0.20}{0.85 \cdot 0.05 \cdot 0.90 \cdot 0.95 + 0.15 \cdot 0.70 \cdot 0.25 \cdot 0.20} = 16.6555$$

Exercise 5.3.13

```
PH = 0.01*[80 70 75];
PHc = 0.01*[85 80 70];
pH = 2/3;
P  = ckn(PH,2)*pH + ckn(PHc,2)*(1 - pH)
P =  0.8577
```

Exercise 5.3.14

$$\frac{P(H|E_1E_2)}{P(H^c|E_1E_2)} = \frac{P(HE_1E_2S) + P(HE_1E_2S^c)}{P(H^cE_1E_2S) + P(H^cE_1E_2S^c)}$$
$$= \frac{P(S)P(H|S)P(E_1|S)P(E_2|S) + P(S^c)P(H|S^c)P(E_1|S^c)P(E_2|S^c)}{P(S)P(H^c|S)P(E_1|S)P(E_2|S) + P(S^c)P(H^c|S^c)P(E_1|S^c)P(E_2|S^c)}$$
$$= \frac{0.80 \cdot 0.95 \cdot 0.90 \cdot 0.95 + 0.20 \cdot 0.30 \cdot 0.20 \cdot 0.25}{0.80 \cdot 0.05 \cdot 0.90 \cdot 0.95 + 0.20 \cdot 0.70 \cdot 0.20 \cdot 0.25} = 15.8447$$

This page titled 5.3: Problems on Conditional Independence is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

6: RANDOM VARIABLES AND PROBABILITIES

6.1: Random Variables and Probabilities
6.2: Problems on Random Variables and Probabilities

This page titled 6: Random Variables and Probabilities is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Random Variables and Probabilities

Probability associates with an event a number which indicates the likelihood of the occurrence of that event on any trial. An event is modeled as the set of those possible outcomes of an experiment which satisfy a property or proposition characterizing the event.

Often, each outcome is characterized by a number. The experiment is performed. If the outcome is observed as a physical quantity, the size of that quantity (in prescribed units) is the entity actually observed. In many nonnumerical cases, it is convenient to assign a number to each outcome. For example, in a coin flipping experiment, a "head" may be represented by a 1 and a "tail" by a 0. In a Bernoulli trial, a success may be represented by a 1 and a failure by a 0. In a sequence of trials, we may be interested in the number of successes in a sequence of n component trials. One could assign a distinct number to each card in a deck of playing cards. Observations of the result of selecting a card could be recorded in terms of individual numbers. In each case, the associated number becomes a property of the outcome.

Random variables as functions

We consider in this chapter *real* random variables (i.e., real-valued random variables). In the chapter on Random Vectors and Joint Distributions, we extend the notion to vector-valued random quantites. The fundamental idea of a real *random variable* is the assignment of a real number to each elementary outcome ω in the basic space Ω. Such an assignment amounts to determining a *function* X, whose domain is Ω and whose range is a subset of the real line **R**. Recall that a real-valued function on a domain (say an interval I on the real line) is characterized by the assignment of a real number y to each element x (argument) in the domain. For a real-valued function of a real variable, it is often possible to write a formula or otherwise state a rule describing the assignment of the value to each argument. Except in special cases, we cannot write a formula for a random variable X. However, random variables share some important general properties of functions which play an essential role in determining their usefulness.

Mappings and inverse mappings

There are various ways of characterizing a function. Probably the most useful for our purposes is as a *mapping* from the *domain* Ω to the *codomain* **R**. We find the mapping diagram of Figure 1 extremely useful in visualizing the essential patterns. Random variable X, as a mapping from basic space Ω to the real line **R**, assigns to each element ω a value $t = X(\omega)$. The object point ω is mapped, or carried, into the image point t. Each ω is mapped into exactly one t, although several ω may have the same image point.

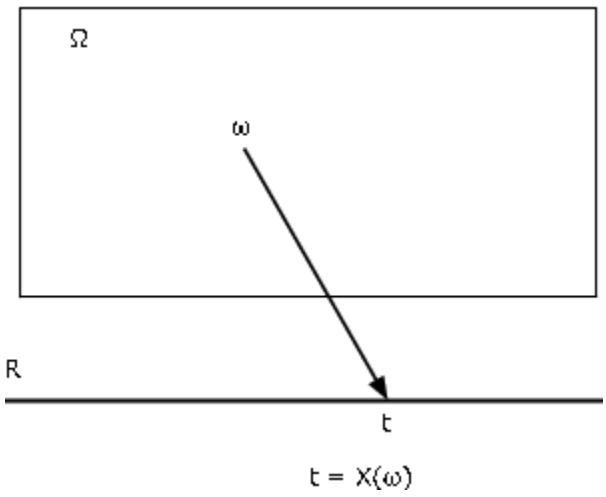


Figure 6.1.1. The basic mapping diagram $t = X(\omega)$.

Associated with a function X as a mapping are the *inverse mapping* X^{-1} and the *inverse images* it produces. Let M be a set of numbers on the real line. By the inverse image of M under the mapping X, we mean the *set* of *all* those $\omega \in \Omega$ which are mapped into M by X (see Figure 2). If X does not take a value in M, the inverse image is the empty set (impossible event). If M includes

the range of X, (the set of all possible values of X), the inverse image is the entire basic space Ω. Formally we write

$$X^{-1}(M) = \{\omega : X(\omega) \in M\}$$

Now we assume the set $X^{-1}(M)$, a subset of Ω, is an event for each M. A detailed examination of that assertion is a topic in *measure theory*. Fortunately, the results of measure theory ensure that we may make the assumption for any X and any subset M of the real line likely to be encountered in practice. The set $X^{-1}(M)$ is the event that X takes a value in M. As an event, it may be assigned a probability.

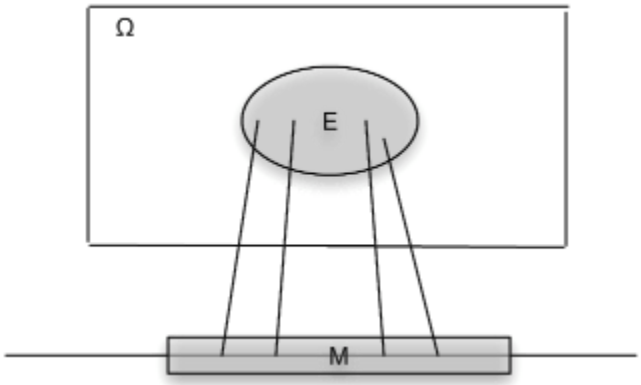


Figure 6.1.2. *E* is the inverse image $X^{-1}(M)$.

Example 6.1.1 Some illustrative examples

1. I_E where E is an event with probability p. Now X takes on only two values, 0 and 1. The event that X take on the value 1 is the set

$$\{\omega : X(\omega) = 1\} = X^{-1}(\{1\}) = E$$

so that $P(\{\omega : X(\omega) = 1\}) = p$. This rather ungainly notation is shortened to $P(X = 1) = p$. Similarly, . Consider any set M. If neither 1 nor 0 is in M, then $X^{-1}(M) = \emptyset$ If 0 is in M, but 1 is not, then $X^{-1}(M) = E^c$ If 1 is in M, but 0 is not, then $X^{-1}(M) = E$ If both 1 and 0 are in M, then $X^{-1}(M) = \Omega$ In this case the class of all events $X^{-1}(M)$ consists of event E, its complement E^c, the impossible event $\emptyset$, and the sure event Ω.

2. Consider a sequence of n Bernoulli trials, with probability p of success. Let S_n be the random variable whose value is the number of successes in the sequence of n component trials. Then, according to the analysis in the section "Bernoulli Trials and the Binomial Distribution"

$$P(S_n = k) = C(n, k)p^k(1 - p)^{n-k} \; 0 \le k \le n$$

Before considering further examples, we note a general property of inverse images. We state it in terms of a random variable, which maps Ω to the real line (see Figure 3).

Preservation of set operations

Let X be a mapping from Ω to the real line **R**. If $M, M_i, i \in J$ are sets of real numbers, with respective inverse images E, E_i, then

$$X^{-1}(M^c) = E^c,\; X^{-1}(\bigcup_{i\in J} M_i) = \bigcup_{i\in J} E_i \text{ and } X^{-1}(\bigcap_{i\in J} M_i) = \bigcap_{i\in J} E_i$$

Examination of simple graphical examples exhibits the plausibility of these patterns. Formal proofs amount to careful reading of the notation. Central to the structure are the facts that each element ω is mapped into only one image point t and that the inverse image of M is the set of *all* those ω which are mapped into image points in M.

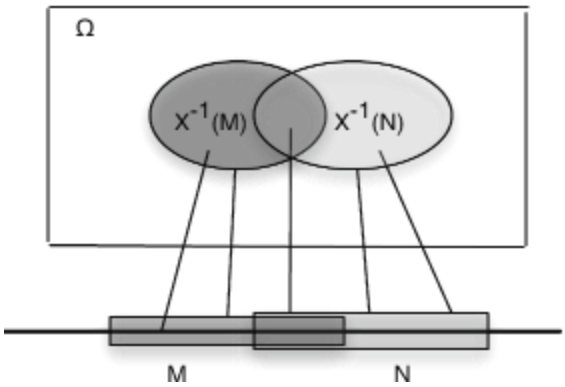


Figure 6.1.3. Preservation of set operations by inverse images.

An easy, but important, consequence of the general patterns is that the inverse images of disjoint M, N are also disjoint. This implies that the inverse of a disjoint union of M_i is a disjoint union of the separate inverse images.

Example 6.1.2 Events determined by a random variable

Consider, again, the random variable S_n which counts the number of successes in a sequence of n Bernoulli trials. Let $n = 10$ and $p = 0.33$. Suppose we want to determine the probability $P(2 < S_{10} \leq 8)$. Let $A_k = \{\omega : S_{10}(\omega) = k\}$, which we usually shorten to $A_k = \{S_{10} = k\}$. Now the A_k form a partition, since we cannot have $\omega \in A_k$ and $\omega \in A_k$ $j \neq k$ (i.e., for any ω, we cannot have two values for $S_n(\omega)$). Now,

$$\{2 < S_{10} \leq 8\} = A_3 \bigvee A_4 \bigvee A_5 \bigvee A_6 \bigvee A_7 \bigvee A_8$$

since S_{10} takes on a value greater than 2 but no greater than 8 iff it takes one of the integer values from 3 to 8. By the additivity of probability,

Mass transfer and induced probability distribution

Because of the abstract nature of the basic space and the class of events, we are limited in the kinds of calculations that can be performed meaningfully with the probabilities on the basic space. We represent probability as mass distributed on the basic space and visualize this with the aid of general Venn diagrams and minterm maps. We now think of the mapping from Ω to **R** as a producing a point-by-point *transfer of the probability mass* to the real line. This may be done as follows:

To any set M on the real line assign probability mass $P_X(M) = P(X^{-1}(M))$

It is apparent that $P_X(M) \geq 0$ and $P_X(\mathbf{R}) = P(\Omega) = 1$. And because of the preservation of set operations by the inverse mapping

$$P_X(\bigvee_{i=1}^{\infty} M_i) = P(X^{-1}(\bigvee_{i=1}^{\infty} M_i)) = P(\bigvee_{i=1}^{\infty} X^{-1}(M_i)) = \sum_{i=1}^{\infty} P(X^{-1}(M_i)) = \sum_{i=1}^{\infty} P_X(M_i)$$

This means that P_X has the properties of a probability measure defined on the subsets of the real line. Some results of measure theory show that this probability is defined uniquely on a class of subsets of **R** that includes any set normally encountered in applications. We have achieved a point-by-point transfer of the probability apparatus to the real line in such a manner that we can make calculations about the random variable X. We call P_X the *probability measure induced by X*. Its importance lies in the fact that $P(X \in M) = P_X(M)$. Thus, *to determine the likelihood that random quantity X will take on a value in set M, we determine how much induced probability mass is in the set M*. This transfer produces what is called the *probability distribution* for *X*. In the chapter "Distribution and Density Functions", we consider useful ways to describe the probability distribution induced by a random variable. We turn first to a special class of random variables.

Simple random variables

We consider, in some detail, random variables which have only a finite set of possible values. These are called *simple* random variables. Thus the term “simple” is used in a special, technical sense. The importance of simple random variables rests on two

facts. For one thing, in practice we can distinguish only a finite set of possible values for any random variable. In addition, any random variable may be approximated as closely as pleased by a simple random variable. When the structure and properties of simple random variables have been examined, we turn to more general cases. Many properties of simple random variables extend to the general case via the approximation procedure.

Representation with the aid of indicator functions

In order to deal with simple random variables clearly and precisely, we must find suitable ways to express them analytically. We do this with the aid of indicator functions. Three basic forms of representation are encountered. These are not mutually exclusive representatons.

Standard or *canonical form*, which displays the possible values and the corresponding events. If *X* takes on distinct values

$\{t_1, t_2, \cdots, t_n\}$ with respective probabilities $\{p_1, p_2, \cdots, p_n\}$

and if $A_i = \{X = t_i\}$, for $1 \le i \le n$, then $\{A_1, A_2, \cdots, A_n\}$ is a partition (i.e., on any trial, exactly one of these events occurs). We call this the *partition determined by* (or, generated by) *X*. We may write

$$X = t_1 I_{A_1} + t_2 I_{A_2} + \cdots + t_n I_{A_n} = \sum_{i=1}^{n} t_i I_{A_i}$$

If $X(\omega) = t_i$, then $\omega \in A_i$, so that $I_{A_i}(\omega) = 1$ and all the other indicator functions have value zero. The summation expression thus picks out the correct value t_i. This is true for any t_i, so the expression represents $X(\omega)$for all ω. The distinct set $\{A, B, C\}$ of the values and the corresponding probabilities $\{p_1, p_2, \cdots, p_n\}$ constitute the *distribution* for *X*. Probability calculations for *X* are made in terms of its distribution. One of the advantages of the canonical form is that it displays the range (set of values), and if the probabilities $\{A, B, C, D\}$ are known, the distribution is determined. *Note* that in canonical form, if one of the t_i has value zero, we include that term. For some probability distributions it may be that $P(A_i) = 0$ for one or more of the t_i. In that case, we call these values *null values*, for they can only occur with probability zero, and hence are practically impossible. In the general formulation, we include possible null values, since they do not affect any probabilitiy calculations.

Example 6.1.3 Successes in Bernoulli trials

As the analysis of Bernoulli trials and the binomial distribution shows (see Section 4.8), canonical form must be

$$S_n = \sum_{k=0}^{n} k I_{A_k} \text{ with } P(A_k) = C(n,k) p^k (1-p)^{n-k},\ 0 \le k \le n$$

For many purposes, both theoretical and practical, canonical form is desirable. For one thing, it displays directly the range (i.e., set of values) of the random variable. The *distribution* consists of the set of values $\{t_k : 1 \le k \le n\}$ paired with the corresponding set of probabilities $\{p_k : 1 \le k \le n\}$, where $p_k = P(A_k) = P(X = t_k)$.

Simple random variable *X* may be represented by a *primitive form*

$$X = c_1 I_{C_1} + c_2 I_{C_2} + \cdots, c_m I_{C_m}, \text{ where } \{C_j : 1 \le j \le m\} \text{ is a partition}$$

Remarks

- If $\{C_j : 1 \le j \le m\}$ is a disjoint class, but $\bigcup_{j=1}^{m} C_j \ne \Omega$, we may append the event $C_{m+1} = [\bigcup_{j=1}^{m} C_j]^c$ and assign value zero to it.
- We say *a* primitive form, since the representation is not unique. Any of the C_j may be partitioned, with the same value c_i associated with each subset formed.
- Canonical form is a special primitive form. Canonical form is unique, and in many ways normative.

Example 6.1.4 Simple random variables in primitive form

- A wheel is spun yielding, on a equally likely basis, the integers 1 through 10. Let C_i be the event the wheel stops at i, $1 \le i \le 10$. Each $P(C_i) = 0.1$. If the numbers 1, 4, or 7 turn up, the player loses ten dollars; if the numbers 2, 5, or 8 turn up, the player gains nothing; if the numbers 3, 6, or 9 turn up, the player gains ten dollars; if the number 10 turns up, the player loses one dollar. The random variable expressing the results may be expressed in primitive form as

$$X = -10I_{C_1} + 0I_{C_2} + 10I_{C_3} - 10I_{C_4} + 0I_{C_5} + 10I_{C_6} - 10I_{C_7} + 0I_{C_8} + 10I_{C_9} - I_{C_{10}}$$

- A store has eight items for sale. The prices are \$3.50, \$5.00, \$3.50, \$7.50, \$5.00, \$5.00, \$3.50, and \$7.50, respectively. A customer comes in. She purchases one of the items with probabilities 0.10, 0.15, 0.15, 0.20, 0.10 0.05, 0.10 0.15. The random variable expressing the amount of her purchase may be written

$$X = 3.5I_{C_1} + 5.0I_{C_2} + 3.5I_{C_3} + 7.5I_{C_4} + 5.0I_{C_5} + 5.0I_{C_6} + 3.5I_{C_7} + 7.5I_{C_8}$$

We commonly have *X* represented in *affine form*, in which the random variable is represented as an affine combination of indicator functions (i.e., a linear combination of the indicator functions plus a constant, which may be zero).

$$X = c_0 + c_1 I_{E_1} + c_2 I_{E_2} + \cdots + c_m I_{E_m} = c_0 + \sum_{j=1}^{m} c_j I_{E_j}$$

In this form, the class $\{E_1, E_2, \cdots, E_m\}$ is not necessarily mutually exclusive, and the coefficients do not display directly the set of possible values. In fact, the E_i often form an independent class. *Remark*. Any primitive form is a special affine form in which $c_0 = 0$ and the E_i form a partition.

Example 6.1.5

Consider, again, the random variable S_n which counts the number of successes in a sequence of n Bernoulli trials. If E_i is the event of a success on the ith trial, then one natural way to express the count is

$$S_n = \sum_{i=1}^{n} I_{E_i}, \text{ with } P(E_i) = p \; 1 \le i \le n$$

This is affine form, with $c_0 = 0$ and $c_i = 1$ for $1 \le i \le n$. In this case, the E_i cannot form a mutually exclusive class, since they form an independent class.

Events generated by a simple random variable: canonical form

We may characterize the class of all inverse images formed by a simple random X in terms of the partition it determines. Consider any set M of real numbers. If t_i in the range of X is in M, then every point $\omega \in A_i$ maps into t_i, hence into M. If the set J is the set of indices i such that $t_i \in M$, then

Only those points ω in $A_M = \bigvee_{i \in J} A_i$ map into M.

Hence, the class of events (i.e., inverse images) determined by X consists of the impossible event $\emptyset$, the sure event Ω, and the union of any subclass of the A_i in the partition determined by X.

Example 6.1.6 Events determined by a simple random variable

Suppose simple random variable X is represented in canonical form by

$X = -2I_A - I_B + 0I_C + 3I_D$

Then the class $\{A, B, C, D\}$ is the partition determined by X and the range of X is $\{-2, -1, 0, 3\}$.

1. If M is the interval [-2, 1], the the values -2, -1, and 0 are in M and $X^{-1}(M) = A \bigvee B \bigvee C$.
2. If M is the set (-2, -1] ∪ [1, 5], then the values -1, 3 are in M and $X^{-1}(M) = B \bigvee D$.
3. The event $\{X \le 1\} = \{X \in (-\infty, 1]\} = X^{-1}(M)$, where $M = (-\infty, 1]$. Since values -2, -1, 0 are in M, the event $\{X \le 1\} = A \bigvee B \bigvee C$.

Determination of the distribution

Determining the partition generated by a simple random variable amounts to determining the canonical form. The distribution is then completed by determining the probabilities of each event $A_k = \{X = t_k\}$.

From a primitive form

Before writing down the general pattern, we consider an illustrative example.

Example 6.1.7 The distribution from a primitive form

Suppose one item is selected at random from a group of ten items. The values (in dollars) and respective probabilities are

c_j	2.00	1.50	2.00	2.50	1.50	1.50	1.00	2.50	2.00	1.50
$P(C_j)$	0.08	0.11	0.07	0.15	0.10	0.09	0.14	0.08	0.08	0.10

By inspection, we find four distinct values: $t_1 = 1.00$, $t_2 = 1.50$, $t_3 = 2.00$, and $t_4 = 2.50$. The value 1.00 is taken on for $\omega \in C_7$, so that $A_1 = C_7$ and $P(A_1) = P(C_7) = 0.14$. Value 1.50 is taken on for $\omega \in C_2, C_5, C_6, C_{10}$ so that

$$A_2 = C_2 \bigvee C_5 \bigvee C_6 \bigvee C_{10} \text{ and } P(A_2) = P(C_2) + P(C_5) + P(C_6) + P(C_{10}) = 0.40$$

Similarly

$$P(A_3) = P(C_1) + P(C_3) + P(C_9) = 0.23 \text{ and } P(A_4) = P(C_4) + P(C_8) = 0.25$$

The distribution for *X* is thus

k	1.00	1.50	2.00	2.50
$P(X = k)$	0.14	0.40	0.23	0.23

The general procedure may be formulated as follows:

If $X = \sum_{j=1}^{m} c_j I_{c_j}$, we identify the set of distinct values in the set $\{c_j : 1 \le j \le m\}$. Suppose these are $t_1 < t_2 < \cdots < t_n$. For any possible value t_i in the range, identify the index set J_i of those j such that $c_j = t_i$ Then the terms

$$\sum_{J_i} c_j I_{c_j} = t_i \sum_{J_i} I_{c_j} = t_i I_{A_i}, \text{ where } A_i = \bigvee_j \in J_i C_j,$$

and

$$P(A_i) = P(X = t_i) = \sum_{j \in J} P(C_j)$$

Examination of this procedure shows that there are two phases:

- Select and sort the distinct values $t_1, t_2, \cdots, t_n$
- Add all probabilities associated with each value t_i to determine $P(X = t_i)$

We use the m-function *csort* which performs these two operations (see Example 4 from "Minterms and MATLAB Calculations").

Example 6.1.8 Use of csort on Example 6.1.7

```
>> C = [2.00 1.50 2.00 2.50 1.50 1.50 1.00 2.50 2.00 1.50];  % Matrix of c_j
>> pc = [0.08 0.11 0.07 0.15 0.10 0.09 0.14 0.08 0.08 0.10]; % Matrix of P(C_j)
>> [X,PX] = csort(C,pc);     % The sorting and consolidating operation
>> disp([X;PX]')             % Display of results
    1.0000    0.1400
    1.5000    0.4000
```

```
    2.0000    0.2300
    2.5000    0.2300
```

For a problem this small, use of a tool such as csort is not really needed. But in many problems with large sets of data the m-function csort is very useful.

From affine form

Suppose X is in affine form,

$$X = c_0 + c_1 I_{E_1} + c_2 I_{E_2} + \cdots + c_m I_{E_m} = c_0 + \sum_{j=1}^{m} c_j I_{E_j}$$

We determine a particular primitive form by determining the value of X on each minterm generated by the class $\{E_j : 1 \le j \le m\}$. We do this in a systematic way by utilizing minterm vectors and properties of indicator functions.

X is constant on each minterm generated by the class $\{E_1, E_2, \cdots, E_m\}$ since, as noted in the treatment of the minterm expansion, each indicator function I_{E_i} is constant on each minterm. We determine the value s_i of X on each minterm M_i. This describes X in a special primitive form

$$X = \sum_{k=0}^{2^m-1} s_i I_{M_i}, \text{ with } P(M_i) = p_i, 0 \le i \le 2^m - 1$$

We apply the csort operation to the matrices of values and minterm probabilities to determine the distribution for X.

We illustrate with a simple example. Extension to the general case should be quite evident. First, we do the problem “by hand” in tabular form. Then we use the m-procedures to carry out the desired operations.

Example 6.1.9 Finding the distribution from affine form

A mail order house is featuring three items (limit one of each kind per customer). Let

- E_1 = the event the customer orders item 1, at a price of 10 dollars.
- E_2 = the event the customer orders item 2, at a price of 18 dollars.
- E_3 = the event the customer orders item 3, at a price of 10 dollars.

There is a mailing charge of 3 dollars per order.

We suppose $\{E_1, E_2, E_3\}$ is independent with probabilities 0.6, 0.3, 0.5, respectively. Let X be the amount a customer who orders the special items spends on them plus mailing cost. Then, in affine form,

$$X = 10 I_{E_1} + 18 I_{E_2} + 10 I_{E_3} + 3$$

We seek first the primitive form, using the minterm probabilities, which may calculated in this case by using the m-function minprob.

1. To obtain the value of X on each minterm we
   - Multiply the minterm vector for each generating event by the coefficient for that event
   - Sum the values on each minterm and add the constant

   To complete the table, list the corresponding minterm probabilities.

i	10 I_{E_1}	18 I_{E_2}	10 I_{E_3}	c	$s - i$	pm_i
0	0	0	0	3	3	0.14
1	0	0	10	3	13	0.14
2	0	18	0	3	21	0.06

3	0	18	10	3	31	0.06
4	10	0	0	3	13	0.21
5	10	0	10	3	23	0.21
6	10	18	0	3	31	0.09
7	10	18	10	3	41	0.09

We then sort on the s_i, the values on the various M_i, to expose more clearly the primitive form for X.

"Primitive form" Values

i	s_i	pm_i
0	3	0.14
1	13	0.14
4	13	0.21
2	21	0.06
5	23	0.21
3	31	0.06
6	31	0.09
7	41	0.09

The primitive form of X is thus

\(X = 3I_{M_0} + 12I_{M_1} + 13I_{M_4} + 21I_{M_2} + 23I_{M_5} + 31I_{M_3} + 31I_{M_6} + 41I_{M_7}

We note that the value 13 is taken on on minterms M_1 and M_4. The probability X has the value 13 is thus $p(1) + p(4)$. Similarly, X has value 31 on minterms M_3 and M_6.

- To complete the process of determining the distribution, we list the sorted values and consolidate by adding together the probabilities of the minterms on which each value is taken, as follows:

k	t_k	p_k
1	3	0.14
2	13	0.14 + 0.21 = 0.35
3	21	0.06
4	23	0.21
5	31	0.06 + 0.09 = 0.15
6	41	0.09

The results may be put in a matrix X of possible values and a corresponding matrix **PX** of probabilities that X takes on each of these values. Examination of the table shows that

$$X = [3\ 13\ 21\ 23\ 31\ 41] \text{ and } PX = [0.14\ 0.35\ 0.06\ 0.21\ 0.15\ 0.09]$$

Matrices X and **PX** describe the distribution for X.

An m-procedure for determining the distribution from affine form

We now consider suitable MATLAB steps in determining the distribution from affine form, then incorporate these in the m-

procedure *canonic* for carrying out the transformation. We start with the random variable in affine form, and suppose we have available, or can calculate, the minterm probabilities.

The procedure uses *mintable* to set the basic minterm vector patterns, then uses a matrix of coefficients, including the constant term (set to zero if absent), to obtain the values on each minterm. The minterm probabilities are included in a row matrix.
Having obtained the values on each minterm, the procedure performs the desired consolidation by using the m-function csort.

Example 6.1.10 Steps in determining the distribution for *X* in Example 6.1.9

```
>> c = [10 18 10 3];                % Constant term is listed last
>> pm = minprob(0.1*[6 3 5]);
>> M  = mintable(3)                 % Minterm vector pattern
M =
     0     0     0     0     1     1     1     1
     0     0     1     1     0     0     1     1
     0     1     0     1     0     1     0     1
% - - - - - - - - - - - - - - -       % An approach mimicking ``hand'' calculation
>> C = colcopy(c(1:3),8)            % Coefficients in position
C =
    10    10    10    10    10    10    10    10
    18    18    18    18    18    18    18    18
    10    10    10    10    10    10    10    10
>> CM = C.*M                        % Minterm vector values
CM =
     0     0     0     0    10    10    10    10
     0     0    18    18     0     0    18    18
     0    10     0    10     0    10     0    10
>> cM = sum(CM) + c(4)              % Values on minterms
cM =
     3    13    21    31    13    23    31    41
% - - - - - - - - - - - -  -         % Practical MATLAB procedure
>> s = c(1:3)*M + c(4)
s =
     3    13    21    31    13    23    31    41
>> pm = 0.14  0.14  0.06  0.06  0.21  0.21  0.09  0.09   % Extra zeros deleted
>> const = c(4)*ones(1,8);}
```

```
>> disp([CM;const;s;pm]')           % Display of primitive form
     0     0     0   3    3    0.14  % MATLAB gives four decimals
     0     0    10   3   13    0.14
     0    18     0   3   21    0.06
     0    18    10   3   31    0.06
    10     0     0   3   13    0.21
    10     0    10   3   23    0.21
    10    18     0   3   31    0.09
    10    18    10   3   41    0.09
>> [X,PX] = csort(s,pm);             % Sorting on s, consolidation of  pm
>> disp([X;PX]')                    % Display of final result
     3    0.14
    13    0.35
    21    0.06
    23    0.21
    31    0.15
    41    0.09
```

The two basic steps are combined in the m-procedure *canonic*, which we use to solve the previous problem.

Example 6.1.11 Use of canonic on the variables of Example 6.1.10

```
>> c = [10 18 10 3]; % Note that the constant term 3 must be included last
>> pm = minprob([0.6 0.3 0.5]);
>> canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
>> disp(XDBN)
    3.0000    0.1400
   13.0000    0.3500
   21.0000    0.0600
   23.0000    0.2100
   31.0000    0.1500
   41.0000    0.0900
```

With the distribution available in the matrices X (set of values) and **PX** (set of probabilities), we may calculate a wide variety of quantities associated with the random variable.

We use two key devices:

1. Use relational and logical operations on the matrix of values X to determine a matrix M which has ones for those values which meet a prescribed condition. $P(X \in M)$: PM = M*PX'
2. Determine $G = g(X) = [g(X_1)g(X_2) \cdots g(X_n)]$ by using array operations on matrix X. We have two alternatives:
   a. Use the matrix G, which has values $g(t_i)$ for each possible value t_i for X, or,
   b. Apply csort to the pair (G, PX) to get the distribution for $Z = g(X)$. This distribution (in value and probability matrices) may be used in exactly the same manner as that for the original random variable X.

Example 6.1.12 Continuation of Example 6.1.11

Suppose for the random variable X in Example 6.11 it is desired to determine the probabilities

$P(15 \leq X \leq 35)$, $P(|X - 20| \leq 7)$, and $(X - 10)(X - 25) > 0)$

```
>> M = (X>=15)&(X<=35);
M = 0   0    1    1    1    0    % Ones for minterms on which 15 <= X <= 35
>> PM = M*PX'                     % Picks out and sums those minterm probs
PM =  0.4200
>> N = abs(X-20)<=7;
N = 0    1    1    1    0    0   % Ones for minterms on which |X - 20| <= 7
>> PN = N*PX'                     % Picks out and sums those minterm probs
PN =  0.6200
>> G = (X - 10).*(X - 25)
G = 154 -36 -44 -26 126 496       % Value of g(t_i) for each possible value
>> P1 = (G>0)*PX'                 % Total probability for those t_i such that
P1 =  0.3800                      % g(t_i) > 0
>> [Z,PZ] = csort(G,PX)           % Distribution for Z = g(X)
Z =  -44   -36   -26   126   154   496
PZ =  0.0600    0.3500    0.2100    0.1500    0.1400    0.0900
>> P2 = (Z>0)*PZ'                 % Calculation using distribution for Z
P2 =  0.3800
```

Example 6.1.13 Alternate formulation of Example 4.3.3 from "Composite Trials"

Ten race cars are involved in time trials to determine pole positions for an upcoming race. To qualify, they must post an average speed of 125 mph or more on a trial run. Let E_i be the event the ith car makes qualifying speed. It seems reasonable to suppose the class $\{E_i : 1 \le i \le 10\}$ is independent. If the respective probabilities for success are 0.90, 0.88, 0.93, 0.77, 0.85, 0.96, 0.72, 0.83, 0.91, 0.84, what is the probability that k or more will qualify (k = 6,7,8,9,10)?

Solution

Let $X = \sum_{i=1}^{10} I_{E_i}$

```
>> c = [ones(1,10) 0];
>> P = [0.90, 0.88, 0.93, 0.77, 0.85, 0.96, 0.72, 0.83, 0.91, 0.84];
>> canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  minprob(P)
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
>> k = 6:10;
>> for i = 1:length(k)
    Pk(i) = (X>=k(i))*PX';
end
>> disp(Pk)
    0.9938    0.9628    0.8472    0.5756    0.2114
```

This solution is not as convenient to write out. However, with the distribution for X as defined, a great many other probabilities can be determined. This is particularly the case when it is desired to compare the results of two independent races or “heats.” We consider such problems in the study of Independent Classes of Random Variables.

A function form for canonic

One disadvantage of the procedure canonic is that it always names the output X and **PX**. While these can easily be renamed, frequently it is desirable to use some other name for the random variable from the start. A function form, which we call **canonicf**, is useful in this case.

Example 6.1.14 Alternate solution of Example 6.1.13, using canonicf

```
>> c = [10 18 10 3];
>> pm = minprob(0.1*[6 3 5]);
>> [Z,PZ] = canonicf(c,pm);
>> disp([Z;PZ]')                % Numbers as before, but the distribution
    3.0000    0.1400            % matrices are now named Z and PZ
   13.0000    0.3500
   21.0000    0.0600
   23.0000    0.2100
   31.0000    0.1500
   41.0000    0.0900
```

General random variables

The distribution for a simple random variable is easily visualized as point mass concentrations at the various values in the range, and the class of events determined by a simple random variable is described in terms of the partition generated by X (i.e., the class of those events of the form $A_i = [X = t_i]$ for each t_i in the range). The situation is conceptually the same for the general case, but the details are more complicated. If the random variable takes on a continuum of values, then the probability mass

distribution may be spread smoothly on the line. Or, the distribution may be a mixture of point mass concentrations and smooth distributions on some intervals. The class of events determined by X is the set of all inverse images $X^{-1}(M)$ for M any member of a general class of subsets of subsets of the real line known in the mathematical literature as the *Borel sets*. There are technical mathematical reasons for not saying *M* is *any* subset, but the class of Borel sets is general enough to include any set likely to be encountered in applications—certainly at the level of this treatment. The Borel sets include any interval and any set that can be formed by complements, countable unions, and countable intersections of Borel sets. This is a type of class known as a *sigma algebra* of events. Because of the preservation of set operations by the inverse image, the class of events determined by random variable X is also a sigma algebra, and is often designated $\sigma(X)$. There are some technical questions concerning the probability measure P_X induced by X, hence the distribution. These also are settled in such a manner that there is no need for concern at this level of analysis. However, some of these questions become important in dealing with random processes and other advanced notions increasingly used in applications. Two facts provide the freedom we need to proceed with little concern for the technical details.

> $X^{-1}(M)$ is an event for every Borel set M iff for every semi-infinite interval $(-\infty, t]$ on the real line $X^{-1}((-\infty, t])$ is an event.
> The induced probability distribution is determined uniquely by its assignment to all intervals of the form $(-\infty, t]$.

These facts point to the importance of the distribution function introduced in the next chapter.

Another fact, alluded to above and discussed in some detail in the next chapter, is that any general random variable can be approximated as closely as pleased by a simple random variable. We turn in the next chapter to a description of certain commonly encountered probability distributions and ways to describe them analytically.

This page titled 6.1: Random Variables and Probabilities is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Problems on Random Variables and Probabilities

Exercise 6.2.1

The following simple random variable is in canonical form:

$$X = -3.75I_A - 1.13I_B + 0I_C + 2.6I_D.$$

Express the events $\{X \in (-4, 2]\}$, $\{X \in (0, 3]\}$, $\{X \in (-\infty, 1]\}$, and $\{X \geq 0\}$ in terms of A, B, C, and D.

Answer available at the end of this section.

Exercise 6.2.2

Random variable X, in canonical form, is given by $X = -2I_A - I_B + I_C + 2I_D + 5I_E$.

Express the events $\{X \in [2, 3)\}$, $\{X \leq 0\}$, $\{X < 0\}$, $\{|X - 2| \leq 3\}$, and $\{X^2 \geq 4\}$, in terms of $A, B, C, D, and E$.

Answer available at the end of this section.

Exercise 6.2.3

The class $\{C_j : 1 \leq j \leq 10\}$ is a partition. Random variable X has values {1, 3, 2, 3, 4, 2, 1, 3, 5, 2} on C_1 through C_{10}, respectively. Express X\) in canonical form.

Answer available at the end of this section.

Exercise 6.2.4

The class $\{C_j : 1 \leq j \leq 10\}$ in Exercise has respective probabilities 0.08, 0.13, 0.06, 0.09, 0.14, 0.11, 0.12, 0.07, 0.11, 0.09. Determine the distribution for X

Answer available at the end of this section.

Exercise 6.2.5

A wheel is spun yielding on an equally likely basis the integers 1 through 10. Let C_i be the event the wheel stops at i, $1 \leq i \leq 10$. Each $P(C_i) = 0.1$. If the numbers 1, 4, or 7 turn up, the player loses ten dollars; if the numbers 2, 5, or 8 turn up, the player gains nothing; if the numbers 3, 6, or 9 turn up, the player gains ten dollars; if the number 10 turns up, the player loses one dollar. The random variable expressing the results may be expressed in primitive form as

$$X = -10I_{C_1} + 0I_{C_2} + 10I_{C_3} - 10I_{C_4} + 0I_{C_5} + 10I_{C_6} - 10I_{C_7} + 0I_{C_8} + 10I_{C_9} - I_{C_{10}}$$

- Determine the distribution for X, (a) by hand, (b) using MATLAB.
- Determine $P(X < 0)$, $P(X > 0)$.

Answer

```
01 p = 0.1*ones(1,10);
```

```
c = [-10 0 10 -10 0 10 -10 0 10 -1];
[X,PX] = csort(c,p);
disp([X;PX]')
  -10.0000    0.3000
   -1.0000    0.1000
         0    0.3000
   10.0000    0.3000
Pneg = (X<0)*PX'
Pneg =  0.4000
Ppos = (X>0)*PX'
Ppos =  0.300
```

Exercise 6.2.6

A store has eight items for sale. The prices are \$3.50, \$5.00, \$3.50, \$7.50, \$5.00, \$5.00, \$3.50, and \$7.50, respectively. A customer comes in. She purchases one of the items with probabilities 0.10, 0.15, 0.15, 0.20, 0.10 0.05, 0.10 0.15. The random variable expressing the amount of her purchase may be written

$$X = 3.5I_{C_1} + 5.0I_{C_2} + 3.5I_{C_3} + 7.5I_{C_4} + 5.0I_{C_5} + 5.0I_{C_6} + 3.5I_{C_7} + 7.5I_{C_8}$$

Determine the distribution for X (a) by hand, (b) using MATLAB.

Answer

```
p = 0.01*[10 15 15 20 10  5 10 15];
c = [3.5 5 3.5 7.5 5 5 3.5 7.5];
[X,PX] = csort(c,p);
disp([X;PX]')
    3.5000    0.3500
    5.0000    0.3000
    7.5000    0.3500
```

Exercise 6.2.7

Suppose X, Y in canonical form are

$X = 2I_{A_1} + 3I_{A_2} + 5I_{A_3}$ $Y = I_{B_1} + 2I_{B_2} + 3I_{B_3}$

The $P(A_i)$ are 0.3, 0.6, 0.1, respectively, and the $P(B_j)$ are 0.2 0.6 0.2. Each pair $\{A_i, B_j\}$ is independent. Consider the random variable $Z = X + Y$. Then $Z = 2 + 1$ on A_1B_1, $Z = 3 + 3$ on A_2B_3, etc. Determine the value of Z on each A_iB_j and determine the corresponding $P(A_iB_j)$. From this, determine the distribution for Z.

Answer available at the end of this section.

Exercise 6.2.8

For the random variables in Exercise, let $W = XY$. Determine the value of W on each A_iB_j and determine the distribution of W.

Answer available at the end of this section.

Exercise 6.2.9

A pair of dice is rolled.

a. Let X be the minimum of the two numbers which turn up. Determine the distribution for X
b. Let Y be the maximum of the two numbers. Determine the distribution for Y.
c. Let Z be the sum of the two numbers. Determine the distribution for Z.
d. Let W be the absolute value of the difference. Determine its distribution.

Answer

```
t = 1:6;
c = ones(6,6);
[x,y] = meshgrid(t,t)
x =  1     2     3     4     5     6     % x-values in each position
     1     2     3     4     5     6
     1     2     3     4     5     6
     1     2     3     4     5     6
     1     2     3     4     5     6
     1     2     3     4     5     6
y =  1     1     1     1     1     1     % y-values in each position
     2     2     2     2     2     2
     3     3     3     3     3     3
     4     4     4     4     4     4
     5     5     5     5     5     5
     6     6     6     6     6     6
m = min(x,y);                           % min in each position
M = max(x,y);                           % max in each position
s = x + y;                              % sum x+y in each position
d = abs(x - y);                         % |x - y| in each position
[X,fX] = csort(m,c)                     % sorts values and counts occurrences
X =   1     2     3     4     5     6
fX = 11     9     7     5     3     1    % PX = fX/36
[Y,fY] = csort(M,c)
Y =   1     2     3     4     5     6
fY =  1     3     5     7     9    11    % PY = fY/36
[Z,fZ] = csort(s,c)
Z =   2     3     4     5     6     7     8     9    10    11    12
fZ =  1     2     3     4     5     6     5     4     3     2     1  %PZ = fZ/36
[W,fW] = csort(d,c)
W =   0     1     2     3     4     5
fW =  6    10     8     6     4     2    % PW = fW/36
```

Exercise 6.2.10

Minterm probabilities $p(0)$ through $p(15)$ for the class $\{A, B, C, D\}$ are, in order,

0.072 0.048 0.018 0.012 0.168 0.112 0.042 0.028 0.062 0.048 0.028 0.010 0.170 0.110 0.040 0.

Determine the distribution for random variable

$$X = -5.3I_A - 2.5I_B + 2.3I_C + 4.2I_D - 3.7$$

Answer available at the end of this section.

Exercise 6.2.11

On a Tuesday evening, the Houston Rockets, the Orlando Magic, and the Chicago Bulls all have games (but not with one another). Let *A* be the event the Rockets win, B be the event the Magic win, and C be the event the Bulls win. Suppose the class $\{A, B, C\}$ is independent, with respective probabilities 0.75, 0.70 0.8. Ellen's boyfriend is a rabid Rockets fan, who does not like the Magic. He wants to bet on the games. She decides to take him up on his bets as follows:

- \$10 to 5 on the Rockets --- i.e. She loses five if the Rockets win and gains ten if they lose
- \$10 to 5 against the Magic
- even \$5 to 5 on the Bulls.

Ellen's winning may be expressed as the random variable

$$X = -5I_A + 10I_{A^c} + 10I_B - 5I_{B^c} - 5I_C + 5I_{C^c} = -15I_A + 15I_B - 10I_C + 10$$

Determine the distribution for X. What are the probabilities Ellen loses money, breaks even, or comes out ahead?

Answer

```
P = 0.01*[75 70 80];
c = [-15 15 -10 10];
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  minprob(P)
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
disp(XDBN)
  -15.0000    0.1800
   -5.0000    0.0450
         0    0.4800
   10.0000    0.1200
   15.0000    0.1400
   25.0000    0.0350
PXneg = (X<0)*PX'
PXneg =  0.2250
PX0 = (X==0)*PX'
PX0 =    0.4800
PXpos = (X>0)*PX'
PXpos =  0.2950
```

Exercise 6.2.12

The class $\{A, B, C, D\}$ has minterm probabilities

$$pm = 0.001 * [5\ 7\ 6\ 8\ 9\ 14\ 22\ 33\ 21\ 32\ 50\ 75\ 86\ 129\ 201\ 302]$$

- Determine whether or not the class is independent.
- The random variable $X = I_A + I_B + I_C + I_D$counts the number of the events which occur on a trial. Find the distribution for *X* and determine the probability that two or more occur on a trial. Find the probability that one or three of these occur on a trial.

Answer

```
npr06_12
Minterm probabilities in pm, coefficients in c
```

```
a = imintest(pm)
The class is NOT independent
Minterms for which the product rule fails
a =
     1     1     1     1
     1     1     1     1
     1     1     1     1
     1     1     1     1
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
XDBN =
         0    0.0050
    1.0000    0.0430
    2.0000    0.2120
    3.0000    0.4380
    4.0000    0.3020
P2 = (X>=2)*PX'
P2 =  0.9520
P13 = ((X==1)|(X==3))*PX'
P13 =  0.4810
```

Exercise 6.2.13

James is expecting three checks in the mail, for \$20, \$26, and \$33 dollars. Their arrivals are the events A, B, C. Assume the class is independent, with respective probabilities 0.90, 0.75, 0.80. Then

$$X = 20I_A + 26I_B + 33I_C$$

represents the total amount received. Determine the distribution for X. What is the probability he receives at least \$50? Less than \$30?

Answer

```
c = [20 26 33 0];
P = 0.01*[90 75 80];
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  minprob(P)
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
disp(XDBN)
         0    0.0050
   20.0000    0.0450
   26.0000    0.0150
   33.0000    0.0200
   46.0000    0.1350
   53.0000    0.1800
   59.0000    0.0600
   79.0000    0.5400
P50 = (X>=50)*PX'
P50 =  0.7800
P30 = (X <30)*PX'
P30 =  0.0650
```

Exercise 6.2.14

A gambler places three bets. He puts down two dollars for each bet. He picks up three dollars (his original bet plus one dollar) if he wins the first bet, four dollars if he wins the second bet, and six dollars if he wins the third. His net winning can be represented by the random variable

$$X = 3I_A + 4I_B + 6I_C - 6, \text{ with } P(A) = 0.5,\ P(B) = 0.4,\ P(C) = 0.3$$

Assume the results of the games are independent. Determine the distribution for X.

Answer

```
c = [3 4 6 -6];
P = 0.1*[5 4 3];
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  minprob(P)
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
dsp(XDBN)
   -6.0000    0.2100
   -3.0000    0.2100
   -2.0000    0.1400
         0    0.0900
    1.0000    0.1400
    3.0000    0.0900
    4.0000    0.0600
    7.0000    0.0600
```

Exercise 6.2.15

Henry goes to a hardware store. He considers a power drill at $35, a socket wrench set at $56, a set of screwdrivers at $18, a vise at $24, and hammer at $8. He decides independently on the purchases of the individual items, with respective probabilities 0.5, 0.6, 0.7, 0.4, 0.9. Let X be the amount of his total purchases. Determine the distribution for X.

Answer available at the end of this section.

Exercise 6.2.16

A sequence of trials (not necessarily independent) is performed. Let E_i be the event of success on the ith component trial. We associate with each trial a "payoff function" $X_i = aI_{E_i} + bI_{E_i^c}$. Thus, an amount a is earned if there is a success on the trial and an amount b (usually negative) if there is a failure. Let S_n be the number of successes in the n trials and W be the net payoff. Show that $W = (a - b)S_n + bn$.

Answer available at the end of this section.

Exercise 6.2.17

A marker is placed at a reference position on a line (taken to be the origin); a coin is tossed repeatedly. If a head turns up, the marker is moved one unit to the right; if a tail turns up, the marker is moved one unit to the left.

a. Show that the position at the end of ten tosses is given by the random variable

$$X = \sum_{i=1}^{10} I_{E_i} - \sum_{i=1}^{10} I_{E_i^c} = 2\sum_{i=1}^{10} I_{E_i} - 10 = 2S_{10} - 10$$

where E_i is the event of a head on the ith toss and S_{10} is the number of heads in ten trials.

- After ten tosses, what are the possible positions and the probabilities of being in each?

Answer

$$X_i = I_{E_i} - I_{E_i^c} = I_{E_i} - (1 - I_{E_i}) = 2I_{E_i} - 1$$

$$X = \sum_{i=1}^{10} X_i = 2\sum_{i=1}^{n} I_{E_i} - 10$$

```
S = 0:10;
PS = ibinom(10,0.5,0:10);
X = 2*S - 10;
disp([X;PS]')
  -10.0000    0.0010
   -8.0000    0.0098
   -6.0000    0.0439
   -4.0000    0.1172
   -2.0000    0.2051
         0    0.2461
    2.0000    0.2051
    4.0000    0.1172
    6.0000    0.0439
    8.0000    0.0098
   10.0000    0.0010
```

Exercise 6.2.18

Margaret considers five purchases in the amounts 5, 17, 21, 8, 15 dollars with respective probabilities 0.37, 0.22, 0.38, 0.81, 0.63. Anne contemplates six purchases in the amounts 8, 15, 12, 18, 15, 12 dollars, with respective probabilities 0.77, 0.52, 0.23, 0.41, 0.83, 0.58. Assume that all eleven possible purchases form an independent class.

a. Determine the distribution for X, the amount purchased by Margaret.
b. Determine the distribution for Y, the amount purchased by Anne.
c. Determine the distribution for $Z = X + Y$, the total amount the two purchase.

Suggestion for part (c). Let MATLAB perform the calculations.

Answer

```
[r,s] = ndgrid(X,Y);
[t,u] = ndgrid(PX,PY);
z = r + s;
pz = t.*u;
[Z,PZ] = csort(z,pz);
```

```
% file npr06_18.m
cx = [5 17 21 8 15 0];
cy = [8 15 12 18 15 12 0];
pmx = minprob(0.01*[37 22 38 81 63]);
pmy = minprob(0.01*[77 52 23 41 83 58]);
npr06_18
```

```
[X,PX] = canonicf(cx,pmx);  [Y,PY] = canonicf(cy,pmy);
[r,s] = ndgrid(X,Y);   [t,u] = ndgrid(PX,PY);
z = r + s;   pz = t.*u;
[Z,PZ] = csort(z,pz);
a = length(Z)
a  =  125              % 125 different values
plot(Z,cumsum(PZ))  % See figure     Plotting details omitted
```

Answers

Exercise 6.2.1

- $A \bigvee B \bigvee C$
- D
- $A \bigvee B \bigvee C$
- C
- $C \bigvee D$

Exercise 6.2.2

- D
- $A \bigvee B$
- $A \bigvee B$
- $B \bigvee C \bigvee D \bigvee E$
- $A \bigvee D \bigvee E$

Exercise 6.2.3

```
T = [1 3 2 3 4 2 1 3 5 2];
[X,I] = sort(T)
X =   1   1   2   2   2   3   3   3   4   5
I =   1   7   3   6  10   2   4   8   5   9
```

$$X = I_A + 2I_B + 3I_C + 4I_D + 5I_E$$
$$A = C_1 \bigvee C_7,\ B = C_3 \bigvee C_6 \bigvee C_{10},\ C = C_2 \bigvee C_4 \bigvee C_8,\ D = C_5,\ E = C_9$$

Exercise 6.2.4

```
T = [1 3 2 3 4 2 1 3 5 2];
pc = 0.01*[8 13 6 9 14 11 12 7 11 9];
[X,PX] = csort(T,pc);
disp([X;PX]')
    1.0000    0.2000
    2.0000    0.2600
    3.0000    0.2900
    4.0000    0.1400
    5.0000    0.1100
```

Exercise 6.2.7

```
A = [2 3 5];
B = [1 2 3];
a = rowcopy(A,3);
b = colcopy(B,3);
Z =a + b               % Possible values of sum Z = X + Y
Z = 3     4     6
    4     5     7
    5     6     8
PA = [0.3 0.6 0.1];
PB = [0.2 0.6 0.2];
 pa= rowcopy(PA,3);
 pb = colcopy(PB,3);
 P = pa.*pb            % Probabilities for various values
P =  0.0600    0.1200    0.0200
     0.1800    0.3600    0.0600
```

```
     0.0600    0.1200    0.0200
[Z,PZ] = csort(Z,P);
 disp([Z;PZ]')         % Distribution for Z = X + Y
    3.0000    0.0600
    4.0000    0.3000
    5.0000    0.4200
    6.0000    0.1400
    7.0000    0.0600
    8.0000    0.0200
```

Exercise 6.2.8

```
XY = a.*b
XY = 2     3     5               % XY values
     4     6    10
     6     9    15


       W         PW              % Distribution for W = XY
    2.0000    0.0600
    3.0000    0.1200
    4.0000    0.1800
    5.0000    0.0200
    6.0000    0.4200
    9.0000    0.1200
   10.0000    0.0600
   15.0000    0.0200
```

Exercise 6.2.10

```
% file npr06_10.m
% Data for Exercise 6.2.10.
pm = [ 0.072 0.048 0.018 0.012 0.168 0.112 0.042 0.028 ...
       0.062 0.048 0.028 0.010 0.170 0.110 0.040 0.032];
c  = [-5.3 -2.5 2.3 4.2 -3.7];
disp('Minterm probabilities are in pm, coefficients in c')
npr06_10
Minterm probabilities are in pm, coefficients in c
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
XDBN
XDBN =
  -11.5000    0.1700
   -9.2000    0.0400
   -9.0000    0.0620
   -7.3000    0.1100
   -6.7000    0.0280
   -6.2000    0.1680
   -5.0000    0.0320
   -4.8000    0.0480
   -3.9000    0.0420
   -3.7000    0.0720
   -2.5000    0.0100
   -2.0000    0.1120
   -1.4000    0.0180
    0.3000    0.0280
```

```
    0.5000    0.0480
    2.8000    0.0120
```

Exercise 6.2.15

```
c = [35 56 18 24 8 0];
P = 0.1*[5 6 7 4 9];
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  minprob(P)
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
disp(XDBN)
         0    0.0036
    8.0000    0.0324
   18.0000    0.0084
   24.0000    0.0024
   26.0000    0.0756
   32.0000    0.0216
   35.0000    0.0036
   42.0000    0.0056
   43.0000    0.0324
   50.0000    0.0504
   53.0000    0.0084
   56.0000    0.0054
   59.0000    0.0024
   61.0000    0.0756
   64.0000    0.0486
   67.0000    0.0216
   74.0000    0.0126
   77.0000    0.0056
   80.0000    0.0036
   82.0000    0.1134
   85.0000    0.0504
   88.0000    0.0324
   91.0000    0.0054
   98.0000    0.0084
   99.0000    0.0486
  106.0000    0.0756
  109.0000    0.0126
  115.0000    0.0036
  117.0000    0.1134
  123.0000    0.0324
  133.0000    0.0084
  141.0000    0.0756
```

Exercise 6.2.16

$$X_i = aI_{E_i} + b(1 - I_{E_i}) = (a - b)I_{E_i} + b$$
$$W = \sum_{i=1}^{n} X_i = (a - b)\sum_{i=1}^{n} I_{E_i} + bn = (a - b)S_n + bn$$

This page titled 6.2: Problems on Random Variables and Probabilities is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

7: DISTRIBUTION AND DENSITY FUNCTIONS

7.1: Distribution and Density Functions
7.2: Distribution Approximations
7.3: Problems on Distribution and Density Functions

This page titled 7: Distribution and Density Functions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Distribution and Density Functions

In the unit on Random Variables and Probability we introduce real random variables as mappings from the basic space Ω to the real line. The mapping induces a transfer of the probability mass on the basic space to subsets of the real line in such a way that the probability that X takes a value in a set M is exactly the mass assigned to that set by the transfer. To perform probability calculations, we need to describe analytically the distribution on the line. For simple random variables this is easy. We have at each possible value of X a point mass equal to the probability X takes that value. For more general cases, we need a more useful description than that provided by the induced probability measure P_X.

The Distribution Function

In the theoretical discussion on Random Variables and Probability, we note that the probability distribution induced by a random variable X is determined uniquely by a consistent assignment of mass to semi-infinite intervals of the form $(-\infty, t]$ for each real t. This suggests that a natural description is provided by the following.

Definition

The *distribution function* F_X for random variable X is given by

$$F_X(t) P(X \leq t) = P(X \in (-\infty, t]) \ \forall t \in R$$

In terms of the mass distribution on the line, this is the probability mass *at or to the left of the point t*. As a consequence, F_X has the following properties:

- **(F1) :** F_X must be a nondecreasing function, for if $t > s$ there must be at least as much probability mass at or to the left of t as there is for s.
- **(F2) :** F_X is continuous from the right, with a jump in the amount p_0 at t_0 iff $P(X = t_0) = p_0$. If the point t approaches t_0 from the left, the interval does not include the probability mass at t_0 until t reaches that value, at which point the amount at or to the left of *t* increases ("jumps") by amount p_0; on the other hand, if t approaches t_0 from the right, the interval includes the mass p_0 all the way to and including t_0, but drops immediately as t moves to the left of t_0.
- **(F3) :** Except in very unusual cases involving random variables which may take “infinite” values, the probability mass included in $(-\infty, t]$ must increase to one as *t* moves to the right; as t moves to the left, the probability mass included must decrease to zero, so that

  $$F_X(-\infty) = \lim_{t \to -\infty} F_X(t) = 0 \tag{7.1.1}$$

  and

  $$F_X(\infty) = \lim_{t \to \infty} F_X(t) = 1 \tag{7.1.2}$$

A distribution function determines the probability mass in each semiinfinite interval $(\infty, t]$. According to the discussion referred to above, this determines uniquely the induced distribution.

The distribution function F_X for a simple random variable is easily visualized. The distribution consists of point mass p_i at each point t_i in the range. To the left of the smallest value in the range, $F_X(t) = 0$; as *t* increases to the smallest value t_1, $F_X(t)$ remains constant at zero until it jumps by the amount p_1 ... $F_X(t)$ remains constant at p_1 until t increases to t_2, where it jumps by an amount p_2 to the value $p_1 + p_2$. This continues until the value of $F_X(t)$ reaches 1 at the largest value t_n. The graph of F_X is thus a step function, continuous from the right, with a jump in the amount p_i at the corresponding point t_i in the range. A similar situation exists for a discrete-valued random variable which may take on an infinity of values (e.g., the geometric distribution or the Poisson distribution considered below). In this case, there is always some probability at points to the right of any t_i, but this must become vanishingly small as t increases, since the total probability mass is one.

The procedure *ddbn* may be used to plot the distribution function for a simple random variable from a matrix *X* of values and a corresponding matrix **PX** of probabilities.

Example 7.1.1: Graph of FX for a simple random variable

```
>> c = [10 18 10 3];            % Distribution for X in Example 6.5.1
>> pm = minprob(0.1*[6 3 5]);
>> canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
>> ddbn                         % Circles show values at jumps
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX
%  Printing details   See Figure 7.1
```

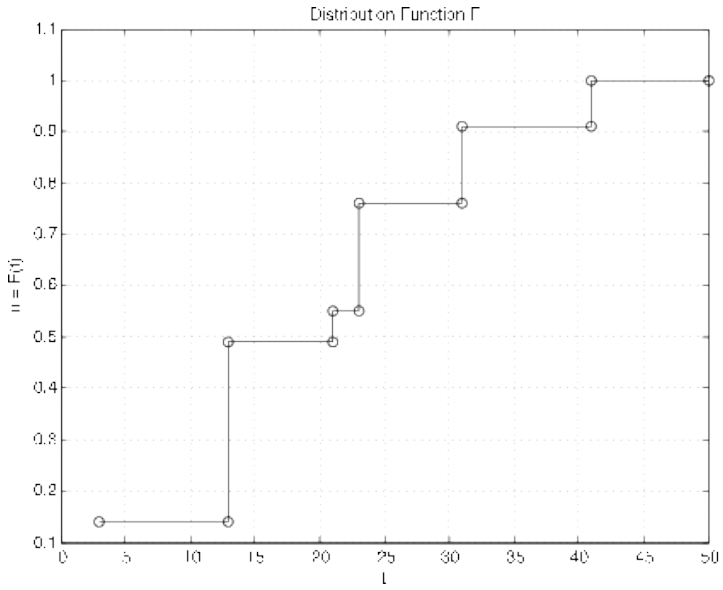


Figure 7.1.1. Distribution function for Example 7.1.1

Description of some common discrete distributions

We make repeated use of a number of common distributions which are used in many practical situations. This collection includes several distributions which are studied in the chapter "Random Variables and Probabilities".

Indicator function. $X = I_E P(X = 1) = P(E) = pP(X = 0) = q = 1 - p$. The distribution function has a jump in the amount q at $t = 0$ and an additional jump of p to the value 1 at $t = 1$.

Simple random variable $X = \sum_{t_i} I_{A_i}$ (canonical form)

$$P(X = t_i) = P(A_i) = p_1$$

The distribution function is a step function, continuous from the right, with jump of p_i at $t = t_i$ (See Figure 7.1.1 for Example 7.1.1)

Binomial (n, p). This random variable appears as the number of successes in a sequence of n Bernoulli trials with probability p of success. In its simplest form

$$X = \sum_{i=1}^{n} I_{E_i} \text{ with } \{E_i : 1 \le i \le n\} \text{ independent}$$

$$P(E_i) = p \; P(X = k) = C(n, k) p^k q^{n-k}$$

As pointed out in the study of Bernoulli sequences in the unit on Composite Trials, two m-functions *ibinom* and*cbinom* are available for computing the individual and cumulative binomial probabilities.

Geometric (p) There are two related distributions, both arising in the study of continuing Bernoulli sequences. The first counts the number of failures *before* the first success. This is sometimes called the "waiting time." The event $\{X = k\}$ consists of a sequence of k failures, then a success. Thus

$$P(X = k) = q^k p,\ 0 \leq k$$

The second designates the component trial on which the first success occurs. The event $\{Y = k\}$ consists of $k - 1$ failures, then a success on the kth component trial. We have

$$P(Y = k) = q^{k-1} p,\ 1 \leq k$$

We say X has the geometric distribution with parameter (p), which we often designate by X geometric (p). Now $Y = X + 1$ or $Y - 1 = X$. For this reason, it is customary to refer to the distribution for the number of the trial for the first success by saying $Y - 1$ geometric (p). The probability of k or more failures before the first success is $P(X \geq k) = q^k$. Also

$$P(X \geq n + k | X \geq n) = \frac{P(X \geq n + k)}{P(X \geq n)} = q^{n+k}/q^n = q^k = P(X \geq k)$$

This suggests that a Bernoulli sequence essentially "starts over" on each trial. If it has failed n times, the probability of failing an additional k or more times before the next success is the same as the initial probability of failing k or more times before the first success.

Example 7.1.2: The geometric distribution

A statistician is taking a random sample from a population in which two percent of the members own a BMW automobile. She takes a sample of size 100. What is the probability of finding no BMW owners in the sample?

Solution

The sampling process may be viewed as a sequence of Bernoulli trials with probability $p = 0.02$ of success. The probability of 100 or more failures before the first success is $0.98^{100} = 0.1326$ or about 1/7.5.

Negative binomial (m, p). X is the number of failures before the mth success. It is generally more convenient to work with $Y = X + m$, the number of the trial on which the mth success occurs. An examination of the possible patterns and elementary combinatorics show that

$$P(Y = k) = C(k - 1, m - 1) p^m q^{k-m},\ m \leq k$$

There are *m*–1 successes in the first $k - 1$ trials, then a success. Each combination has probability $p^m q^{k-m}$. We have an m-function *nbinom* to calculate these probabilities.

Example 7.1.3: A game of chance

A player throws a single six-sided die repeatedly. He scores if he throws a 1 or a 6. What is the probability he scores five times in ten or fewer throws?

```
>> p = sum(nbinom(5,1/3,5:10))
p  =  0.2131
```

An *alternate solution* is possible with the use of the *binomial distribution*. The mth success comes not later than the *k*th trial iff the number of successes in k trials is greater than or equal to m.

```
>> P = cbinom(10,1/3,5)
P  =  0.2131
```

Poisson (μ). This distribution is assumed in a wide variety of applications. It appears as a counting variable for items arriving with exponential interarrival times (see the relationship to the gamma distribution below). For large n and small p (which may not be a value found in a table), the binomial distribution is approximately Poisson (np). Use of the generating function (see Transform Methods) shows the sum of independent Poisson random variables is Poisson. The Poisson distribution is integer valued, with

$$P(X = k) = e^{-\mu}\frac{\mu^k}{k!} \text{ (0 \le k\)}$$

Although Poisson probabilities are usually easier to calculate with scientific calculators than binomial probabilities, the use of tables is often quite helpful. As in the case of the binomial distribution, we have two m-functions for calculating Poisson probabilities. These have advantages of speed and parameter range similar to those for ibinom and cbinom.

$P(X = k)$ is calculated by `P = ipoisson(mu,k)`, where k is a row or column vector of integers and the result P is a row matrix of the probabilities.
$P(X \geq k)$ is calculated by `P = cpoisson(mu,k)`, where k is a row or column vector of integers and the result P is a row matrix of the probabilities.

Example 7.1.4: Poisson counting random variable

The number of messages arriving in a one minute period at a communications network junction is a random variable $N \sim$ Poisson (130). What is the probability the number of arrivals is greater than equal to 110, 120, 130, 140, 150, 160 ?

```
>> p = cpoisson(130,110:10:160)
p  =  0.9666  0.8209  0.5117  0.2011  0.0461  0.0060
```

The descriptions of these distributions, along with a number of other facts, are summarized in the table DATA ON SOME COMMON DISTRIBUTIONS in Appendix C.

The Density Function

If the probability mass in the induced distribution is spread smoothly along the real line, with no point mass concentrations, there is a *probability density* function f_X which satisfies

$$P(X \in M) = P_X(M) = \int_M f_X(t)\, dt \text{ (are under the graph of } f_X \text{ over } M)$$

At each t, $f_X(t)$ is the mass per unit length in the probability distribution. The density function has three characteristic properties:

$$\text{(f1) } f_X \geq 0 \text{ (f2) } \int_R f_X = 1 \text{ (f3) } F_X(t) = \int_{-\infty}^{t} f_X$$

A random variable (or distribution) which has a density is called *absolutely continuous*. This term comes from measure theory. We often simply abbreviate as *continuous* distribution.

Remarks

1. There is a technical mathematical description of the condition “spread smoothly with no point mass concentrations.” And strictly speaking the integrals are Lebesgue integrals rather than the ordinary Riemann kind. But for practical cases, the two agree, so that we are free to use ordinary integration techniques.
2. By the fundamental theorem of calculus

$$f_X(t) = F_X'(t) \text{ at every point of continuity of } f_X$$

- Any integrable, nonnegative function f with $\int f = 1$ determines a distribution function F, which in turn determines a

probability distribution. If $\int f \neq 1$, multiplication by the appropriate positive constant gives a suitable f. An argument based on the Quantile Function shows the existence of a random variable with that distribution.

- In the literature on probability, it is customary to omit the indication of the region of integration when integrating over the whole line. Thus

$$\int g(t) f_X(t) dt = \int_R g(t) f_X(t) dt$$

The first expression is *not* an indefinite integral. In many situations, f_X will be zero outside an interval. Thus, the integrand effectively determines the region of integration.

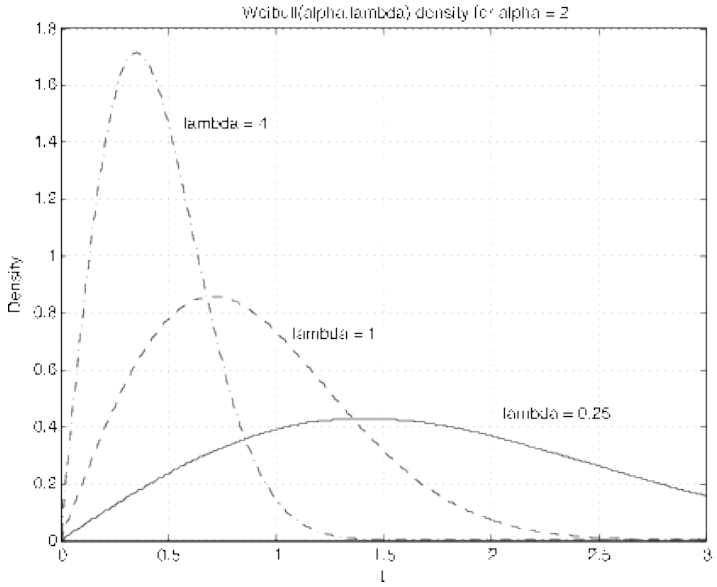


Figure 7.1.2. The Weibull density for $\alpha = 2$, $\lambda = 0.25, 1, 4$.

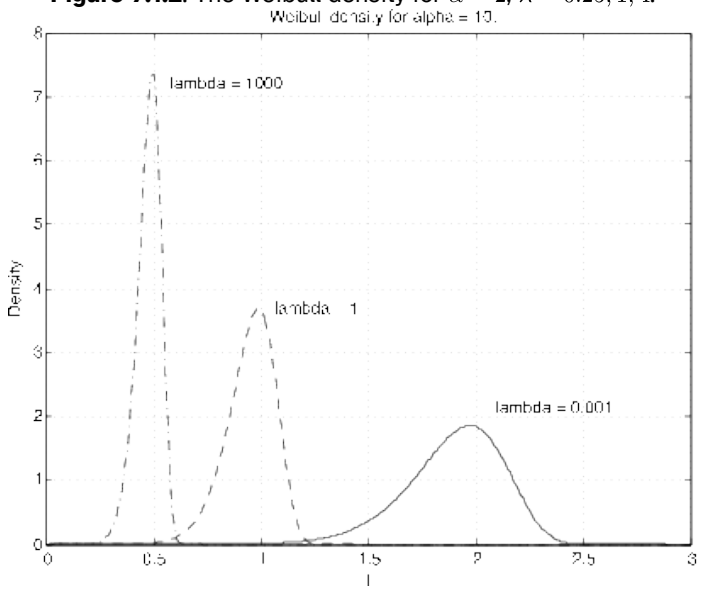


Figure 7.1.3. The Weibull density for $\alpha = 10$, $\lambda = 0.001, 1, 1000$.

Some common absolutely continuous distributions

Uniform (a, b).
Mass is spread uniformly on the interval $[a, b]$. It is immaterial whether or not the end points are included, since probability associated with each individual point is zero. The probability of any subinterval is proportional to the length of the subinterval. The probability of being in any two subintervals of the same length is the same. This distribution is used to model situations in

which it is known that X takes on values in $[a, b]$ but is equally likely to be in any subinterval of a given length. The density must be constant over the interval (zero outside), and the distribution function increases linearly with t in the interval. Thus,

$$f_X(t) = \frac{1}{b-a} \ (a < t < b) \text{ (zero outside the interval)}$$

The graph of F_X rises linearly, with slope 1/$(b - a)$ from zero at $t = a$ to one at $t = b$.

Symmetric triangular $(-a, a)$, $f_X(t) = \begin{cases} (a+t)/a^2 & -a \le t < 0 \\ (a-t)/a^2 & 0 \le t \le a \end{cases}$.

This distribution is used frequently in instructional numerical examples because probabilities can be obtained geometrically. It can be shifted, with a shift of the graph, to different sets of values. It appears naturally (in shifted form) as the distribution for the sum or difference of two independent random variables uniformly distributed on intervals of the same length. This fact is established with the use of the moment generating function (see Transform Methods). More generally, the density may have a triangular graph which is not symmetric.

Example 7.1.5: Use of a triangular distribution

Suppose X symmetric triangular (100, 300). Determine $P(120 < X \le 250)$.

Remark. Note that in the continuous case, it is immaterial whether the end point of the intervals are included or not.

Solution

To get the area under the triangle between 120 and 250, we take one minus the area of the right triangles between 100 and 120 and between 250 and 300. Using the fact that areas of similar triangles are proportional to the square of any side, we have

$$P = 1 - \frac{1}{2}((20/100)^2 + (50/100)^2) = 0.855$$

Exponential (λ) $f_X(t) = \lambda e^{-\lambda t}$ $t \ge 0$ (zero elsewhere).
Integration shows $F_X(t) = 1 - e^{-\lambda t}$ (t \ge 0\) (zero elsewhere). We note that $P(X > 0) = 1 - F_X(t) = e^{-\lambda t}$ $t \ge 0$. This leads to an extremely important property of the exponential distribution. Since $X > t + h$, $h > 0$ implies $X > t$, we have

$$P(X > t + h | X > t) = P(X > t + h)/P(X > t) = e^{-\lambda(t+h)}/e^{-\lambda t} = e^{-\lambda h} = P(X > h)$$

Because of this property, the exponential distribution is often used in reliability problems. Suppose X represents the time to failure (i.e., the life duration) of a device put into service at $t = 0$. If the distribution is exponential, this property says that if the device survives to time t (i.e., $X > t$) then the (conditional) probability it will survive h more units of time is the same as the original probability of surviving for h units of time. Many devices have the property that they do not wear out. Failure is due to some stress of external origin. Many solid state electronic devices behave essentially in this way, once initial "burn in" tests have removed defective units. Use of Cauchy's equation (Appendix B) shows that the exponential distribution is the only continuous distribution with this property.

Gamma distribution (α, λ) $f_X(t) = \dfrac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}$ $t \ge 0$ (zero elsewhere)

We have an m-function *gammadbn* to determine values of the distribution function for X gamma (α, λ). Use of moment generating functions shows that for $\alpha = n$, a random variable X gamma (n, λ) has the same distribution as the sum of n independent random variables, each exponential $(lambda)$. A relation to the Poisson distribution is described in Sec 7.5.

Example 7.1.6: An arrival problem

On a Saturday night, the times (in hours) between arrivals in a hospital emergency unit may be represented by a random quantity which is exponential $(\lambda = 3)$. As we show in the chapter Mathematical Expectation, this means that the average interarrival time is 1/3 hour or 20 minutes. What is the probability of ten or more arrivals in four hours? In six hours?

Solution

The time for ten arrivals is the sum of ten interarrival times. If we suppose these are independent, as is usually the case, then the time for ten arrivals is gamma (10, 3).

```
>> p = gammadbn(10,3,[4 6])
p  =  0.7576    0.9846
```

Normal, or Gaussian (μ, σ^2) $f_X(t) = \dfrac{1}{\sigma\sqrt{2\pi}}$ exp $(-\dfrac{1}{2}(\dfrac{t-\mu}{\sigma})^2)$ $\forall t$

We generally indicate that a random variable X has the normal or gaussian distribution by writing X $N(\mu, \sigma^2)$, putting in the actual values for the parameters. The gaussian distribution plays a central role in many aspects of applied probability theory, particularly in the area of statistics. Much of its importance comes from the *central limit theorem* (CLT), which is a term applied to a number of theorems in analysis. Essentially, the CLT shows that the distribution for the sum of a sufficiently large number of independent random variables has approximately the gaussian distribution. Thus, the gaussian distribution appears naturally in such topics as theory of errors or theory of noise, where the quantity observed is an additive combination of a large number of essentially independent quantities. Examination of the expression shows that the graph for $f_X(t)$ is symmetric about its maximum at $t = \mu$ .. The greater the parameter σ^2, the smaller the maximum value and the more slowly the curve decreases with distance from μ.. Thus parameter μ. locates the center of the mass distribution and σ^2 is a measure of the spread of mass about μ. The parameter μ is called the *mean value* and σ^2 is the *variance*. The parameter σ, the positive square root of the variance, is called the *standard deviation*. While we have an explicit formula for the density function, it is known that the distribution function, as the integral of the density function, cannot be expressed in terms of elementary functions. The usual procedure is to use tables obtained by numerical integration.

Since there are two parameters, this raises the question whether a separate table is needed for each pair of parameters. It is a remarkable fact that this is not the case. We need only have a table of the distribution function for X $N(0, 1)$. This is refered to as the *standardized normal* distribution. We use φ and ϕ for the standardized normal density and distribution functions, respectively.

Standardized normal $varphi(t) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$ so that the distribution function is $\phi(t) = \int_{-\infty}^{t} \varphi(u)du$.

The graph of the density function is the well known bell shaped curve, symmetrical about the origin (see Figure 7.1.4). The symmetry about the origin contributes to its usefulness.

$P(X \le t) = \phi(t)$ = area under the curve to the left of t

Note that the area to the left of $t = -1.5$ is the same as the area to the right of $t = 1.5$, so that $\phi(-2) = 1 - \phi(2)$. The same is true for any t, so that we have

$$\phi(-t) = 1 - \phi(t) \ \forall t$$

This indicates that we need only a table of values of $\phi(t)$ for $t > 0$ to be able to determine $\phi(t)$ for any t. We may use the symmetry for any case. Note that $\phi(0) = 1/2$,

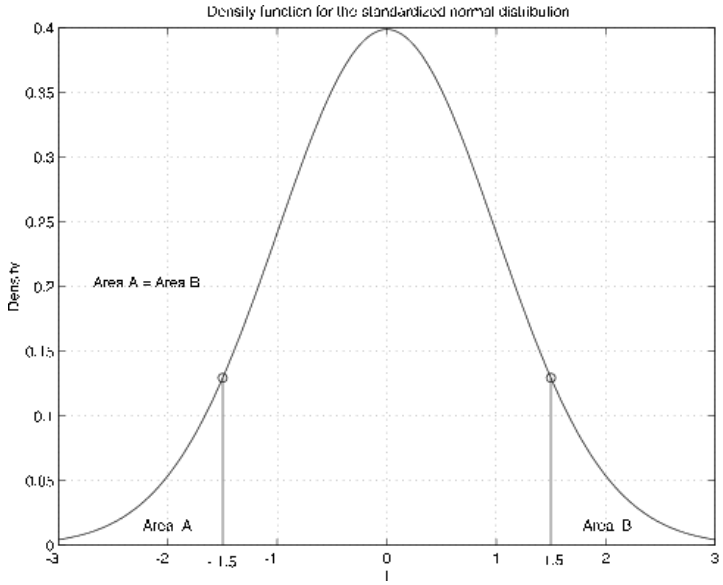


Figure 7.1.4. The standardized normal distribution.

Example 7.1.7: Standardized normal calculations

Suppose $X \sim N(0,1)$. Determine $P(-1 \leq X \leq 2)$ and $P(|X| > 1)$

Solution

1. $P(-1 \leq X \leq 2) = \phi(2) - \phi(-1) = \phi(2) - [1 - \phi(1)] = \phi(2) + \phi(1) - 1$
2. $P(|X| > 1) = P(X > 1) + P(X < -1) = 1 - \phi(1) + \phi(-1) = 2[1 - \phi(1)]$

From a table of standardized normal distribution function (see Appendix D), we find

$\phi(2) = 0.9772$ and $\phi(1) = 0.8413$ which gives $P(-1 \leq X \leq 2) = 0.8185$ and $P(|X| > 1) = 0.3174$

General gaussian distribution

For $X \sim N(\mu, \sigma^2)$, the density maintains the bell shape, but is shifted with different spread and height. Figure 7.1.5 shows the distribution function and density function for $X \sim N(2, 0.1)$. The density is centered about $t = 2$. It has height 1.2616 as compared with 0.3989 for the standardized normal density. Inspection shows that the graph is narrower than that for the standardized normal. The distribution function reaches 0.5 at the mean value 2.

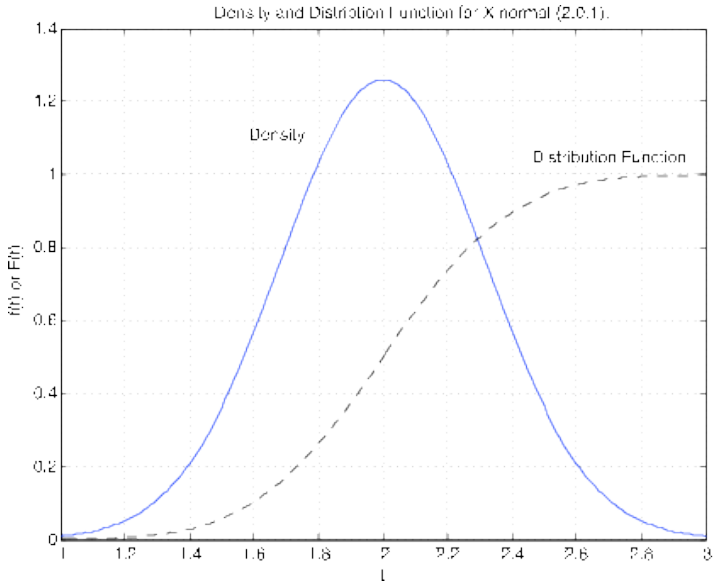


Figure 7.1.5. The normal density and distribution functions for X $N(2, 0.1)$.

A change of variables in the integral shows that the table for standardized normal distribution function can be used for any case.

$$F_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)dx = \int_{-infty}^{t} \varphi(\frac{x-\mu}{\sigma})\frac{1}{\sigma}dx$$

Make the change of variable and corresponding formal changes

$$u = \frac{x-\mu}{\sigma} \; du = \frac{1}{\sigma}dx \; x = t \; u = \frac{t-\mu}{\sigma}$$

to get

$$F_X(t) = \int_{-\infty}^{(t-\mu)/\sigma} \varphi(u)du = \phi(\frac{t-\mu}{\sigma})$$

Example 7.1.8: General gaussian calculation

Suppose X N(3,16) (i.e., $\mu = 3$ and $\sigma^2 = 16$). Determine $P(-1 \leq X \leq 11)$ and $P(|X-3| > 4)$.

Solution

1. $F_X(11) - F_X(-1) = \phi(\frac{11-3}{4}) - \phi(\frac{-1-3}{4}) = \phi(2) - \phi(-1) = 0.8185$
2. $P(X-3 < -4) + P(X-3 > 4) = F_X(-1) + [1 - F_X(7)] = \phi(-1) + 1 - \phi(1) = 0.3174$

In each case the problem reduces to that in Example.

We have m-functions *gaussian* and *gaussdensity* to calculate values of the distribution and density function for any reasonable value of the parameters.
The following are solutions of example 7.1.7 and example 7.1.8, using the m-function gaussian.

Example 7.1.9: Example 7.1.7 and Example 7.1.8 (continued)

```
01 >> P1 = gaussian(0,1,2) - gaussian(0,1,-1)
02 P1 =  0.8186
03 >> P2 = 2*(1 - gaussian(0,1,1))
```

```
P2 =  0.3173
>> P1 = gaussian(3,16,11) - gaussian(3,16,-1)
P2 =  0.8186
>> P2 = gaussian(3,16,-1)) + 1 - (gaussian(3,16,7)
P2 =  0.3173
```

The differences in these results and those above (which used tables) are due to the roundoff to four places in the tables.

Beta (r, s), $r > 0$, $s > 0$. $f_X(t) = \dfrac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} t^{r-1}(1-t)^{s-1}$ $0 < t < 1$

Analysis is based on the integrals

$$\int_0^1 u^{r-1}(1-u)^{s-1}du = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} \text{ with } \Gamma(t+1) = t\Gamma(t)$$

Figure 7.6 and Figure 7.7 show graphs of the densities for various values of r, s. The usefulness comes in approximating densities on the unit interval. By using scaling and shifting, these can be extended to other intervals. The special case $r = s = 1$ gives the uniform distribution on the unit interval. The Beta distribution is quite useful in developing the Bayesian statistics for the problem of sampling to determine a population proportion. If r, s are integers, the density function is a polynomial. For the general case we have two m-functions, *beta* and *betadbn* to perform the calculatons.

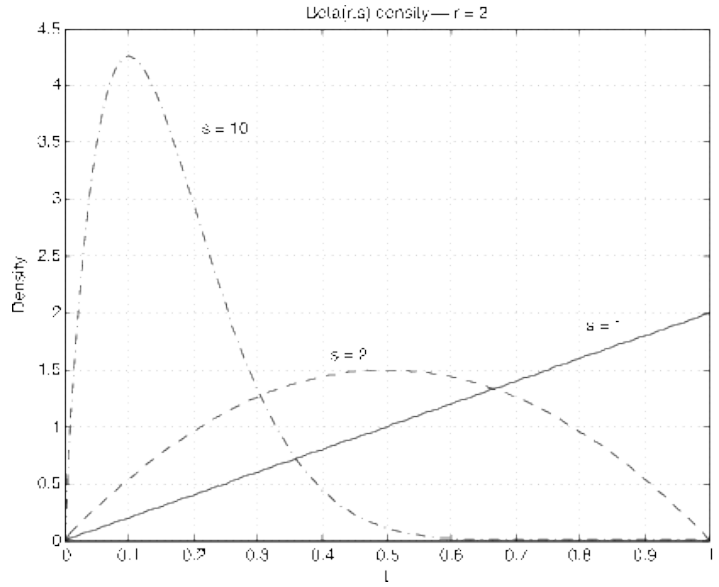


Figure 7.6. The Beta (r, s) density for $r = 2, s = 1, 2, 10$.

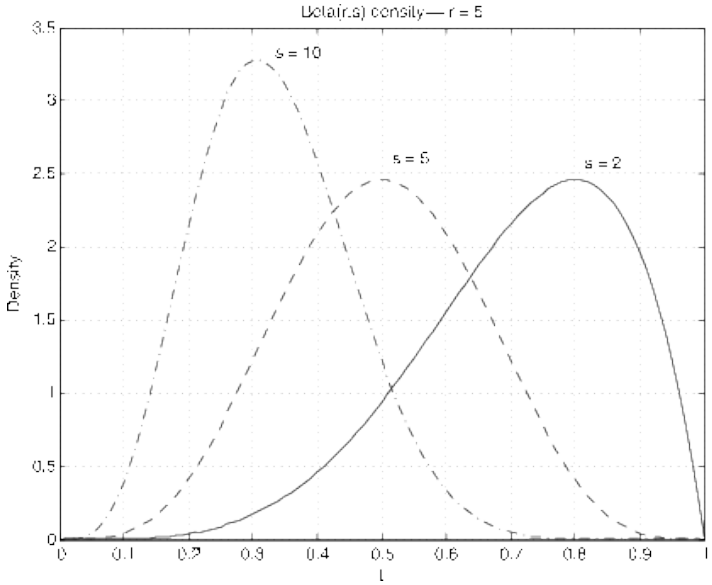


Figure 7.7. The Beta (r, s) density for $r = 5, s = 2, 5, 10$.

Weibull (α, λ, v) $F_X(t) = 1 - e^{-\lambda(t-v)^\alpha}$ $\alpha > 0$, $\lambda > 0$, $v \geq 0$, $t \geq v$

The parameter v is a shift parameter. Usually we assume $v = 0$. Examination shows that for α=1 the distribution is exponential (λ). The parameter α provides a distortion of the time scale for the exponential distribution. Figure 7.6 and Figure 7.7 show graphs of the Weibull density for some representative values of α and λ $(v = 0)$. The distribution is used in reliability theory. We do not make much use of it. However, we have m-functions *weibull* (density) and *weibulld* (distribution function) for shift parameter $v = 0$ only. The shift can be obtained by subtracting a constant from the t values.

This page titled 7.1: Distribution and Density Functions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

7.2: Distribution Approximations

Binomial, Poisson, gamma, and Gaussian distributions

The Poisson approximation to the binomial distribution

The following approximation is a classical one. We wish to show that for small p and sufficiently large n

$$P(X=k) = C(n,k)p^k(1-p)^{n-k} \approx e^{-np}\frac{np}{k!} \tag{7.2.1}$$

Suppose $p = \mu/n$ with n large and $\mu/n < 1$. Then,

$$P(X=k) = C(n,k)(\mu/n)^k(1-\mu/n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k}(1-\frac{\mu}{n})^{-k}(1-\frac{\mu}{n})^n\frac{\mu^k}{k!} \tag{7.2.2}$$

The first factor in the last expression is the ratio of polynomials in n of the same degree k, which must approach one as n becomes large. The second factor approaches one as n becomes large. According to a well known property of the exponential

$$(1-\frac{\mu}{n})^n \to e^{-\mu} \tag{7.2.3}$$

as $n \to \infty$.

The result is that for large n, $P(X=k) \approx e^{-\mu}\frac{\mu^k}{k!}$, where $\mu = np$.

The Poisson and Gamma Distributions

Suppose Y Poisson (λt). Now X gamma (α, λ) iff

$$P(X \le t) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^1 x^{\alpha-1}e^{-\lambda x}\,dx = \frac{1}{\Gamma(\alpha)}\int_0^t (\lambda x)^{\alpha-1}e^{\lambda x}d(\lambda x) = \frac{1}{\Gamma(\alpha)}\int_0^{\lambda t} u^{\alpha-1}e^{-\mu}\,du \tag{7.2.4}$$

A well known definite integral, obtained by integration by parts, is

$$int_\alpha^\infty t^{n-1}e^{-t}dt = \Gamma(n)e^{-a}\sum_{k=1}^{n-1}\frac{a^k}{k!} \tag{7.2.5}$$

with $\Gamma(n) = (n-1)!$.

Noting that $1 = e^{-a}e^a = e^{-a}\sum_{k=0}^\infty \frac{a^k}{k!}$ we find after some simple algebra that

$$\frac{1}{\Gamma(n)}\int_0^a t^{n-1}e^{-t}\,dt = e^{-a}\sum_{k=n}^\infty \frac{a^k}{k!} \tag{7.2.6}$$

For $a = \lambda t$ and $\alpha = n$, we have the following equality iff X gamma (α, λ)

$$P(X \le t) = \frac{1}{\Gamma(n)}\int_0^{\lambda t} u^{n-1}d^{-u}\,du = e^{-\lambda t}\sum_{k=n}^\infty \frac{(\lambda t)^k}{k!} \tag{7.2.7}$$

Now

$$P(Y \ge n) = e^{-\lambda t}\sum_{k=n}^\infty \frac{(\lambda t)^k}{k!} \tag{7.2.8}$$

iff Y Poisson (λt.

The Gaussian (normal) approximation

The central limit theorem, referred to in the discussion of the Gaussian or normal distribution above, suggests that the binomial and Poisson distributions should be approximated by the Gaussian. The number of successes in *n* trials has the binomial (*n*,*p*) distribution. This random variable may be expressed

$$X = \sum_{i=1}^{n} I_{E_i} \tag{7.2.9}$$

Since the mean value of X is np and the variance is npq, the distribution should be approximately $N(np, npq)$.

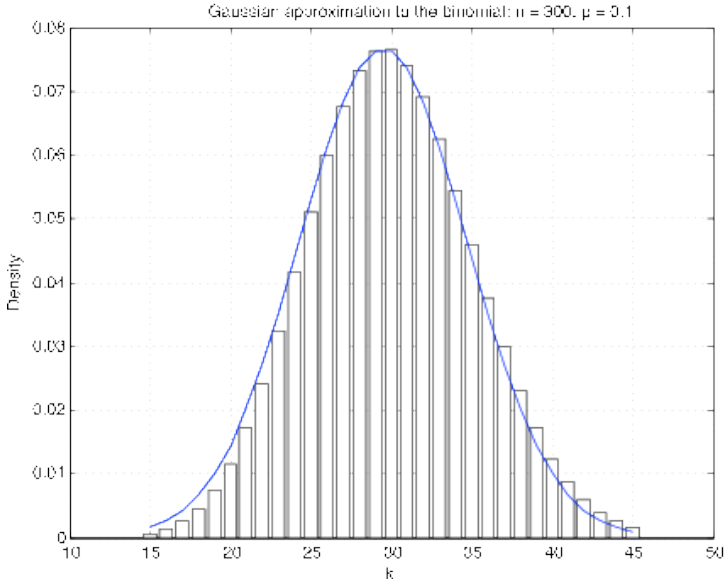


Figure 7.2.8. Gaussian approximation to the binomial.

Use of the generating function shows that the sum of independent Poisson random variables is Poisson. Now if X $)Poisson(\backslash(\mu$), then X may be considered the sum of n independent random variables, each Poisson (μ/n). Since the mean value and the variance are both μ, it is reasonable to suppose that suppose that X is approximately $N(\mu, \mu)$.

It is generally best to compare distribution functions. Since the binomial and Poisson distributions are integer-valued, it turns out that the best Gaussian approximation is obtained by making a "continuity correction." To get an approximation to a density for an integer-valued random variable, the probability at $t = k$ is represented by a rectangle of height p_k and unit width, with k as the midpoint. Figure 1 shows a plot of the "density" and the corresponding Gaussian density for $n = 300$, $p = 0.1$. It is apparent that the Gaussian density is offset by approximately 1/2. To approximate the probability $X \leq k$, take the area under the curve from k + 1/2; this is called the *continuity correction*.

Use of m-procedures to compare

We have two m-procedures to make the comparisons. First, we consider approximation of the

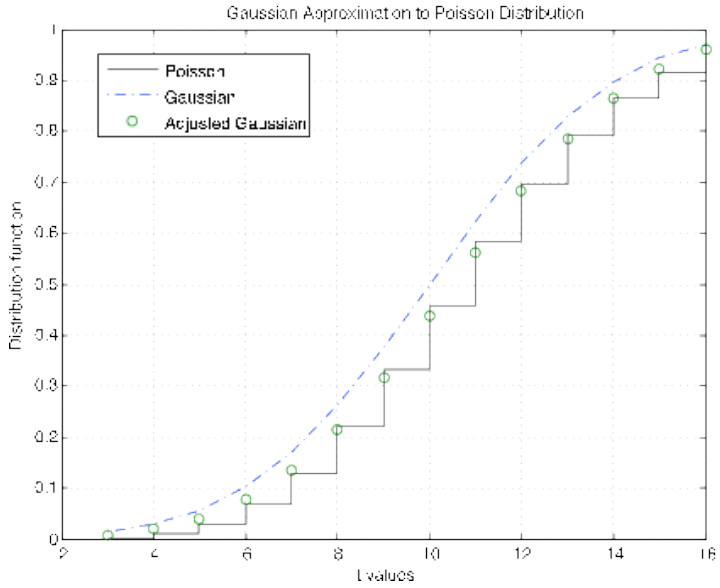


Figure 7.2.9. Gaussian approximation to the Poisson distribution function μ = 10.

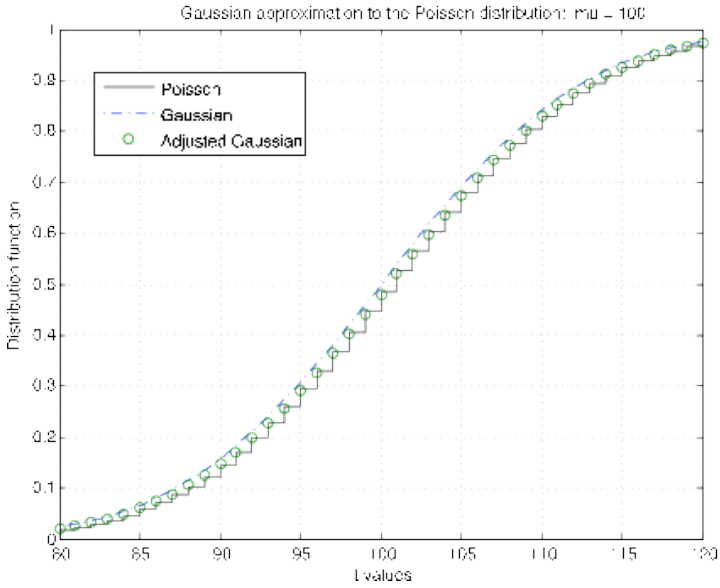


Figure 7.2.10. Gaussian approximation to the Poisson distribution function μ = 100.

Poisson (μ) distribution. The m-procedure *poissapp* calls for a value of μ, selects a suitable range about $k = \mu$ and plots the distribution function for the Poisson distribution (stairs) and the normal (Gaussian) distribution (dash dot) for $N(\mu, \mu)$. In addition, the continuity correction is applied to the gaussian distribution at integer values (circles). Figure 7.2.10 shows plots for μ = 10. It is clear that the continuity correction provides a much better approximation. The plots in Figure 7.2.11 are for μ = 100. Here the continuity correction provides the better approximation, but not by as much as for the smaller μ.

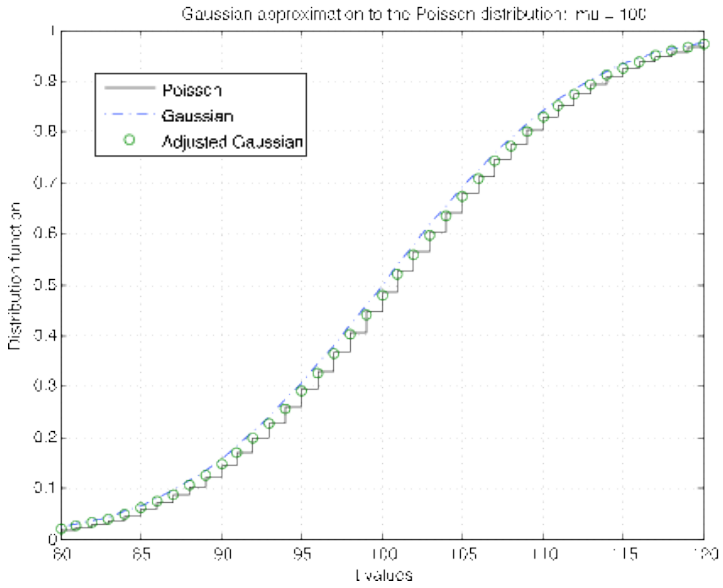


Figure 7.2.11. Poisson and Gaussian approximation to the binomial: n = 1000, p = 0.03.

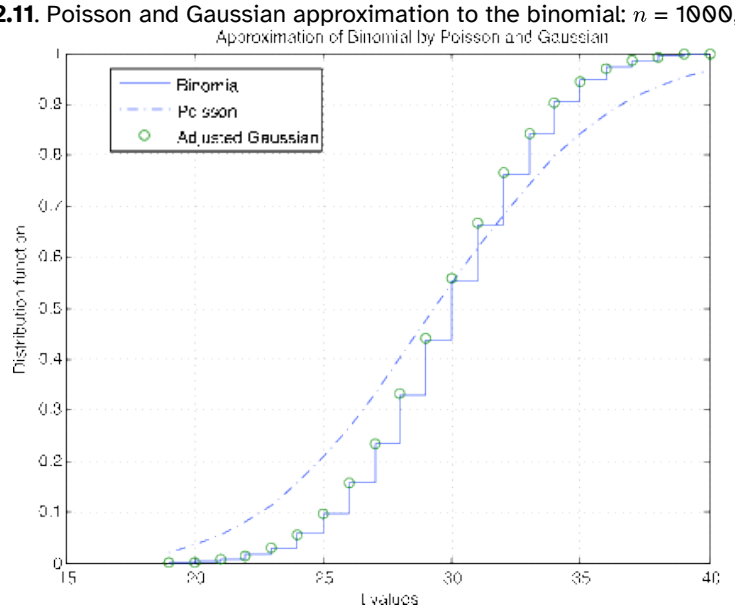


Figure 7.2.12. Poisson and Gaussian approximation to the binomial: n = 50, p = 0.6.

The m-procedure *bincomp* compares the binomial, gaussian, and Poisson distributions. It calls for values of n and p, selects suitable k values, and plots the distribution function for the binomial, a continuous approximation to the distribution function for the Poisson, and continuity adjusted values of the gaussian distribution function at the integer values. Figure 7.2.11 shows plots for $n = 1000$, $p = 0.03$. The good agreement of all three distribution functions is evident. Figure 7.2.12 shows plots for $n = 50, p = 0.6$. There is still good agreement of the binomial and adjusted gaussian. However, the Poisson distribution does not track very well. The difficulty, as we see in the unit Variance, is the difference in variances--npq for the binomial as compared with np for the Poisson.

Approximation of a real random variable by simple random variables

Simple random variables play a significant role, both in theory and applications. In the unit Random Variables, we show how a simple random variable is determined by the set of points on the real line representing the possible values and the corresponding set of probabilities that each of these values is taken on. This describes the distribution of the random variable and makes possible calculations of event probabilities and parameters for the distribution.

A continuous random variable is characterized by a set of possible values spread continuously over an interval or collection of

intervals. In this case, the probability is also spread smoothly. The distribution is described by a probability density function, whose value at any point indicates "the probability per unit length" near the point. A simple approximation is obtained by subdividing an interval which includes the range (the set of possible values) into small enough subintervals that the density is approximately constant over each subinterval. A point in each subinterval is selected and is assigned the probability mass in its subinterval. The combination of the selected points and the corresponding probabilities describes the distribution of an approximating simple random variable. Calculations based on this distribution approximate corresponding calculations on the continuous distribution.

Before examining a general approximation procedure which has significant consequences for later treatments, we consider some illustrative examples.

Example 7.2.10: Simple approximation to Poisson

A random variable with the Poisson distribution is unbounded. However, for a given parameter value μ, the probability for $k \geq n$, n sufficiently large, is negligible. Experiment indicates $n = \mu + 6\sqrt{\mu}$ (i.e., six standard deviations beyond the mean) is a reasonable value for $5 \leq \mu \leq 200$.

Solution

```
>> mu = [5 10 20 30 40 50 70 100 150 200];
>> K = zeros(1,length(mu));
>> p = zeros(1,length(mu));
>> for i = 1:length(mu)
     K(i) = floor(mu(i)+ 6*sqrt(mu(i)));
     p(i) = cpoisson(mu(i),K(i));
end
>> disp([mu;K;p*1e6]')
    5.0000   18.0000    5.4163  % Residual probabilities are 0.000001
   10.0000   28.0000    2.2535  % times the numbers in the last column.
   20.0000   46.0000    0.4540  % K is the value of k needed to achieve
   30.0000   62.0000    0.2140  % the residual shown.
   40.0000   77.0000    0.1354
   50.0000   92.0000    0.0668
   70.0000  120.0000    0.0359
  100.0000  160.0000    0.0205
  150.0000  223.0000    0.0159
  200.0000  284.0000    0.0133
```

An m-procedure for discrete approximation

If X is bounded, absolutely continuous with density functon f_X, the m-procedure *tappr* sets up the distribution for an approximating simple random variable. An interval containing the range of X is divided into a specified number of equal subdivisions. The probability mass for each subinterval is assigned to the midpoint. If dx is the length of the subintervals, then the integral of the density function over the subinterval is approximated by $f_X(t_i)dx$. where t_i is the midpoint. In effect, the graph of the density over the subinterval is approximated by a rectangle of length dx and height $f_X(t_i)$. Once the approximating simple distribution is established, calculations are carried out as for simple random variables.

Example 7.2.11: a numerical example

Suppose $f_X(t) = 3t^2$, $0 \leq t \leq 1$. Determine $P(0.2 \leq X \leq 0.9)$.

Solution

In this case, an analytical solution is easy. $F_X(t) = t^3$ on the interval [0, 1], so

$P = 0.9^3 - 0.2^3 = 0.7210$. We use tappr as follows.

```
>> tappr
Enter matrix [a b] of x-range endpoints  [0 1]
Enter number of x approximation points  200
Enter density as a function of t  3*t.^2
Use row matrices X and PX as in the simple case
>> M = (X >= 0.2)&(X <= 0.9);
>> p = M*PX'
p  =  0.7210
```

Because of the regularity of the density and the number of approximation points, the result agrees quite well with the theoretical value.

The next example is a more complex one. In particular, the distribution is not bounded. However, it is easy to determine a bound beyond which the probability is negligible.

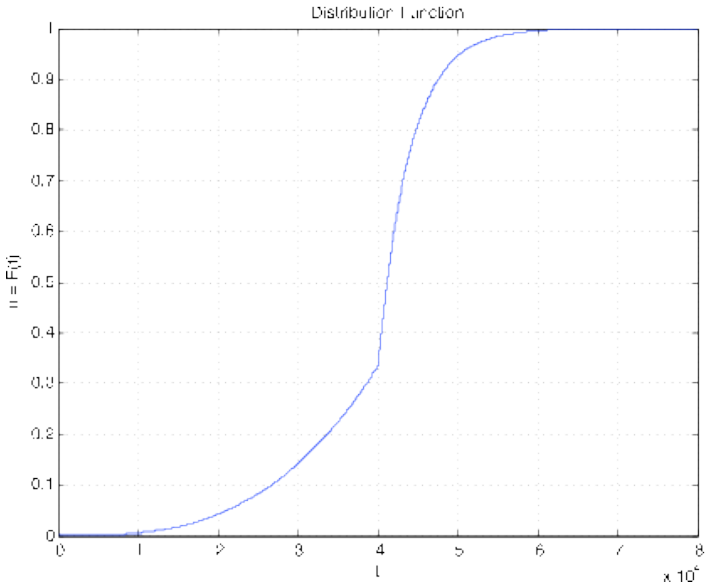


Figure 7.2.13. Distribution function for Example 7.2.12.

Example 7.2.12: Radial tire mileage

The life (in miles) of a certain brand of radial tires may be represented by a random variable X with density

$$f_X(t) = \begin{cases} t^2/a^3 & \text{for } \ 0 \le t < a \\ (b/a)e^{-k(t-a)} & \text{for } \ a \le t \end{cases}$$

where $a = 40,000$, $b = 20/3$, and $k = 1/4000$. Determine $P(X \ge 45,000$.

```
>> a = 40000;
>> b = 20/3;
>> k = 1/4000;
>> % Test shows cutoff point of 80000 should be satisfactory
>> tappr
Enter matrix [a b] of x-range endpoints  [0 80000]
Enter number of x approximation points  80000/20
```

```
Enter density as a function of t  (t.^2/a^3).*(t < 40000) + ...
(b/a)*exp(k*(a-t)).*(t >= 40000)
Use row matrices X and PX as in the simple case
>> P = (X >= 45000)*PX'
P   =  0.1910            % Theoretical value = (2/3)exp(-5/4) = 0.191003
>> cdbn
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX  % See Figure 7.2.14 for plot
```

In this case, we use a rather large number of approximation points. As a consequence, the results are quite accurate. In the single-variable case, designating a large number of approximating points usually causes no computer memory problem.

The general approximation procedure

We show now that any bounded real random variable may be approximated as closely as desired by a simple random variable (i.e., one having a finite set of possible values). For the unbounded case, the approximation is close except in a portion of the range having arbitrarily small total probability.

We limit our discussion to the bounded case, in which the range of X is limited to a bounded interval $I = [a, b]$. Suppose I is partitioned into n subintervals by points t_i, $1 \leq i \leq n-1$, with $a = t_0$ and $b = t_n$. Let $M_i = [t_{i-1}, t_i)$ be the ith subinterval, $1 \leq i \leq n-1$ and $M_n = [t_{n-1}, t_n]$ (see Figure 7.14). Now random variable X may map into any point in the interval, and hence into any point in each subinterval M_i. Let $E_i X^{-1}(M_i)$ be the set of points mapped into M_i by X. Then the E_i form a partition of the basic space Ω. For the given subdivision, we form a simple random variable X_s as follows. In each subinterval, pick a point s_i, $t_{i-1} \leq s_i \leq t_i$. Consider the simple random variable $X_s = \sum_{i=1}^{n} s_i I_{E_i}$.

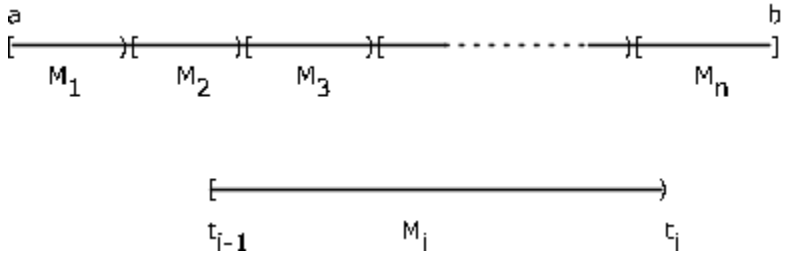


Figure 7.2.14. Partition of the interval I including the range of X

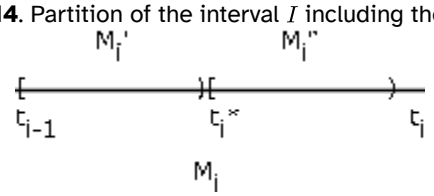


Figure 7.2.15. Refinement of the partition by additional subdividion points.

This random variable is in canonical form. If $\omega \in E_i$, then $X(\omega) \in M_i$ and $X_s(\omega) = s_i$. Now the absolute value of the difference satisfies

$$|X(\omega) - X_s(\omega)| < t_i - t_{i-1} \text{ the length of subinterval } M_i$$

Since this is true for each ω and the corresponding subinterval, we have the important fact

$$|X(\omega) - X_s(\omega)| < \text{maximum length of the } M_i$$

By making the subintervals small enough by increasing the number of subdivision points, we can make the difference as small as we please.

While the choice of the s_i is arbitrary in each M_i, the selection of $s_i = t_{i-1}$ (the left-hand endpoint) leads to the property $X_s(\omega) \leq X(\omega) \forall \omega$. In this case, if we add subdivision points to decrease the size of some or all of the M_i, the new simple approximation Y_s satisfies

$$X_s(\omega) = Y_s(\omega) \leq X(\omega) \ \forall \omega$$

To see this, consider $t_i^* \in M_i$(see Figure 7.15). M_i is partitioned into $M_i^{'} \bigcup M_i^{''}$ and E_i is partitioned into $E_i^{'} \bigcup E_i^{''}$. X maps $E_i^{'}$ into $M_i^{'}$ and $E_i^{''}$ into $M_i^{''}$. Y_s maps $E_i^{'}$ into t_i and maps $E_i^{''}$ into $t_i^{''}$ > t_i\). X_s maps both $E_i^{'}$ and $E_i^{''}$ into t_i. Thus, the asserted inequality must hold for each ω By taking a sequence of partitions in which each succeeding partition refines the previous (i.e. adds subdivision points) in such a way that the maximum length of subinterval goes to zero, we may form a nondecreasing sequence of simple random variables X_n which increase to X for each ω.

The latter result may be extended to random variables unbounded above. Simply let N th set of subdivision points extend from a to N, making the last subinterval $[N, \infty)$. Subintervals from a to N are made increasingly shorter. The result is a nondecreasing sequence $\{X_N : 1 \leq N\}$ of simple random variables, with $X_N(\omega) \to X(\omega)$ as $N \to \infty$, for each $\omega \in \Omega$.

For probability calculations, we simply select an interval I large enough that the probability outside I is negligible and use a simple approximation over I.

This page titled 7.2: Distribution Approximations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

7.3: Problems on Distribution and Density Functions

Exercise 7.3.1

(See Exercises 3 and 4 from "Problems on Random Variables and Probabilities"). The class $\{C_j : 1 \leq j \leq 10\}$ is a partition. Random variable X has values {1, 3, 2, 3, 4, 2, 1, 3, 5, 2} on C_1 through C_{10}, respectively, with probabilities 0.08, 0.13, 0.06, 0.09, 0.14, 0.11, 0.12, 0.07, 0.11, 0.09. Determine and plot the distribution function F_X.

Answer available at the end of this section.

Exercise 7.3.2

(See Exercise 6 from "Problems on Random Variables and Probabilities"). A store has eight items for sale. The prices are \$3.50, \$5.00, \$3.50, \$7.50, \$5.00, \$5.00, \$3.50, and \$7.50, respectively. A customer comes in. She purchases one of the items with probabilities 0.10, 0.15, 0.15, 0.20, 0.10 0.05, 0.10 0.15. The random variable expressing the amount of her purchase may be written

$$X = 3.5I_{C_1} + 5.0I_{C_2} + 3.5I_{C_3} + 7.5I_{C_4} + 5.0I_{C_5} + 5.0I_{C_6} + 3.5I_{C_7} + 7.5I_{C_8}$$

Determine and plot the distribution function for X.

Answer available at the end of this section.

Exercise 7.3.3

(See Exercise 12 from "Problems on Random Variables and Probabilities"). The class $\{A, B, C, D\}$ has minterm probabilities

$$pm = 0.001 * [5\ 7\ 6\ 8\ 9\ 14\ 22\ 33\ 21\ 32\ 50\ 75\ 86\ 129\ 201\ 302]$$

Determine and plot the distribution function for the random variable $X = I_A + I_B + I_C + I_D$, which counts the number of the events which occur on a trial.

Answer available at the end of this section.

Exercise 7.3.4

Suppose *a* is a ten digit number. A wheel turns up the digits 0 through 9 with equal probability on each spin. On ten spins what is the probability of matching, in order, *k* or more of the ten digits in a, $0 \leq k \leq 10$? Assume the initial digit may be zero.

Answer available at the end of this section.

Exercise 7.3.5

In a thunderstorm in a national park there are 127 lightning strikes. Experience shows that the probability of of a lightning strike starting a fire is about 0.0083. What is the probability that k fires are started, $k = 0$,1,2,3?

Answer available at the end of this section.

Exercise 7.3.6

A manufacturing plant has 350 special lamps on its production lines. On any day, each lamp could fail with probability $p = 0.0017$. These lamps are critical, and must be replaced as quickly as possible. It takes about one hour to replace a lamp, once it has failed. What is the probability that on any day the loss of production time due to lamp failaures is k or fewer hours, $k = 0, 1, 2, 3, 4, 5$?

Answer available at the end of this section.

Exercise 7.3.7

Two hundred persons buy tickets for a drawing. Each ticket has probability 0.008 of winning. What is the probability of k or fewer winners, $k = 2, 3, 4$?

Answer available at the end of this section.

Exercise 7.3.8

Two coins are flipped twenty times. What is the probability the results match (both heads or both tails) k times, $0 \leq k \leq 20$?

Answer available at the end of this section.

Exercise 7.3.9

Thirty members of a class each flip a coin ten times. What is the probability that at least five of them get seven or more heads?

Answer available at the end of this section.

Exercise 7.3.10

For the system in Exercise 6, call a day in which one or more failures occur among the 350 lamps a "service day." Since a Bernoulli sequence "starts over" at any time, the sequence of service/nonservice days may be considered a Bernoulli sequence with probability p_1, the probability of one or more lamp failures in a day.

a. Beginning on a Monday morning, what is the probability the first service day is the first, second, third, fourth, fifth day of the week?
b. What is the probability of no service days in a seven day week?

Answer

p1 = 1 - (1 - 0.0017)^350 = 0.4487 k = 1:5; (prob given day is a service day)

a.
```
P = p1*(1 - p1).^(k-1) = 0.4487  0.2474  0.1364  0.0752  0.0414
```

b.
```
P0 = (1 - p1)^7 = 0.0155
```

Exercise 7.3.11

For the system in Exercise 6 and Exercise 10 assume the plant works seven days a week. What is the probability the third service day occurs by the end of 10 days? Solve using the negative binomial distribution; repeat using the binomial distribution.

Answer available at the end of this section.

Exercise 7.3.12

A residential College plans to raise money by selling "chances" on a board. Fifty chances are sold. A player pays \$10 to play; he or she wins \$30 with probability $p = 0.2$. The profit to the College is

$$X = 50 \cdot 10 - 30N, \text{ where } N \text{ is the number of winners}$$

Determine the distribution for X and calculate $P(X > 0)$, $P(X \geq 200)$, and $P(X \geq 300)$

Answer available at the end of this section.

Exercise 7.3.13

A single six-sided die is rolled repeatedly until either a one or a six turns up. What is the probability that the first appearance of either of these numbers is achieved by the fifth trial or sooner?

Answer available at the end of this section.

Exercise 7.3.14

Consider a Bernoulli sequence with probability $p = 0.53$ of success on any component trial.

a. The probability the fourth success will occur no later than the tenth trial is determined by the negative binomial distribution. Use the procedure nbinom to calculate this probability .
b. Calculate this probability using the binomial distribution.

Answer

a. `P = sum(nbinom(4,0.53,4:10)) = 0.8729`
b. `Pa = cbinom(10,0.53,4) = 0.8729`

Exercise 7.3.15

Fifty percent of the components coming off an assembly line fail to meet specifications for a special job. It is desired to select three units which meet the stringent specifications. Items are selected and tested in succession. Under the usual assumptions for Bernoulli trials, what is the probability the third satisfactory unit will be found on six or fewer trials?

Answer available at the end of this section.

Exercise 7.3.16

The number of cars passing a certain traffic count position in an hour has Poisson (53) distribution. What is the probability the number of cars passing in an hour lies between 45 and 55 (inclusive)? What is the probability of *more* than 55?

Answer available at the end of this section.

Exercise 7.3.17

Compare $P(X \leq k)$ and $P(Y \leq k)$ for X binomial(5000, 0.001) and Y Poisson (5), for $0 \leq k \leq 10$. Do this directly with ibinom and ipoisson. Then use the m-procedure bincomp to obtain graphical results (including a comparison with the normal distribution).

Answer available at the end of this section.

Exercise 7.3.18

Suppose X binomial (12, 0.375), Y Poisson (4.5), and Z exponential (1/4.5). For each random variable, calculate and tabulate the probability of a value at least k, for integer values $3 \leq k \leq 8$.

Answer available at the end of this section.

Exercise 7.3.19

The number of noise pulses arriving on a power circuit in an hour is a random quantity having Poisson (7) distribution. What is the probability of having at least 10 pulses in an hour? What is the probability of having at most 15 pulses in an hour?

Answer available at the end of this section.

Exercise 7.3.20

The number of customers arriving in a small specialty store in an hour is a random quantity having Poisson (5) distribution. What is the probability the number arriving in an hour will be between three and seven, inclusive? What is the probability of no more than ten?

Answer available at the end of this section.

Exercise 7.3.21

Random variable X binomial (1000, 0.1).

a. Determine $P(X \geq 80)$, $P(X \geq 100)$, $P(X \geq 120)$
b. Use the appropriate Poisson distribution to approximate these values.

Answer

```
k = [80 100 120];
P = cbinom(1000,0.1,k)
P  =  0.9867    0.5154    0.0220
P1 = cpoisson(100,k)
P1 =  0.9825    0.5133    0.0282
```

Exercise 7.3.22

The time to failure, in hours of operating time, of a televesion set subject to random voltage surges has the exponential (0.002) distribution. Suppose the unit has operated successfully for 500 hours. What is the (conditional) probability it will operate for another 500 hours?

Answer available at the end of this section.

Exercise 7.3.23

For X exponential (λ), determine $P(X \geq 1/\lambda)$, $P(X \geq 2/\lambda)$.

Answer available at the end of this section.

Exercise 7.3.24

Twenty "identical" units are put into operation. They fail independently. The times to failure (in hours) form an iid class, exponential (0.0002). This means the "expected" life is 5000 hours. Determine the probabilities that at least k, for k = 5,8,10,12,15, will survive for 5000 hours.

Answer available at the end of this section.

Exercise 7.3.25

Let T gamma (20, 0.0002) be the total operating time for the units described in Exercise 24.

a. Use the m-function for the gamma distribution to determine $P(T \leq 100,000)$.
b. Use the Poisson distribution to determine $P(T \leq 100,000)$.

Answer

```
P1 = gammadbn(20,0.0002,100000) = 0.5297 P2 = cpoisson(0.0002*100000,20) = 0.5297
```

Exercise 7.3.26

The sum of the times to failure for five independent units is a random variable X gamma (5, 0.15). Without using tables or m-programs, determine $P(Xle25)$.

Answer available at the end of this section.

Exercise 7.3.27

Interarrival times (in minutes) for fax messages on a terminal are independent, exponential ($\lambda = 0.1$). This means the time X for the arrival of the fourth message is gamma(4, 0.1). Without using tables or m-programs, utilize the relation of the gamma to the Poisson distribution to determine $P \leq 30$.

Answer available at the end of this section.

Exercise 7.3.28

Customers arrive at a service center with independent interarrival times in hours, which have exponential (3) distribution. The time X for the third arrival is thus gamma (3, 3). Without using tables or m-programs, determine $P(X \leq 2)$.

Answer available at the end of this section.

Exercise 7.3.29

Five people wait to use a telephone, currently in use by a sixth person. Suppose time for the six calls (in minutes) are iid, exponential (1/3). What is the distribution for the total time Z from the present for the six calls? Use an appropriate Poisson distribution to determine $P(Z \leq 20)$.

Answer available at the end of this section.

Exercise 7.3.30

A random number generator produces a sequence of numbers between 0 and 1. Each of these can be considered an observed value of a random variable uniformly distributed on the interval [0, 1]. They assume their values independently. A sequence of 35 numbers is generated. What is the probability 25 or more are less than or equal to 0.71? (Assume continuity. Do not make a discrete adjustment.)

Answer available at the end of this section.

Exercise 7.3.31

Five "identical" electronic devices are installed at one time. The units fail independently, and the time to failure, in days, of each is a random variable exponential (1/30). A maintenance check is made each fifteen days. What is the probability that at least four are still operating at the maintenance check?

Answer available at the end of this section.

Exercise 7.3.32

Suppose X $N(4, 81)$. That is, X has gaussian distribution with mean μ = 4 and variance σ^2 = 81.

a. Use a table of standardized normal distribution to determine $P(2 < X < 8)$ and $P(|X - 4| \leq 5)$.
b. Calculate the probabilities in part (a) with the m-function gaussian.

Answer

a.

$$P(2 < X < 8) = \phi((8-4)/9) - \phi((2-4)/9) =$$

$$\phi(4/9) + \phi(2/9) - 1 = 0.6712 + 0.5875 - 1 = 0.2587$$

$$P(|X - 4| \leq 5) = 2\phi(5/9) - 1 = 1.4212 - 1 = 0.4212$$

b.

```
P1 = gaussian(4,81,8) - gaussian(4,81,2)
```

```
P1 = 0.2596
P2 = gaussian(4,81,9) - gaussian(4,84,-1)
P2 = 0.4181
```

Exercise 7.3.33

Suppose X $N(5, 81)$. That is, X has gaussian distribution with μ = 5 and σ^2 = 81. Use a table of standardized normal distribution to determine $P(3 < X < 9)$ and $P(|X - 5|le5)$. Check your results using the m-function gaussian.

Answer available at the end of this section.

Exercise 7.3.34

Suppose X $N(3, 64)$. That is, X has gaussian distribution with μ = 3 and σ^2 = 64. Use a table of standardized normal distribution to determine $P(1 < X < 9)$ and $P(|X - 3|le4)$. Check your results with the m-function gaussian.

Answer available at the end of this section.

Exercise 7.3.35

Items coming off an assembly line have a critical dimension which is represented by a random variable N (10, 0.01). Ten items are selected at random. What is the probability that three or more are within 0.05 of the mean value μ.

Answer available at the end of this section.

Exercise 7.3.36

The result of extensive quality control sampling shows that a certain model of digital watches coming off a production line have accuracy, in seconds per month, that is normally distributed with μ = 5 and σ^2 = 300. To achieve a top grade, a watch must have an accuracy within the range of -5 to +10 seconds per month. What is the probability a watch taken from the production line to be tested will achieve top grade? Calculate, using a standardized normal table. Check with the m-function gaussian.

Answer available at the end of this section.

Exercise 7.3.37

Use the m-procedure bincomp with various values of n from 10 to 500 and p from 0.01 to 0.7, to observe the approximation of the binomial distribution by the Poisson.

Answer available at the end of this section.

Exercise 7.3.38

Use the m-procedure poissapp to compare the Poisson and gaussian distributions. Use various values of μ from 10 to 500.

Answer available at the end of this section.

Exercise 7.3.39

Random variable X has density $f_X(t) = \frac{3}{2}t^2$, $-1 \le t \le 1$ (and zero elsewhere).

a. Determine $P(-0.5 \le X < 0.8)$, $P(|X| > 0.5)$, $P(|X - 0.25) \le 0.5)$.
b. Determine an expression for the distribution function.
c. Use the m-procedures tappr and cdbn to plot an approximation to the distribution function.

Answer

a. $\frac{3}{2}\int t^2 = t^3/2$

$$P1 = 0.5 * (0.8^3 - (-0.5)^3) = 0.3185 \; P2 = 2\int_{0.5}^{1} \frac{3}{2}t^2 = (1 - (-0.5)^3) = 7/8$$

$$P3 = P(|X - 0.25| \le 0.5) = P(-0.25 \le X \le 0.75) = \frac{1}{2}[(3/4)^3 - (-1/4)^3] = 7/32$$

b. $F_X(t) = \int_{-1}^{1} f_X = \frac{1}{2}(t^3 + 1)$

```
tappr
Enter matrix [a b] of x-range endpoints  [-1 1]
Enter number of x approximation points  200
Enter density as a function of t  1.5*t.^2
Use row matrices X and PX as in the simple case
cdbn
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX    % See MATLAB plot
```

Exercise 7.3.40

Random variable X has density function $f_X(t) = t - \frac{3}{8}t^3$, $0 \le t \le 2$ (and zero elsewhere).

a. Determine $P(X \le 0.5)$, $P(0.5 \le X < 1.5)$, $P(|X - 1| < 1/4)$.
b. Determine an expression for the distribution function.
c. Use the m-procedures tappr and cdbn to plot an approximation to the distribution function.

Answer

a. $\int (t - \frac{3}{8}t^2) = \frac{t^2}{2} - \frac{t^3}{8}$

$$P1 = 0.5^2/2 - 0.5^3/8 = 7/64 \; P2 = 1.5^2/2 - 1.5^3/8 - 7/64 = 19/32 \; P3 = 79/256)$$

b. $F_X(t) = \frac{t^2}{2} - \frac{t^3}{8}$, $0 \le t \le 2$

```
tappr
Enter matrix [a b] of x-range endpoints  [0 2]
Enter number of x approximation points  200
Enter density as a function of t  t - (3/8)*t.^2
Use row matrices X and PX as in the simple case
cdbn
Enter row matrix of VALUES  X
```

```
08 Enter row matrix of PROBABILITIES  PX    % See MATLAB plot
```

Exercise 7.3.41

Random variable X has density function

$$f_X(t) = \begin{cases} (6/5)t^2 & \text{for } 0 \le t \le 1 \\ (6/5)(2-t) & \text{for } 1 < t \le 2 \end{cases} = I[0,1](t)\frac{6}{5}t^2 + I_{(1,2]}(t)\frac{6}{5}(2-t)$$

a. Determine $P(X \le 0.5)$, $P(0.5 \le X < 1.5)$, $P(|X-1| < 1/4)$.
b. Determine an expression for the distribution function.
c. Use the m-procedures tappr and cdbn to plot an approximation to the distribution function.

Answer available at the end of this section.

Answers

Exercise 7.3.1

```
01 T = [1 3 2 3 4 2 1 3 5 2];
02 pc = 0.01*[8 13 6 9 14 11 12 7 11 9];
03 [X,PX] = csort(T,pc);
04 ddbn
05 Enter row matrix of VALUES  X
06 Enter row matrix of PROBABILITIES  PX    % See MATLAB plot
```

Exercise 7.3.2

```
01 T = [3.5 5 3.5 7.5 5 5 3.5 7.5];
02 pc = 0.01*[10 15 15 20 10 5 10 15];
03 [X,PX] = csort(T,pc);
04 ddbn
05 Enter row matrix of VALUES  X
06 Enter row matrix of PROBABILITIES  PX    % See MATLAB plot
```

Exercise 7.3.3

```
01 npr06_12
02 Minterm probabilities in pm, coefficients in c
03 T = sum(mintable(4)); % Alternate solution.  See Exercise 6.2.12 from "Problems on Random
   Variables and Probabilities"
04 [X,PX] = csort(T,pm);
05 ddbn
06 Enter row matrix of VALUES  X
07 Enter row matrix of PROBABILITIES  PX    % See MATLAB plot
```

Exercise 7.3.4

$P =$ cbinom(10, 0.1, 0 : 10).

Exercise 7.3.5

```
P = ibinom(127,0.0083,0:3) P = 0.3470 0.3688 0.1945 0.0678
```

Exercise 7.3.6

P = 1 - chinom(350, 0.0017, 1:6)

```
01 =  0.5513    0.8799    0.9775    0.9968    0.9996    1.0000
```

Exercise 7.3.7

P = 1 - cbinom(200,0.008,3:5) = 0.7838 0.9220 0.9768

Exercise 7.3.8

```
P = ibinom(20,1/2,0:20)
```

Exercise 7.3.9

p = cbinom(10,0.5,7) = 0.1719

```
P = cbinom(30,p,5) = 0.6052
```

Exercise 7.3.11

p1 = 1 - (1 - 0.0017)^350 = 0.4487

- P = sum(nbinom(3,p1,3:10)) = 0.8990
- Pa = cbinom(10,p1,3) = 0.8990

Exercise 7.3.12

```
N = 0:50;
PN = ibinom(50,0.2,0:50);
X = 500 - 30*N;
Ppos = (X>0)*PN'
Ppos =  0.9856
P200 = (X>=200)*PN'
P200 =  0.5836
P300 = (X>=300)*PN'
P300 =  0.1034
```

Exercise 7.3.13

P = 1 - (2/3)^5 = 0.8683

Exercise 7.3.15

P = cbinom(6,0.5,3) = 0.6562

Exercise 7.3.16

P1 = cpoisson(53,45) - cpoisson(53,56) = 0.5224

```
P2 = cpoisson(53,56) = 0.3581
```

Exercise 7.3.17

```
k = 0:10;
Pb = 1 - cbinom(5000,0.001,k+1);
Pp = 1 - cpoisson(5,k+1);
disp([k;Pb;Pp]')
         0    0.0067    0.0067
    1.0000    0.0404    0.0404
    2.0000    0.1245    0.1247
    3.0000    0.2649    0.2650
    4.0000    0.4404    0.4405
    5.0000    0.6160    0.6160
    6.0000    0.7623    0.7622
    7.0000    0.8667    0.8666
    8.0000    0.9320    0.9319
    9.0000    0.9682    0.9682
   10.0000    0.9864    0.9863
```

```
bincomp
Enter the parameter n  5000
Enter the parameter p  0.001
Binomial-- stairs
Poisson--  -.-.
Adjusted Gaussian-- o o o
gtext('Exercise 17')
```

Exercise 7.3.18

```
k = 3:8;
Px = cbinom(12,0.375,k);
```

```
Py = cpoisson(4.5,k);
Pz = exp(-k/4.5);
disp([k;Px;Py;Pz]')
    3.0000    0.8865    0.8264    0.5134
    4.0000    0.7176    0.6577    0.4111
    5.0000    0.4897    0.4679    0.3292
    6.0000    0.2709    0.2971    0.2636
    7.0000    0.1178    0.1689    0.2111
    8.0000    0.0390    0.0866    0.1690
```

Exercise 7.3.19

P1 = cpoisson(7,10) = 0.1695 P2 = 1 - cpoisson(7,16) = 0.9976

Exercise 7.3.20

P1 = cpoisson(5,3) - cpoisson(5,8) = 0.7420

```
P2 = 1 - cpoisson(5,11) = 0.9863
```

Exercise 7.3.22

$P(X > 500 + 500 | X > 500) = P(X > 500) = e^{-0.002 \cdot 500} = 0.3679$

Exercise 7.3.23

$P(X > k\lambda) = e^{-\lambda k/\lambda} = e^{-k}$

Exercise 7.3.24

```
p = exp(-0.0002*5000)
p = 0.3679
k = [5 8 10 12 15];
P = cbinom(20,p,k)
P = 0.9110  0.4655  0.1601  0.0294  0.0006
```

Exercise 7.3.26

$$P(X \leq 25) = P(Y \geq 5), Y \text{ Poisson } (0.15 \cdot 25 = 3.75)$$
$$P(Y \geq 5) = 1 - P(Y \leq 4) = 1 - e^{-3.35}(1 + 3.75 + \frac{3.75^2}{2} + \frac{3.75^3}{3!} + \frac{3.75^4}{24}) = 0.3225$$

Exercise 7.3.27

$$P(X \leq 30) = P(Y \geq 4), Y \text{ poisson } (0.2 \cdot 30 = 3)$$
$$P(Y \geq 4) = 1 - P(Y \leq 3) = 1 - e^{-3}(1 + 3 + \frac{3^2}{2} + \frac{3^3}{3!}) = 0.3528$$

Exercise 7.3.28

$$P(X \leq 2) = P(Y \geq 3), Y \text{ poisson } (3 \cdot 2 = 6)$$
$$P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - e^{-6}(1 + 6 + 36/2) = 0.9380$$

Exercise 7.3.29

Z gamma (6, 1/3).

$$P(Z \leq 20) = P(Y \geq 6), Y \text{ poisson}(1/3 \cdot 20)$$
$$P(Y \geq 6) = \text{cpoisson}(20/3, 6) = 0.6547$$

Exercise 7.3.30

p = cbinom(35,0.71,25) = 0.5620

Exercise 7.3.31

p = exp(-15/30) = 0.6065 P = cbinom(5,p,4) = 0.3483

Exercise 7.3.33

$$P(3 < X < 9) = \phi((9-5)/9) - \phi((3-5)/9) = \phi(4/9) + \phi(2/9) - 1 = 0.6712 + 0.5875 - 1 = 0.2587$$
$$P(|X - 5| \leq 5) = 2\phi(5/9) - 1 = 1.4212 - 1 = 0.4212$$

```
P1 = gaussian(5,81,9) - gaussian(5,81,3)
P1 = 0.2596
P2 = gaussian(5,81,10) - gaussian(5,84,0)
P2 = 0.4181
```

Exercise 7.3.34

$$P(1 < X < 9) = \phi((9-3)/8) - \phi(1-3)/9) =$$
$$\phi(0.75) + \phi(0.25) - 1 = 0.7734 + 0.5987 - 1 = 0.3721$$
$$P(|X - 3| \leq 4) = 2\phi(4/8) - 1 = 1.3829 - 1 = 0.3829$$

```
P1 = gaussian(3,64,9) - gaussian(3,64,1)
P1 = 0.3721
P2 = gaussian(3,64,7) - gaussian(3,64,-1)
P2 = 0.3829
```

Exercise 7.3.35

```
p = gaussian(10,0.01,10.05) - gaussian(10,0.01,9.95)
p =  0.3829
P = cbinom(10,p,3)
P =  0.8036
```

Exercise 7.3.36

$P(-5 \le X \le 10) = \phi(5/\sqrt{300}) + \phi(10/\sqrt{300}) - 1 = \phi(0.289) + \phi(0.577) - 1 = 0.614 + 0.717 - 1 = 0.331$

$P =$ gaussian(5, 300, 10) - gaussian(5, 300, -5) = 0.3317

Exercise 7.3.37

Experiment with the m-procedure bincomp.

Exercise 7.3.38

Experiment with the m-procedure poissapp.

Exercise 7.3.41

a. $P1 = \frac{6}{5}\int_0^{1/2} t^2 = 1/20$ $P2 = \frac{6}{5}\int_{1/2}^{1} t^2 + \frac{6}{5}\int_1^{3/2}(2-t) = 4/5$

$$P3) = \frac{6}{5}\int_{3/4}^{1} t^2 + \frac{6}{5}\int_1^{5/4}(2-t) = 79/160$$

b. $F_X(t) = \int_0^1 f_X = I_{[0,1]}(t)\frac{2}{5}t^3 + I_{(1,2]}(t)[-\frac{7}{5} + \frac{6}{5}(2t - \frac{t^2}{2})]$

c.

```
tappr
Enter matrix [a b] of x-range endpoints  [0 2]
Enter number of x approximation points  400
Enter density as a function of t  (6/5)*(t<=1).*t.^2 + ...
      (6/5)*(t>1).*(2 - t)
Use row matrices X and PX as in the simple case
cdbn
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX    % See MATLAB plot
```

This page titled 7.3: Problems on Distribution and Density Functions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

8: RANDOM VECTORS AND JOINT DISTRIBUTIONS

8.1: Random Vectors and Joint Distributions
8.2: Random Vectors and MATLAB
8.3: Problems on Random Vectors and Joint Distributions

This page titled 8: Random Vectors and Joint Distributions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Random Vectors and Joint Distributions

A single, real-valued random variable is a function (mapping) from the basic space Ω to the real line. That is, to each possible outcome ω of an experiment there corresponds a real value $t = X(\omega)$. The mapping induces a probability mass distribution on the real line, which provides a means of making probability calculations. The distribution is described by a distribution function F_X. In the absolutely continuous case, with no point mass concentrations, the distribution may also be described by a probability density function f_X. The probability density is the linear density of the probability mass along the real line (i.e., mass per unit length). The density is thus the derivative of the distribution function. For a simple random variable, the probability distribution consists of a point mass p_i at each possible value t_i of the random variable. Various m-procedures and m-functions aid calculations for simple distributions. In the absolutely continuous case, a simple approximation may be set up, so that calculations for the random variable are approximated by calculations on this simple distribution.

Often we have more than one random variable. Each can be considered separately, but usually they have some probabilistic ties which must be taken into account when they are considered jointly. We treat the joint case by considering the individual random variables as *coordinates of a random vector*. We extend the techniques for a single random variable to the multidimensional case. To simplify exposition and to keep calculations manageable, we consider a pair of random variables as coordinates of a two-dimensional random vector. The concepts and results extend directly to any finite number of random variables considered jointly.

Random variables considered jointly; random vectors

As a starting point, consider a simple example in which the probabilistic interaction between two random quantities is evident.

Example 8.1.1: A selection problem

Two campus jobs are open. Two juniors and three seniors apply. They seem equally qualified, so it is decided to select them by chance. Each combination of two is equally likely. Let X be the number of juniors selected (possible values 0, 1, 2) and Y be the number of seniors selected (possible values 0, 1, 2). However there are only three possible pairs of values for (X, Y): (0, 2), (1, 1), or (2, 0). Others have zero probability, since they are impossible. Determine the probability for each of the possible pairs.

Solution

There are $C(5, 2) = 10$ equally likely pairs. Only one pair can be both juniors. Six pairs can be one of each. There are $C(3, 2) = 3$ ways to select pairs of seniors. Thus

$$P(X = 0, Y = 2) = 3/10,\ P(X = 1, Y = 1) = 6/10,\ P(X = 2, Y = 0) = 1/10$$

These probabilities add to one, as they must, since this exhausts the mutually exclusive possibilities. The probability of any other combination must be zero. We also have the distributions for the random variables conisidered individually.

$$X = [0\ 1\ 2]\ PX = [3/10\ 6/10\ 1/10]\ Y = [0\ 1\ 2]\ PY = [1/10\ 6/10\ 3/10]$$

We thus have a *joint distribution* and two individual or *marginal distributions*.

We formalize as follows:

A pair $\{X, Y\}$ of random variables considered jointly is treated as the pair of coordinate functions for a two-dimensional random vector $W = (X, Y)$. To each $\omega \in \Omega$, W assigns the pair of real numbers (t, u), where $X(\omega) = t$ and $Y(\omega) = u$. If we represent the pair of values $\{t, u\}$ as the point (t, u) on the plane, then $W(\omega) = (t, u)$, so that

$$W = (X, Y) : \Omega \to \mathbf{R}^2$$

is a mapping from the basic space Ω to the plane R^2. Since W is a function, all mapping ideas extend. The inverse mapping W^{-1} plays a role analogous to that of the inverse mapping X^{-1} for a real random variable. A two-dimensional vector *W* is a *random*

vector iff $W^{-1}(Q)$ is an event for each reasonable set (technically, each Borel set) on the plane.

A fundamental result from measure theory ensures

$W = (X, Y)$ is a random vector iff each of the coordinate functions X and Y is a random variable.

In the selection example above, we model X (the number of juniors selected) and Y (the number of seniors selected) as random variables. Hence the vector-valued function

Induced distribution and the joint distribution function

In a manner parallel to that for the single-variable case, we obtain a mapping of probability mass from the basic space to the plane. Since $W^{-1}(Q)$ is an event for each reasonable set Q on the plane, we may assign to Q the probability mass

$$P_{XY}(Q) = P[W^{-1}(Q)] = P[(X, Y)^{-1}(Q)]$$

Because of the preservation of set operations by inverse mappings as in the single-variable case, the mass assignment determines P_{XY} as a probability measure on the subsets of the plane R^2. The argument parallels that for the single-variable case. The result is the *probability distribution induced by* $W = (X, Y)$. To determine the probability that the vector-valued function $W = (X, Y)$ takes on a (vector) value in region Q, we simply determine how much induced probability mass is in that region.

Example 8.1.2: Induced distribution and probability calculations

To determine $P(1 \leq X \leq, Y > 0)$, we determine the region for which the first coordinate value (which we call t) is between one and three and the second coordinate value (which we call u) is greater than zero. This corresponds to the set Q of points on the plane with $1 \leq t \leq 3$ and $u > 0$. Geometrically, this is the strip on the plane bounded by (but not including) the horizontal axis and by the vertical lines $t = 1$ and $t = 3$(included). The problem is to determine how much probability mass lies in that strip. How this is achieved depends upon the nature of the distribution and how it is described.

As in the single-variable case, we have a distribution function.

Definition: Joint Distribution Function

The **joint distribution function** F_{XY} for $W = (X, Y)$ is given by

$$F_{XY}(t, u) = P(X \leq t, Y \leq u) \quad \forall (t, u) \in R^2 \tag{8.1.1}$$

This means that $F_{XY}(t, u)$ is equal to the probability mass in the region Q_{tu} on the plane such that the first coordinate is less than or equal to t and the second coordinate is less than or equal to u. Formally, we may write

$$F_{XY}(t, u) = P[(X, Y) \in Q_{tu}], \text{ where } \backslash Q\_\{tu\} = \backslash\{(r, s) : r \backslash le\ t, s \backslash le\ u\backslash\}\backslash)$$

Now for a given point (a, b), the region Q_{ab} is the set of points (t, u) on the plane which are on or to the left of the vertical line through $(t,$ 0)*and* on or below the horizontal line through (0, u) (see Figure 1 for specific point $t = a, u = b$). We refer to such regions as semiinfinite intervals on the plane.

The theoretical result quoted in the real variable case extends to ensure that a distribution on the plane is determined uniquely by consistent assignments to the semiinfinite intervals Q_{tu}. Thus, the induced distribution is determined completely by the joint distribution function.

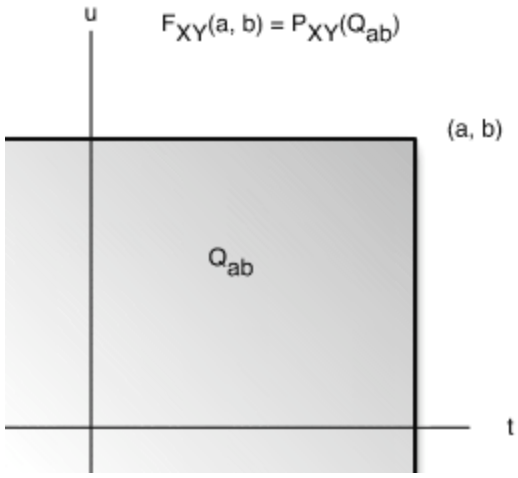


Figure 8.1.1. The region Q_{ab} for the value $F_{XY}(a, b)$.

Distribution function for a discrete random vector

The induced distribution consists of point masses. At point (t_i, u_j) in the range of $W = (X, Y)$ there is probability mass $P_{ij} = P[W = (t, u_j)] = P(X = t_i, Y = u_j)$. As in the general case, to determine $[P(X, Y) \in Q]$ we determine how much probability mass is in the region. In the discrete case (or in any case where there are point mass concentrations) one must be careful to note whether or not the boundaries are included in the region, should there be mass concentrations on the boundary.

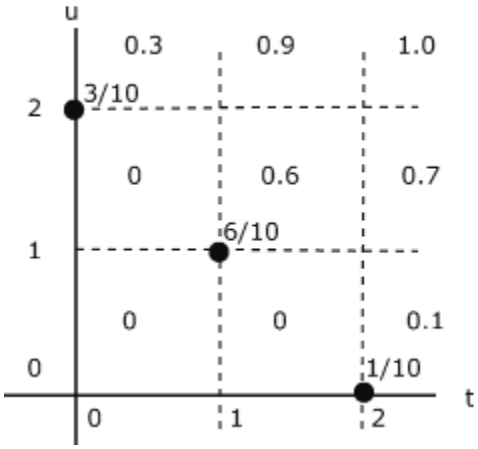


Figure 8.1.2. The joint distribution for Example **8.1.3**.

Example 8.1.3: distribution function for the selection problem in Example 8.1.1

The probability distribution is quite simple. Mass 3/10 at (0,2), 6/10 at (1,1), and 1/10 at (2,0). This distribution is plotted in Figure 8.2. To determine (and visualize) the joint distribution function, think of moving the point (t, u) on the plane. The region \Q_{tu}\) is a giant "sheet" with corner at $t, u)$. The value of $F_{XY}(t, u)$ is the amount of probability covered by the sheet. This value is constant over any grid cell, including the left-hand and lower boundariies, and is the value taken on at the lower left-hand corner of the cell. Thus, if (t, u) is in any of the three squares on the lower left hand part of the diagram, no probability mass is covered by the sheet with corner in the cell. If (t, u) is on or in the square having probability 6/10 at the lower left-hand corner, then the sheet covers that probability, and the value of $F_{XY}(t, u) = 6/10$. The situation in the other cells may be checked out by this procedure.

Distribution function for a mixed distribution

Example 8.1.4: A Mixed Distribution

The pair $\{X, Y\}$ produces a mixed distribution as follows (see Figure 8.3)

Point masses 1/10 at points (0,0), (1,0), (1,1), (0,1)

Mass 6/10 spread uniformly over the unit square with these vertices

The joint distribution function is zero in the second, third, and fourth quadrants.

- If the point (t, u) is in the square or on the left and lower boundaries, the sheet covers the point mass at (0,0) plus 0.6 times the area covered within the square. Thus in this region

$$F_{XY}(t, u) = \frac{1}{10}(1 + 6tu)$$

- If the pont (t, u) is above the square (including its upper boundary) but to the left of the line $t = 1$, the sheet covers two point masses plus the portion of the mass in the square to the left of the vertical line through (t, u). In this case

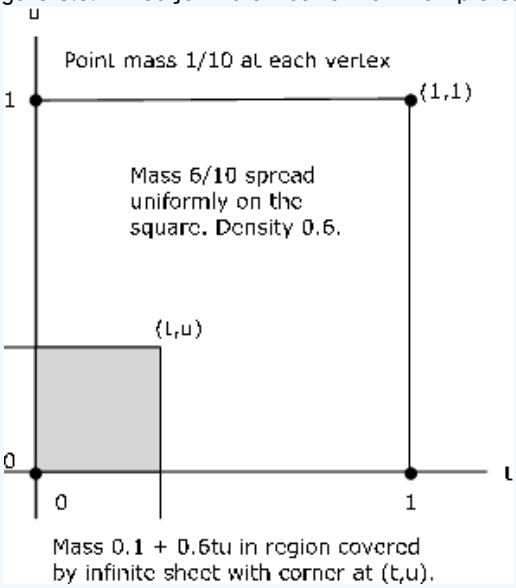
$$F_{XY}(t, u) = \frac{1}{10}(2 + 6t)$$

- If the point (t, u) is to the right of the square (including its boundary) with $0 \le u < 1$, the sheet covers two point masses and the portion of the mass in the square below the horizontal line through (t, u), to give

F_{XY} (t, u) = \dfrac{1}{10} (2 + 6u)\)

- If (t, u) is above and to the right of the square (i.e., both $1 \le t$ and $1 \le u$). then all probability mass is covered and $F_{XY}(t, u) = 1$ in this region.

Figure 8.3. Mixed joint distribution for Example 8.4.



Marginal Distributions

If the joint distribution for a random vector is known, then the distribution for each of the component random variables may be determined. These are known as **marginal distributions**. In general, the converse is not true. However, if the component random variables form an independent pair, the treatment in that case shows that the marginals determine the joint distribution.

To begin the investigation, note that

$$F_X(t) = P(X \le t) = P(X \le t, Y < \infty)$$ i.e. Y can take any of its possible values.

Thus

$$F_X(t) = F_{XY}(t, \infty) = \lim_{u \to \infty} F_{XY}(t, u)$$

This may be interpreted with the aid of Figure 8.1.4. Consider the sheet for point (t, u).

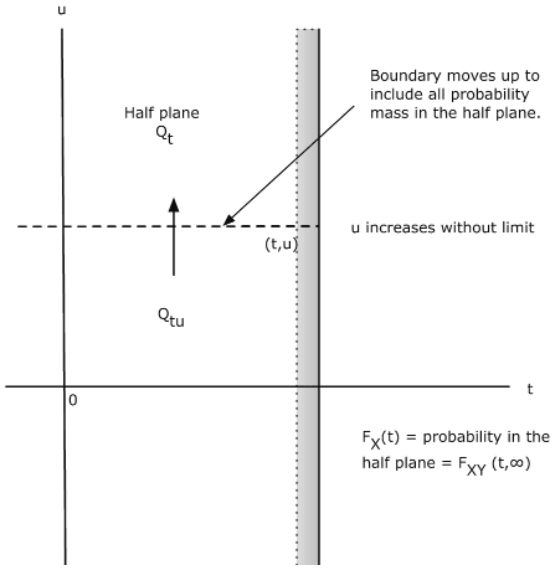


Figure 8.1.4. Construction for obtaining the marginal distribution for X.

If we push the point up vertically, the upper boundary of Q_{tu} is pushed up until eventually all probability mass on or to the left of the vertical line through (t, u) is included. This is the total probability that $X \le t$. Now $F_X(t)$ describes probability mass on the line. The probability mass described by $F_X(t)$ is the same as the total joint probability mass on or to the left of the vertical line through (t, u). We may think of the mass in the half plane being projected onto the horizontal line to give the *marginal* distribution for X. A parallel argument holds for the marginal for Y.

$$F_Y(u) = P(Y \le u) = F_{XY}(\infty, u) =$$ mass on or below horizontal line through (t, u)

This mass is projected onto the vertical axis to give the marginal distribution for Y.

Marginals for a joint discrete distribution

Consider a joint simple distribution.

$$P(X = t_i) = \sum_{j=1}^{m} P(X = t_i, Y = u_j)$$ and $$P(Y = u_j) = \sum_{i=1}^{n} P(X = t_i, Y = u_j)$$

Thus, all the probability mass on the vertical line through $(t_i, 0)$ is projected onto the point t_i on a horizontal line to give $P(X = t_i)$. Similarly, all the probability mass on a horizontal line through $(0, u_j)$ is projected onto the point u_j on a vertical line to give $P(Y = u_j)$.

Example 8.1.5: Marginals for a discrete distribution

The pair $\{X, Y\}$ produces a joint distribution that places mass 2/10 at each of the five points

(0, 0), (1, 1), (2, 0), (2, 2), (3, 1) (See Figure 8.1.5)

The marginal distribution for X has masses 2/10, 2/10, 4/10, 2/10 at points $t = 0$, 1, 2, 3, respectively. Similarly, the marginal distribution for *Y* has masses 4/10, 4/10, 2/10 at points $u = 0$, 1, 2, respectively.

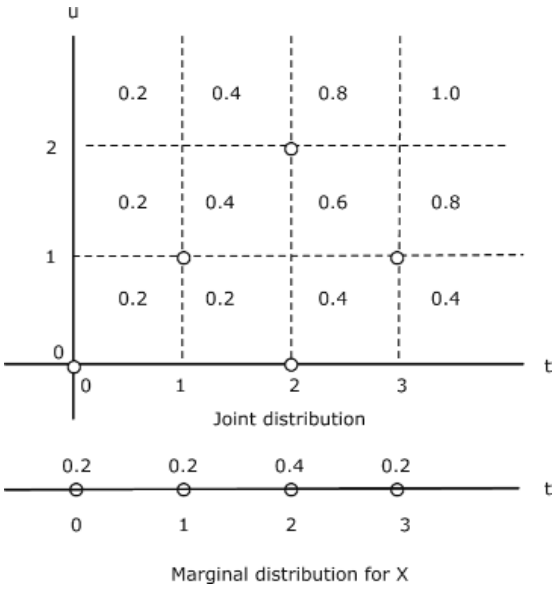


Figure 8.1.5. Marginal distribution for Example 8.1.1.

Consider again the joint distribution in Example 8.4. The pair $\{X, Y\}$ produces a mixed distribution as follows:

Point masses 1/10 at points (0,0), (1,0), (1,1), (0,1)

Mass 6/10 spread uniformly over the unit square with these vertices

The construction in Figure 8.1.6 shows the graph of the marginal distribution function F_X. There is a jump in the amount of 0.2 at $t = 0$, corresponding to the two point masses on the vertical line. Then the mass increases linearly with t, slope 0.6, until a final jump at $t = 1$ in the amount of 0.2 produced by the two point masses on the vertical line. At $t = 1$, the total mass is "covered" and $F_X(t)$ is constant at one for $t \geq 1$. By symmetry, the marginal distribution for Y is the same.

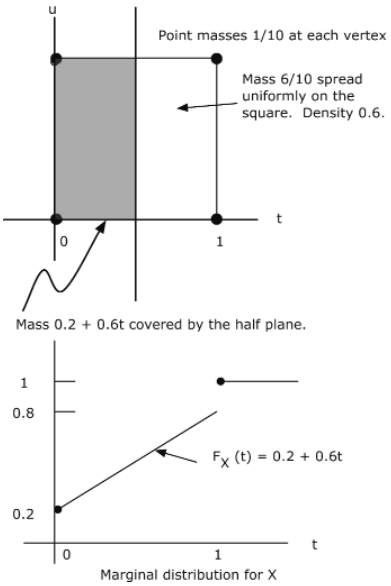


Figure 8.1.6. Marginal distribution for Example 8.1.6

This page titled 8.1: Random Vectors and Joint Distributions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

8.2: Random Vectors and MATLAB

m-procedures for a pair of simple random variables

We examine, first, calculations on a pair of simple random variables X, Y considered jointly. These are, in effect, two components of a random vector $W = (X, Y)$, which maps from the basic space Ω to the plane. The induced distribution is on the (t, u)-plane. Values on the horizontal axis (t-axis) correspond to values of the first coordinate random variable X and values on the vertical axis (u-axis) correspond to values of Y. We extend the computational strategy used for a single random variable.

First, let us review the one-variable strategy. In this case, data consist of values t_i and corresponding probabilities arranged in matrices

$$X = [t_1, t_2, \cdots, t_n] \text{ and } PX = [P(X = t_1), P(X = t_2), \cdots, P(X = t_n)]$$

To perform calculations on $Z = g(X)$, we we use array operations on X to form a matrix

$$G = [g(t_1)g(t_2) \cdots g(t_n)]$$

which has $g(t_i)$ in a position corresponding to $P(X = t_i)$ in matrix PX.

Basic problem. Determine $P(g(X) \in M)$, where M is some prescribed set of values.

- Use relational operations to determine the *positions* for which $g(t_i) \in M$. These will be in a zero-one matrix N, with ones in the desired positions.
- Select the $P(X = t_i)$ in the corresponding positions and sum. This is accomplished by one of the MATLAB operations to determine the inner product of N and PX

We extend these techniques and strategies to a pair of simple random variables, considered jointly.

The data for a pair $\{X, Y\}$ of random variables are the values of X and Y, which we may put in row matrices

$$X = [t_1 t_2 \cdots t_n] \text{ and } Y = [u_1 u_2 \cdots u_m]$$

and the joint probabilities $P(X = t_i, Y = u_j)$ in a matrix P. We usually represent the distribution graphically by putting probability mass $P(X = t_i, Y = u_j)$ at the point (t_i, u_j) on the plane. This joint probability may is represented by the matrix P with elements arranged corresponding to the mass points on the plane. Thus

$$P \text{ has elememt } P(X = t_i, Y = u_j) \text{ at the } (t_i, u_j) \text{ position}$$

To perform calculations, we form computational matrices t and u such that — t has element t_i at each (t_i, u_j) position (i.e., at each point on the ith column from the left) — u has element u_j at each (t_i, u_j) position (i.e., at each point on the jth row from the bottom) MATLAB array and logical operations on t, u, P perform the specified operations on t_i, u_j, and $P(X = t_i, Y = u_j)$ at each (t_i, u_j) position, in a manner analogous to the operations in the single-variable case.

Formation of the *t* and *u* matrices is achieved by a basic setup m-procedure called *jcalc*. The data for this procedure are in three matrices: $X = [t_1, t_2, \cdots, t_n]$ is the set of values for random variable X $Y = [u_1, u_2, \cdots, u_m]$ is the set of values for random variable Y, and $P = [p_{ij}]$, where $p_{ij} = P(X = t_i, Y = u_j)$. We arrange the joint probabilities as on the plane, with X-values increasing to the right and *Y*-values increasing upward. This is different from the usual arrangement in a matrix, in which values of the second variable increase downward. The m-procedure takes care of this inversion. The m-procedure forms the matrices t and u, utilizing the MATLAB function *meshgrid*, and computes the marginal distributions for X and Y. In the following example, we display the various steps utilized in the setup procedure. Ordinarily, these intermediate steps would not be displayed.

Example 8.2.7: Setup and basic calculations

```
>> jdemo4                          % Call for data in file jdemo4.m
>> jcalc                           % Call for setup procedure
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
>> disp(P)                         % Optional call for display of P
    0.0360    0.0198    0.0297    0.0209    0.0180
    0.0372    0.0558    0.0837    0.0589    0.0744
    0.0516    0.0774    0.1161    0.0817    0.1032
    0.0264    0.0270    0.0405    0.0285    0.0132
>> PX                              % Optional call for display of PX
PX =  0.1512    0.1800    0.2700    0.1900    0.2088
>> PY                              % Optional call for display of PY
PY =  0.1356    0.4300    0.3100    0.1244
- - - - - - - - - - -              % Steps performed by jcalc
>> PX = sum(P)                     % Calculation of PX as performed by jcalc
PX =  0.1512    0.1800    0.2700    0.1900    0.2088
>> PY = fliplr(sum(P'))            % Calculation of PY (note reversal)
PY = 0.1356    0.4300    0.3100    0.1244
>> [t,u] = meshgrid(X,fliplr(Y)); % Formation of t, u matrices (note reversal)
>> disp(t)                         % Display of calculating matrix t
    -3     0     1     3     5    % A row of X-values for each value of Y
    -3     0     1     3     5
    -3     0     1     3     5
    -3     0     1     3     5
>> disp(u)                         % Display of calculating matrix u
     2     2     2     2     2    % A column of Y-values (increasing
     1     1     1     1     1    % upward) for each value of X
     0     0     0     0     0
    -2    -2    -2    -2    -2
```

Suppose we wish to determine the probability $P(X^2 - 3Y \geq 1)$. Using array operations on t and u, we obtain the matrix $G = [g(t_i, u_j)]$.

```
>> G = t.^2 - 3*u                  % Formation of G = [g(t_i,u_j)] matrix
G  = 3    -6    -5     3    19
     6    -3    -2     6    22
     9     0     1     9    25
    15     6     7    15    31
>> M = G >=  1                     % Positions where G >= 1
M =  1     0     0     1     1
     1     0     0     1     1
     1     0     1     1     1
     1     1     1     1     1
>> pM = M.*P                     % Selection of probabilities
pM =
    0.0360         0         0    0.0209    0.0180
    0.0372         0         0    0.0589    0.0744
    0.0516         0    0.1161    0.0817    0.1032
    0.0264    0.0270    0.0405    0.0285    0.0132
>> PM = total(pM)                % Total of selected probabilities
PM =  0.7336                     % P(g(X,Y) >= 1)
```

In Example 8.1.3 from "Random Vectors and Joint Distributions" we note that the joint distribution function F_{XY} is constant over any grid cell, including the left-hand and lower boundaries, at the value taken on at the lower left-hand corner of the cell. These lower left-hand corner values may be obtained systematically from the joint probability matrix *P* by a two step operation.

- Take cumulative sums upward of the columns of P.
- Take cumulative sums of the rows of the resultant matrix.

This can be done with the MATLAB function cumsum, which takes column cumulative sums downward. By flipping the matrix and transposing, we can achieve the desired results.

Example 8.2.8: *F*Calculation of $_{XY}$ values for Example 8.3 from "Random Vectors and Joint Distributions"

```
>> P = 0.1*[3 0 0; 0 6 0; 0 0 1];
>> FXY = flipud(cumsum(flipud(P)))  % Cumulative column sums upward
FXY =
    0.3000    0.6000    0.1000
         0    0.6000    0.1000
         0         0    0.1000
>> FXY = cumsum(FXY')'              % Cumulative row sums
FXY =
    0.3000    0.9000    1.0000
         0    0.6000    0.7000
         0         0    0.1000
```

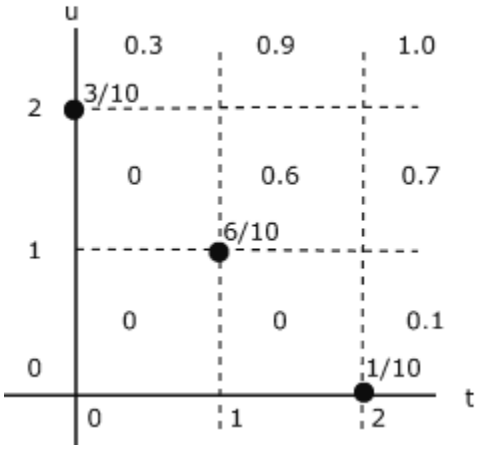


Figure 8.2.7. The joint distribution for Example 8.1.3 in "Random Vectors and Joint Distributions'.

Comparison with Example 8.3 from "Random Vectors and Joint Distributions" shows agreement with values obtained by hand. The two step procedure has been incorprated into an m-procedure *jddbn*. As an example, return to the distribution in Example Example 8.7

Example 8.2.9: Joint distribution function for example 8.7

```
>> jddbn
Enter joint probability matrix (as on the plane)  P
To view joint distribution function, call for FXY
>> disp(FXY)
    0.1512    0.3312    0.6012    0.7912    1.0000
    0.1152    0.2754    0.5157    0.6848    0.8756
    0.0780    0.1824    0.3390    0.4492    0.5656
    0.0264    0.0534    0.0939    0.1224    0.1356
```

These values may be put on a grid, in the same manner as in Figure 8.1.2 for Example 8.1.3 in "Random Vectors and Joint Distributions".

As in the case of canonic for a single random variable, it is often useful to have a function version of the procedure jcalc to provide the freedom to name the outputs conveniently. *function* `[x,y,t,u,px,py,p] = jcalcf(X,Y,P)` The quantities x, y, t, u, px, py , and p may be given any desired names.

Joint absolutely continuous random variables

In the single-variable case, the condition that there are no point mass concentrations on the line ensures the existence of a probability density function, useful in probability calculations. A similar situation exists for a joint distribution for two (or more) variables. For any joint mapping to the plane which assigns zero probability to each set with zero area (discrete points, line or curve segments, and countable unions of these) there is a density function.

Definition

If the joint probability distribution for the pair $\{X, Y\}$ assigns zero probability to every set of points with zero area, then there exists a *joint density function* f_{XY} with the property

$$P[(X, Y) \in Q] = \int \int_Q f_{XY}$$

We have three properties analogous to those for the single-variable case:

$$\text{(f1) } f_{XY} \geq 0 \text{ (f2) } \int \int_{R^2} f_{XY} = 1 \text{ (f3) } F_{XY}(t, u) = \int_{-\infty}^{1} \int_{-\infty}^{u} f_{XY}$$

At every continuity point for f_{XY}, the density is the second partial

$$f_{XY}(t, u) = \frac{\partial^2 F_{XY}(t, u)}{\partial t \partial u}$$

Now

$$F_X(t) = F_{XY}(t, \infty) = \int_{-\infty}^{t} \int_{-\infty}^{\infty} f_{XY}(r, s) ds dr$$

A similar expression holds for $F_Y(u)$. Use of the fundamental theorem of calculus to obtain the derivatives gives the result

$$f_X(t) = \int_{-\infty}^{\infty} f_{XY}(t, s) ds \text{ and } f_Y(u) = \int_{-\infty}^{\infty} f_{XY}(r, u) du$$

Marginal densities. Thus, to obtain the marginal density for the first variable, integrate out the second variable in the joint density, and similarly for the marginal for the second variable.

Example 8.2.10: Marginal density functions

Let $f_{XY}(t, u) = 8tu \; 0 \leq u \leq t \leq 1$. This region is the triangle bounded by $u = 0, u = t$, and $t = 1$(see Figure 8.2.8)

$$f_X(t) = \int f_{XY}(t, u) du = 8t \int_0^1 u du = 4t^3, \; 0 \leq t \leq 1$$

$$f_Y(u) = \int f_{XY}(t, u) dt = 8u \int_u^1 t dt = 4u(1 - u^2), \; 0 \leq u \leq 1$$

$P(0.5 \leq X \leq 0.75, Y > 0.5) = P[(X, Y) \in Q]$ where Q is the common part of the triangle with the strip between $t = 0.5$ and $t = 0.75$ and above the line $u = 0.5$. This is the small triangle bounded by $u = 0.5$, $u = t$, and $t = 0.75$. Thus

$$p = 8 \int_{1/2}^{3/4} \int_{1/2}^{t} t u du dt = 25/256 \approx 0.0977$$

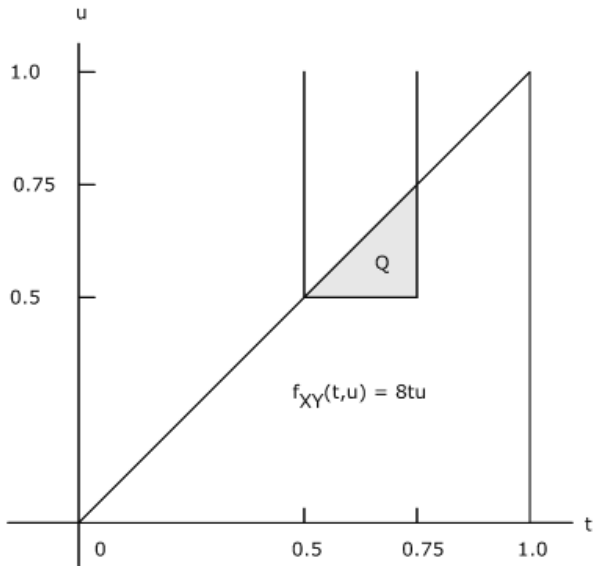


Figure 8.2.8. Distribution for Example 8.2.10

Example 8.2.11: Marginal distribution with compound expression

The pair $\{X, Y\}$ has joint density \f_{XY}(t, u) = \dfrac{6}{37} (t + 2u)\) on the region bounded by $t = 0, t = 2, u = 0$ and $u = \max\{1, t\}$ (see Figure 8.9). Determine the marginal density f_X.

Solution

Examination of the figure shows that we have different limits for the integral with respect to u for $0 \le t \le 1$ and for $1 < t \le 2$.

- For $0 \le t \le 1$

$$f_x(t) = \frac{6}{37} \int_0^1 (t + 2u)du = \frac{6}{37}(t + 1)$$

- For $1 < t \le 2$

$$f_X(t) = \frac{6}{37} \int_0^1 (t + 2u)du = \frac{12}{37}t^2$$

We may combine these into a single expression in a manner used extensively in subsequent treatments. Suppose $M = [0, 1]$ and $N = (1, 2]$. Then $I_M(t) = 1$ for $t \in M$ (i.e., $0 \le t \le 1$) and zero elsewhere. Likewise, $I_N(t) = 1$ for $t \in N$ and zero elsewhere. We can, therefore express f_X by

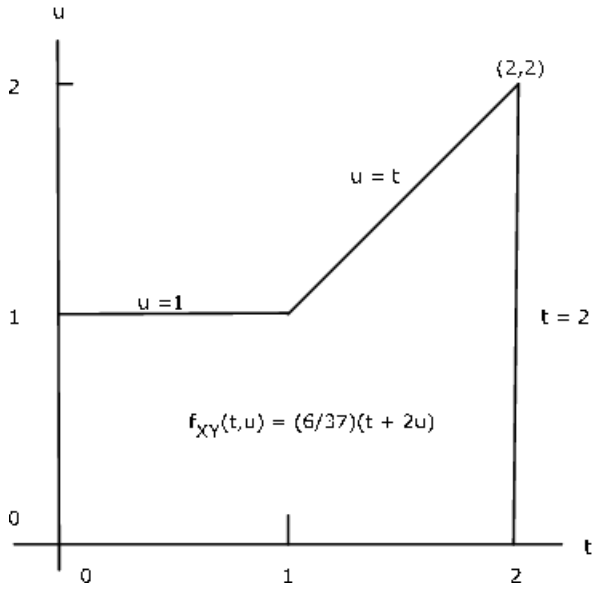


Figure 8.2.9. Marginal distribution for Example 8.2.11

Discrete approximation in the continuous case

For a pair $\{X, Y\}$ with joint density f_{XY}, we approximate the distribution in a manner similar to that for a single random variable. We then utilize the techniques developed for a pair of simple random variables. If we have n approximating values t_i for X and m approximating values u_j for Y, we then have $n \cdot m$ pairs (t_i, u_j), corresponding to points on the plane. If we subdivide the horizontal axis for values of X, with constant increments dx, as in the single-variable case, and the vertical axis for values of Y, with constant increments dy, we have a grid structure consisting of rectangles of size $dx \cdot dy$. We select t_i and u_j at the midpoint of its increment, so that the point (t_i, u_j) is at the midpoint of the rectangle. If we let the approximating pair be $\{X^*, Y^*\}$, we assign

$$p_{ij} = P((X^*, Y^*) = (t_i, u_j)) = P(X^* = t_i, Y^* = u_j) = P((X, Y) \text{ in } ij\text{th rectangle})$$

As in the one-variable case, if the increments are small enough,

$$P((X, Y) \in ij\text{th rectangle}) \approx dx \cdot dy \cdot f_{XY}(t_i, u_j)$$

The m-procedure *tuappr* calls for endpoints of intervals which include the ranges of X and Y and for the numbers of subintervals on each. It then prompts for an expression for $f_{XY}(t, u)$, from which it determines the joint probability distribution. It calculates the marginal approximate distributions and sets up the calculating matrices t and u as does the m-process jcalc for simple random variables. Calculations are then carried out as for any joint simple pair.

Example 8.2.12: Approximation to a joint continuous distribution

$$f_{XY}(t, u) = 3 \text{ on } 0 \leq u \leq t^2 \leq 1$$

Determine $P(X \leq 0.8, Y > 0.1)$.

```
>> tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  3*(u <= t.^2)
Use array operations on X, Y, PX, PY, t, u, and P
>> M = (t <= 0.8)&(u > 0.1);
```

```
>> p = total(M.*P)          % Evaluation of the integral with
p =   0.3355                % Maple gives 0.3352455531
```

The discrete approximation may be used to obtain approximate plots of marginal distribution and density functions.

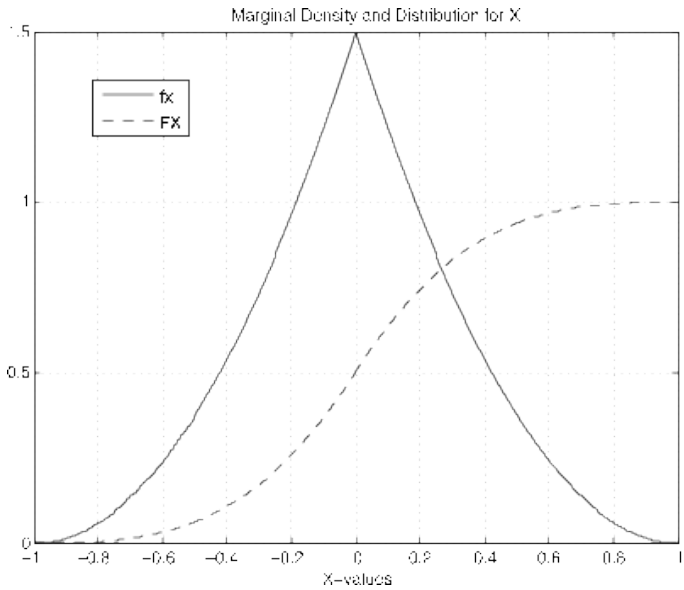


Figure 8.2.10. Marginal density and distribution function for Example 8.2.13

Example 8.2.13: Approximate plots of marginal density and distribution functions

$f_{XY}(t,u) = 3u$ on the triangle bounded by $u = 0$, $u \leq 1 + t$, and $u \leq 1 - t$.

```
>> tuappr
Enter matrix [a b] of X-range endpoints  [-1 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density  3*u.*(u<=min(1+t,1-t))
Use array operations on X, Y, PX, PY, t, u, and P
>> fx = PX/dx;                % Density for X  (see Figure 8.2.10)
                              % Theoretical (3/2)(1 - |t|)^2
>> fy = PY/dy;                % Density for Y
>> FX = cumsum(PX);           % Distribution function for X (Figure 8.2.10)
>> FY = cumsum(PY);           % Distribution function for Y
>> plot(X,fx,X,FX)            % Plotting details omitted
```

These approximation techniques useful in dealing with functions of random variables, expectations, and conditional expectation and regression.

This page titled 8.2: Random Vectors and MATLAB is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

8.3: Problems on Random Vectors and Joint Distributions

Exercise 8.3.1

Two cards are selected at random, without replacement, from a standard deck. Let X be the number of aces and Y be the number of spades. Under the usual assumptions, determine the joint distribution and the marginals.

Answer available at the end of this section.

Exercise 8.3.2

Two positions for campus jobs are open. Two sophomores, three juniors, and three seniors apply. It is decided to select two at random (each possible pair equally likely). Let X be the number of sophomores and Y be the number of juniors who are selected. Determine the joint distribution for the pair $\{X, Y\}$ and from this determine the marginals for each.

Answer available at the end of this section.

Exercise 8.3.3

A die is rolled. Let X be the number that turns up. A coin is flipped X times. Let Y be the number of heads that turn up. Determine the joint distribution for the pair $\{X, Y\}$. Assume $P(X = k) = 1/6$ for $1 \leq k \leq 6$ and for each k, $P(Y = j|X = k)$ has the binomial (k, 1/2) distribution. Arrange the joint matrix as on the plane, with values of Y increasing upward. Determine the marginal distribution for Y. (For a MATLAB based way to determine the joint distribution see Example 14.1.7 from "Conditional Expectation, Regression")

Answer available at the end of this section.

Exercise 8.3.4

As a variation of Exercise 8.3.3., Suppose a pair of dice is rolled instead of a single die. Determine the joint distribution for the pair $\{X, Y\}$ and from this determine the marginal distribution for Y.

Answer available at the end of this section.

Exercise 8.3.5

Suppose a pair of dice is rolled. Let X be the total number of spots which turn up. Roll the pair an additional X times. Let Y be the number of sevens that are thrown on the X rolls. Determine the joint distribution for the pair $\{X, Y\}$ and from this determine the marginal distribution for Y. What is the probability of three or more sevens?

Answer available at the end of this section.

Exercise 8.3.6

The pair $\{X, Y\}$ has the joint distribution (in m-file npr08_06.m):

$$X = [-2.3\ -0.7\ 1.1\ 3.9\ 5.1]\ Y = = [1.3\ 2.5\ 4.1\ 5.3]$$

$$P = \begin{bmatrix} 0.0483 & 0.0357 & 0.0420 & 0.0399 & 0.0441 \\ 0.0437 & 0.0323 & 0.0380 & 0.0361 & 0.0399 \\ 0.0713 & 0.0527 & 0.0620 & 0.0609 & 0.0551 \\ 0.0667 & 0.0493 & 0.0580 & 0.0651 & 0.0589 \end{bmatrix}$$

Determine the marginal distribution and the corner values for F_{XY}. Determine $P(X + Y > 2)$ and $P(X \geq Y)$.

Answer available at the end of this section.

Exercise 8.3.7

The pair $\{X, Y\}$ has the joint distribution (in m-file npr08_07.m):

$$P(X = i, Y = u)$$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189
-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

Determine the marginal distributions and the corner values for F_{XY}. Determine $P(1 \leq X \leq 4, Y > 4)$ and $P(|X - Y| \leq 2)$.

Answer available at the end of this section.

Exercise 8.3.8

The pair $\{X, Y\}$ has the joint distribution (in m-file npr08_08.m):

$$P(X = t, Y = u)$$

t =	1	3	5	7	9	11	13	15	17	19
u = 12	0.0156	0.0191	0.0081	0.0035	0.0091	0.0070	0.0098	0.0056	0.0091	0.0049
10	0.0064	0.0204	0.0108	0.0040	0.0054	0.0080	0.0112	0.0064	0.0104	0.0056
9	0.0196	0.0256	0.0126	0.0060	0.0156	0.0120	0.0168	0.0096	0.0056	0.0084
5	0.0112	0.0182	0.0108	0.0070	0.0182	0.0140	0.0196	0.0012	0.0182	0.0038
3	0.0060	0.0260	0.0162	0.0050	0.0160	0.0200	0.0280	0.0060	0.0160	0.0040
-1	0.0096	0.0056	0.0072	0.0060	0.0256	0.0120	0.0268	0.0096	0.0256	0.0084
-3	0.0044	0.0134	0.0180	0.0140	0.0234	0.0180	0.0252	0.0244	0.0234	0.0126
-5	0.0072	0.0017	0.0063	0.0045	0.0167	0.0090	0.0026	0.0172	0.0217	0.0223

Determine the marginal distributions. Determine $F_{XY}(10, 6)$ and $P(X > Y)$.

Answer available at the end of this section.

Exercise 8.3.9

Data were kept on the effect of training time on the time to perform a job on a production line. X is the amount of training, in hours, and Y is the time to perform the task, in minutes. The data are as follows (in m-file npr08_09.m):

$$P(X = t, Y = u)$$

t =	1	1.5	2	2.5	3
u = 5	0.039	0.011	0.005	0.001	0.001
4	0.065	0.070	0.050	0.015	0.010
3	0.031	0.061	0.137	0.051	0.033
2	0.012	0.049	0.163	0.058	0.039
1	0.003	0.009	0.045	0.025	0.017

Determine the marginal distributions. Determine $F_{XY}(2, 3)$ and $P(Y/X \geq 1.25)$.

Answer available at the end of this section.

For the joint densities in Exercises 10-22 below

a. Sketch the region of definition and determine analytically the marginal density functions f_X and f_Y.
b. Use a discrete approximation to plot the marginal density f_X and the marginal distribution function F_X.
c. Calculate analytically the indicated probabilities.
d. Determine by discrete approximation the indicated probabilities.

Exercise 8.3.10

$f_{XY}(t, u) = 1$ for $0 \leq t \leq 1$, $0 \leq u \leq 2(1 - t)$.

$$P(X > 1/2, Y > 1), P(0 \leq X \leq 1/2, Y > 1/2), P(Y \leq X)$$

Answer available at the end of this section.

Exercise 8.3.11

$f_{XY}(t, u) = 1/2$ on the square with vertices at (1, 0), (2, 1), (1, 2), (0, 1).

$$P(X > 1, Y > 1), P(X \leq 1/2, 1 < Y), P(Y \leq X)$$

Answer available at the end of this section.

Exercise 8.3.12

$f_{XY}(t, u) = 4t(1 - u)$ for $0 \leq t \leq 1$, $0 \leq u \leq 1$.

$$P(1/2 < X < 3/4, Y > 1/2),\ P(X \leq 1/2, Y > 1/2),\ P(Y \leq X)$$

Answer available at the end of this section.

Exercise 8.3.13

$f_{XY}(t,u) = \frac{1}{8}(t+u)$ for $0 \le t \le 2$, $0 \le u \le 2$.

$$P(X > 1/2, Y > 1/2), P(0 \le X \le 1, Y > 1), P(Y \le X)$$

Answer available at the end of this section.

Exercise 8.3.14

$f_{XY}(t,u) = 4ue^{-2t}$ for $0 \le t, 0 \le u \le 1$

$$P(X \le 1, Y > 1), P(X > 0, 1/2 < Y < 3/4), P(X < Y)$$

Answer available at the end of this section.

Exercise 8.3.15

$f_{XY}(t,u) = \frac{3}{88}(2t + 3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1 + t$.

$$F_{XY}(1,1), P(X \le 1, Y > 1), P(|X - Y| < 1)$$

Answer available at the end of this section.

Exercise 8.3.16

$f_{XY}(t,u) = 12t^2u$ on the parallelogram with vertices (-1, 0), (0, 0), (1, 1), (0, 1).

$$P(X \le 1/2, Y > 0), P(X < 1/2, Y \le 1/2), P(Y \ge 1/2)$$

Answer available at the end of this section.

Exercise 8.3.17

$f_{XY}(t,u) = \frac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2 - t\}$

$$P(X \le 1, Y \le 1), P(X > 1), P(X < Y)$$

Answer available at the end of this section.

Exercise 8.3.18

$f_{XY}(t,u) = \frac{3}{23}(t + 2u)$ for $0 \le t \le 2$, $0 \le u \le \max\{2 - t, t\}$

$$P(X \ge 1, Y \ge 1), P(Y \le 1), P(Y \le X)$$

Answer available at the end of this section.

Exercise 8.3.19

$f_{XY}(t,u) = \dfrac{12}{179}(3t^2 + u)$, for $0 \le t \le 2$, $0 \le u \le \min\{1+t, 2\}$

$$P(X \ge 1, Y \ge 1), P(X \le 1, Y \le 1), P(Y < X)$$

Answer available at the end of this section.

Exercise 8.3.20

$f_{XY}(t,u) = \dfrac{12}{227}(3t + 2tu)$ for $0 \le t \le 2$, $0 \le u \le \min\{1+t, 2\}$

$$P(X \le 1/2, Y \le 3/2), P(X \le 1.5, Y > 1), P(Y < X)$$

Answer available at the end of this section.

Exercise 8.3.21

$f_{XY}(t,u) = \dfrac{2}{13}(t + 2u)$ for $0 \le t \le 2$, $0 \le u \le \min\{2t, 3-t\}$

$$P(X < 1), P(X \ge 1, Y \le 1), P(Y \le X/2)$$

Answer available at the end of this section.

Exercise 8.3.22

$f_{XY}(t,u) = I_{[0,1]}(t)\dfrac{3}{8}(t^2 + 2u) + I_{(1,2]}(t)\dfrac{9}{14}t^2u^2$ for $0 \le u \le 1$.

$$P(1/2 \le X \le 3/2, Y \le 1/2)$$

Answer available at the end of this section.

Answers

Exercise 8.3.1

Let X be the number of aces and Y be the number of spades. Define the events AS_i, A_i, S_i, and N_i, $i = 1, 2$ of drawing ace of spades, other ace, spade (other than the ace), and neither on the i selection. Let $P(i,k) = P(X = i, Y = k)$.

$P(0,0) = P(N_1N_2) = \dfrac{36}{52}\cdot\dfrac{35}{51} = \dfrac{1260}{2652}$

$P(0,1) = P(N_1S_2 \bigvee S_1N_2) = \dfrac{36}{52}\cdot\dfrac{12}{51} + \dfrac{12}{52}\cdot\dfrac{36}{51} = \dfrac{864}{2652}$

$P(0,2) = P(S_1S_2) = \dfrac{12}{52}\cdot\dfrac{11}{51} = \dfrac{132}{2652}$

Double subscripts: use braces to clarify

$P(1,1) = P(A_1S_2 \bigvee S_1A_2 \bigvee AS_1N_2 \bigvee N_1AS_2) = \dfrac{3}{52}\cdot\dfrac{12}{51} + \dfrac{12}{52}\cdot\dfrac{3}{51} + \dfrac{1}{52}\cdot\dfrac{36}{51} + \dfrac{36}{52}\cdot\dfrac{1}{51} = \dfrac{144}{2652}$

$P(1,2) = P(AS_1S_2 \bigvee S_1AS_2) = \dfrac{1}{52}\cdot\dfrac{12}{51} + \dfrac{12}{52}\cdot\dfrac{1}{51} = \dfrac{24}{2652}$

$P(2,0) = P(A_1A_2) = \dfrac{3}{52}\cdot\dfrac{2}{51} = \dfrac{6}{2652}$

$P(2,1) = P(AS_1A_2 \bigvee A_1AS_2) = \dfrac{1}{52}\cdot\dfrac{3}{51} + \dfrac{3}{52}\cdot\dfrac{1}{51} = \dfrac{6}{2652}$

$P(2,2) = P(\emptyset) = 0$

```
01 % type npr08_01
```

```
% file npr08_01.m
% Solution for Exercise 8.3.1.
X = 0:2;
Y = 0:2;
Pn = [132  24   0; 864 144  6; 1260 216 6];
P = Pn/(52*51);
disp('Data in Pn, P, X, Y')

npr08_01         % Call for mfile
Data in Pn, P, X, Y    % Result
PX = sum(P)
PX =  0.8507    0.1448    0.0045
PY = fliplr(sum(P'))
PY =  0.5588    0.3824    0.0588
```

Exercise 8.3.2

Let A_i, B_i, C_i be the events of selecting a sophomore, junior, or senior, respectively, on the ith trial. Let X be the number of sophomores and Y be the number of juniors selected.

Set $P(i,k) = P(X = i, Y = k)$

$$P(0,0) = P(C_1C_2) = \frac{3}{8}\cdot\frac{2}{7} = \frac{6}{56}$$

$$P(0,1) = P(B_1C_2) + P(C_1B_2) = \frac{3}{8}\cdot\frac{3}{7} + \frac{3}{8}\cdot\frac{3}{7} = \frac{18}{56}$$

$$P(0,2) = P(B_1B_2) = \frac{3}{8}\cdot\frac{2}{7} = \frac{6}{56}$$

$$P(1,0) = P(A_1C_2) + P(C_1A_2) = \frac{2}{8}\cdot\frac{3}{7} + \frac{3}{8}\cdot\frac{2}{7} = \frac{12}{56}$$

$$P(1,1) = P(A_1B_2) + P(B_1A_2) = \frac{2}{8}\cdot\frac{3}{7} + \frac{3}{8}\cdot\frac{2}{7} = \frac{12}{56}$$

$$P(2,0) = P(A_1A_2) = \frac{2}{8}\cdot\frac{1}{7} = \frac{2}{56}$$

$$P(1,2) = P(2,1) = P(2,2) = 0$$

$PX =$ [30/56 24/56 2/56] $PY =$ [20/56 30/56 6/56]

```
% file npr08_02.m
% Solution for Exercise 8.3.2.
X = 0:2;
Y = 0:2;
Pn = [6 0 0; 18 12 0; 6 12 2];
P = Pn/56;
disp('Data are in X, Y,Pn, P')
npr08_02
Data are in X, Y,Pn, P
PX = sum(P)
PX =  0.5357    0.4286    0.0357
PY = fliplr(sum(P'))
PY =  0.3571    0.5357    0.1071
```

Exercise 8.3.3

$P(X = i, Y = k) = P(X = i)P(Y = k|X = i) = (1/6)P(Y = k|X = i)$.

```
% file npr08_03.m
% Solution for Exercise 8.3.3.
X = 1:6;
Y = 0:6;
P0 = zeros(6,7);       % Initialize
for i = 1:6            % Calculate rows of Y probabilities
    P0(i,1:i+1) = (1/6)*ibinom(i,1/2,0:i);
end
P = rot90(P0);         % Rotate to orient as on the plane
PY = fliplr(sum(P'));  % Reverse to put in normal order
disp('Answers are in X, Y, P, PY')
npr08_03            % Call for solution m-file
```

```
Answers are in X, Y, P, PY
disp(P)
         0         0         0         0         0    0.0026
         0         0         0         0    0.0052    0.0156
         0         0         0    0.0104    0.0260    0.0391
         0         0    0.0208    0.0417    0.0521    0.0521
         0    0.0417    0.0625    0.0625    0.0521    0.0391
    0.0833    0.0833    0.0625    0.0417    0.0260    0.0156
    0.0833    0.0417    0.0208    0.0104    0.0052    0.0026
disp(PY)
     0.1641  0.3125  0.2578  0.1667  0.0755  0.0208  0.0026
```

Exercise 8.3.4

```
% file npr08_04.m
% Solution for Exercise 8.3.4.
X = 2:12;
Y = 0:12;
PX = (1/36)*[1 2 3 4 5 6 5 4 3 2 1];
P0 = zeros(11,13);
for i = 1:11
    P0(i,1:i+2) = PX(i)*ibinom(i+1,1/2,0:i+1);
end
P = rot90(P0);
PY = fliplr(sum(P'));
disp('Answers are in X, Y, PY, P')
npr08_04
Answers are in X, Y, PY, P
disp(P)
  Columns 1 through 7
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0         0
         0         0         0         0         0         0    0.0005
         0         0         0         0         0    0.0013    0.0043
         0         0         0         0    0.0022    0.0091    0.0152
         0         0         0    0.0035    0.0130    0.0273    0.0304
         0         0    0.0052    0.0174    0.0326    0.0456    0.0380
         0    0.0069    0.0208    0.0347    0.0434    0.0456    0.0304
    0.0069    0.0208    0.0312    0.0347    0.0326    0.0273    0.0152
    0.0139    0.0208    0.0208    0.0174    0.0130    0.0091    0.0043
    0.0069    0.0069    0.0052    0.0035    0.0022    0.0013    0.0005
  Columns 8 through 11
         0         0         0    0.0000
         0         0    0.0000    0.0001
         0    0.0001    0.0003    0.0004
    0.0002    0.0008    0.0015    0.0015
    0.0020    0.0037    0.0045    0.0034
    0.0078    0.0098    0.0090    0.0054
    0.0182    0.0171    0.0125    0.0063
    0.0273    0.0205    0.0125    0.0054
    0.0273    0.0171    0.0090    0.0034
    0.0182    0.0098    0.0045    0.0015
    0.0078    0.0037    0.0015    0.0004
    0.0020    0.0008    0.0003    0.0001
    0.0002    0.0001    0.0000    0.0000
disp(PY)
```

```
  Columns 1 through 7
    0.0269    0.1025    0.1823    0.2158    0.1954    0.1400    0.0806
  Columns 8 through 13
    0.0375    0.0140    0.0040    0.0008    0.0001    0.0000
```

Exercise 8.3.5

```
% file npr08_05.m
% Data and basic calculations for Exercise 8.3.5.
PX = (1/36)*[1 2 3 4 5 6 5 4 3 2 1];
X = 2:12;
Y = 0:12;
P0 = zeros(11,13);
for i = 1:11
  P0(i,1:i+2) = PX(i)*ibinom(i+1,1/6,0:i+1);
end
P = rot90(P0);
PY = fliplr(sum(P'));
disp('Answers are in X, Y, P, PY')
npr08_05
Answers are in X, Y, P, PY
disp(PY)
  Columns 1 through 7
    0.3072    0.3660    0.2152    0.0828    0.0230    0.0048    0.0008
  Columns 8 through 13
    0.0001    0.0000    0.0000    0.0000    0.0000    0.0000
```

Exercise 8.3.6

```
npr08_06
Data are in X, Y, P
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
disp([X;PX]')
   -2.3000    0.2300
   -0.7000    0.1700
    1.1000    0.2000
    3.9000    0.2020
    5.1000    0.1980

disp([Y;PY]')
    1.3000    0.2980
    2.5000    0.3020
    4.1000    0.1900
    5.3000    0.2100
jddbn
Enter joint probability matrix (as on the plane)  P
To view joint distribution function, call for FXY
disp(FXY)
    0.2300    0.4000    0.6000    0.8020    1.0000
    0.1817    0.3160    0.4740    0.6361    0.7900
    0.1380    0.2400    0.3600    0.4860    0.6000
    0.0667    0.1160    0.1740    0.2391    0.2980
P1 = total((t+u>2).*P)
P1 =  0.7163
P2 = total((t>=u).*P)
```

```
P2 =  0.2799
```

Exercise 8.3.7

```
npr08_07
Data are in X, Y, P
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
disp([X;PX]')
   -3.1000    0.1500
   -0.5000    0.2200
    1.2000    0.3300
    2.4000    0.1200
    3.7000    0.1100
    4.9000    0.0700
disp([Y;PY]')
   -3.8000    0.1929
   -2.0000    0.3426
    4.1000    0.2706
    7.5000    0.1939
jddbn
Enter joint probability matrix (as on the plane)  P
To view joint distribution function, call for FXY
disp(FXY)
    0.1500    0.3700    0.7000    0.8200    0.9300    1.0000
    0.1410    0.3214    0.5920    0.6904    0.7564    0.8061
    0.0915    0.2719    0.4336    0.4792    0.5089    0.5355
    0.0510    0.0994    0.1720    0.1852    0.1852    0.1929
M = (1<=t)&(t<=4)&(u>4);
P1 = total(M.*P)
P1 =  0.3230
P2 = total((abs(t-u)<=2).*P)
P2 =  0.3357
```

Exercise 8.3.8

```
npr08_08
Data are in X, Y, P
jcalc
- - - - - - - - -
 Use array operations on matrices X, Y, PX, PY, t, u, and P
disp([X;PX]')
    1.0000    0.0800
    3.0000    0.1300
    5.0000    0.0900
    7.0000    0.0500
    9.0000    0.1300
   11.0000    0.1000
   13.0000    0.1400
   15.0000    0.0800
   17.0000    0.1300
   19.0000    0.0700
disp([Y;PY]')
   -5.0000    0.1092
   -3.0000    0.1768
   -1.0000    0.1364
```

```
    3.0000    0.1432
    5.0000    0.1222
    9.0000    0.1318
   10.0000    0.0886
   12.0000    0.0918
F = total(((t<=10)&(u<=6)).*P)
F =   0.2982
P = total((t>u).*P)
P =   0.7390
```

Exercise 8.3.9

```
npr08_09
Data are in X, Y, P
jcalc
- - - - - - - - - - - -
 Use array operations on matrices X, Y, PX, PY, t, u, and P
disp([X;PX]')
    1.0000    0.1500
    1.5000    0.2000
    2.0000    0.4000
    2.5000    0.1500
    3.0000    0.1000
disp([Y;PY]')
    1.0000    0.0990
    2.0000    0.3210
    3.0000    0.3130
    4.0000    0.2100
    5.0000    0.0570
F = total(((t<=2)&(u<=3)).*P)
F =   0.5100
P = total((u./t>=1.25).*P)
P =   0.5570
```

Exercise 8.3.10

Region is triangle with vertices (0, 0), (1, 0), (0, 2).

$$f_X(t) = \int_0^{2(1-t)} du = 2(1-t),\ 0 \le t \le 1$$
$$f_Y(u) = \int_0^{1-u/2} dt = 1 - u/2,\ 0 \le u \le 2$$
$M1 = \{(t,u) : t > 1/2, u > 1\}$ lies outside the trianlge $P((X,Y) \in M1) = 0$
$M2 = \{(t,u) : 0 \le t \le 1/2, u > 1/2\}$ has area in the triangle = 1/2
$M3$ = the region in the triangle under $u = t$, which has area 1/3

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  400
Enter expression for joint density  (t<=1)&(u<=2*(1-t))
Use array operations on X, Y, PX, PY, t, u, and P
fx = PX/dx;
FX = cumsum(PX);
plot(X,fx,X,FX)          % Figure not reproduced
M1 = (t>0.5)&(u>1);
P1 = total(M1.*P)
P1 =  0                  % Theoretical = 0
M2 = (t<=0.5)&(u>0.5);
P2 = total(M2.*P)
P2 =  0.5000             % Theoretical = 1/2
P3 = total((u<=t).*P)
```

```
P3 =  0.3350            % Theoretical = 1/3
```

Exercise 8.3.11

The region is bounded by lines $u = 1 + t$, $u = 1 - t$, $u = 3 - t$, and $u = t - 1$

$f_X(t) = I_{[0,1]}(t)0.5\int_{1-t}^{1+t} du + I_{(1,2]}(t)0.5\int_{t-1}^{3-t} du = I_{(1,2]}(t)(2-t) = f_Y(t)$ by symmetry

$M1 = \{(t,u) : t > 1, u > 1\}$ has area in the trangle = 1/2, so $PM1 = 1/4$

$M2 = \{(t,u) : t \le 1/2, u > 1\}$ has area in the trangle = 1/8\), so $PM2 = 1/16$

$M3 = \{(t,u) : u \le t\}$ has area in the trangle = 1, so $PM3 = 1/2$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  0.5*(u<=min(1+t,3-t))& ...
  (u>=max(1-t,t-1))
Use array operations on X, Y, PX, PY, t, u, and P
fx = PX/dx;
FX = cumsum(PX);
plot(X,fx,X,FX)          % Plot not shown
M1 = (t>1)&(u>1);
PM1 = total(M1.*P)
PM1 =  0.2501            % Theoretical = 1/4
M2 = (t<=1/2)&(u>1);
PM2 = total(M2.*P)
PM2 =  0.0631            % Theoretical = 1/16 = 0.0625
M3 = u<=t;
PM3 = total(M3.*P)
PM3 =  0.5023            % Theoretical = 1/2
```

Exercise 8.3.12

Region is the unit square,

$f_X(t) = \int_0^1 4t(1-u)du = 2t,\ 0 \le t \le 1$

$f_Y(u) = \int_0^1 4t(1-u)dt = 2(1-u),\ 0 \le u \le 1$

$P1 = \int_{1/2}^{3/4}\int_{1/2}^{1} 4t(1-u)dudt = 5/64$ $P2 = \int_0^{1/2}\int_{1/2}^{1} 4t(1-u)dudt = 1/16$

$P3 = \int_0^1\int_0^t 4t(1-u)dudt = 5/6$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  4*t.*(1 - u)
Use array operations on X, Y, PX, PY, t, u, and P
fx = PX/dx;
FX = cumsum(PX);
plot(X,fx,X,FX)           % Plot not shown
M1 = (1/2<t)&(t<3/4)&(u>1/2);
P1 = total(M1.*P)
P1 =  0.0781              % Theoretical = 5/64 = 0.0781
M2 = (t<=1/2)&(u>1/2);
P2 = total(M2.*P)
P2 =  0.0625              % Theoretical = 1/16 = 0.0625
M3 = (u<=t);
P3 = total(M3.*P)
P3 =  0.8350              % Theoretical = 5/6 = 0.8333
```

Exercise 8.3.13

Region is the square $0 \le t \le 2$, $0 \le u \le 2$

$f_X(t) = \frac{1}{8}\int_0^2 (t+u) = \frac{1}{4}(t+1) = f_Y(t),\ 0 \le t \le 2$

$$P1 = \int_{1/2}^{2}\int_{1/2}^{2}(t+u)dudt = 45/64\ P2 = \int_{0}^{1}\int_{1}^{2}(t+u)dudt = 1/4$$
$$P3 = \int_{0}^{2}\int_{0}^{1}(t+u)dudt = 1/2$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (1/8)*(t+u)
Use array operations on X, Y, PX, PY, t, u, and P
fx = PX/dx;
FX = cumsum(PX);
plot(X,fx,X,FX)
M1 = (t>1/2)&(u>1/2);
P1 = total(M1.*P)
P1 =  0.7031              % Theoretical = 45/64 = 0.7031
M2 = (t<=1)&(u>1);
P2 = total(M2.*P)
P2 =  0.2500              % Theoretical = 1/4
M3 = u<=t;
P3 = total(M3.*P)
P3 =  0.5025              % Theoretical = 1/2
```

Exercise 8.3.14

Region is strip by $t = 0, u = 0, u = 1$

$$f_X(t) = 2e^{-2t},\ 0 \le t,\ f_Y(u) = 2u,\ 0 \le u \le 1,\ f_{XY} = f_X f_Y$$
$$P1 = 0,\ P2 = \int_{0.5}^{\infty} 2e^{-2t}dt \int_{1/2}^{3/4} 2udu = e^{-1}5/16$$
$$P3 = 4\int_{0}^{1}\int_{t}^{1} ue^{-2t}dudt = \frac{3}{2}e^{-2} + \frac{1}{2} = 0.7030$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 3]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density  4*u.*exp(-2*t)
Use array operations on X, Y, PX, PY, t, u, and P
M2 = (t > 0.5)&(u > 0.5)&(u<3/4);
p2 = total(M2.*P)
p2 =  0.1139            % Theoretical = (5/16)exp(-1) = 0.1150
p3 = total((t<u).*P)
p3 =  0.7047            % Theoretical = 0.7030
```

Exercise 8.3.15

Region bounded by $t = 0, t = 2, u = 0, u = 1 + t$

$$f_X(t) = \frac{3}{88}\int_{0}^{1+t}(2t+3u^2)du = \frac{3}{88}(1+t)(1+4t+t^2) = \frac{3}{88}(1+5t+5t^2+t^3),\ 0 \le t \le 2$$
$$f_Y(u) = I_{[0,1]}(u)\frac{3}{88}\int_{0}^{2}(2t+3u^2)dt + I_{(1,3]}(u)\frac{3}{88}\int_{u-1}^{2}(2t+3u^2)dt =$$
$$I_{[0,1]}(u)\frac{3}{88}(6u^2+4) + I_{(1,3]}(t)\frac{3}{88}(3+2u+8u^2-3u^3)$$
$$F_{XY}(1,1) = \int_{0}^{1}\int_{0}^{1} f_{XY}(t,u)dudt = 3/44$$
$$P1 = \int_{0}^{1}\int_{1}^{1+t} f_{XY}(t,u)dudt = 41/352\ P2 = \int_{0}^{1}\int_{1}^{1+t} f_{XY}(t,u)dudt = 329/352$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 3]
Enter number of X approximation points  200
Enter number of Y approximation points  300
Enter expression for joint density  (3/88)*(2*t+3*u.^2).*(u<=1+t)
Use array operations on X, Y, PX, PY, t, u, and P
fx = PX/dx;
```

```
FX = cumsum(PX);
plot(X,fx,X,FX)
MF = (t<=1)&(u<=1);
F = total(MF.*P)
F =   0.0681            % Theoretical = 3/44 = 0.0682
M1 = (t<=1)&(u>1);
P1 = total(M1.*P)
P1 =  0.1172            % Theoretical = 41/352 = 0.1165
M2 = abs(t-u)<1;
P2 = total(M2.*P)
P2 =  0.9297           % Theoretical = 329/352 = 0.9347
```

Exercise 8.3.16

Region bounded by $u = 0$, $u = t$, $u = 1$, $u = t + 1$

$$f_X(t) = I_{[-1,0]}(t)12\int_0^{t+1} t^2u\,du + I_{(0,1]}(t)12\int_t^1 t^2u\,du = I_{[-1,0]}(t)6t^2(t+1)^2 + I_{(0,1]}(t)6t^2(1-t^2)$$

$$f_Y(u) = 12\int_{u-1}^{t} t^2u\,du + 12u^3 - 12u^2 + 4u,\ 0 \le u \le 1$$

$$P1 = 1 - 12\int_{1/2}^1\int_t^1 t^2u\,du\,dt = 33/80,\ P2 = 12\int_0^{1/2}\int_{u-1}^{u} t^2u\,dt\,du = 3/16$$

$$P3 = 1 - P2 = 13/16$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [-1 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density  12*u.*t.^2.*((u<=t+1)&(u>=t))
Use array operations on X, Y, PX, PY, t, u, and P
p1 = total((t<=1/2).*P)
p1 =  0.4098                % Theoretical = 33/80 = 0.4125
M2 = (t<1/2)&(u<=1/2);
p2 = total(M2.*P)
p2 =  0.1856                % Theoretical = 3/16  = 0.1875
P3 = total((u>=1/2).*P)
P3 =  0.8144                % Theoretical = 13/16 = 0.8125
```

Exercise 8.3.17

Region is bounded by $t = 0, u = 0, u = 2, u = 2 - t$

$$f_X(t) = I_{[0,1]}(t)\frac{24}{11}\int_0^1 tu\,du + I_{(1,2]}(t)\frac{24}{11}\int_0^{2-t} tu\,du =$$

$$I_{[0,1]}(t)\frac{12}{11}t + I_{(1,2]}(t)\frac{12}{11}t(2-t)^2$$

$$f_Y(u) = \frac{24}{11}\int_0^{2-u} tu\,dt = \frac{12}{11}u(u-2)^2,\ 0 \le u \le 1$$

$$P1 = \frac{24}{11}\int_0^1\int_0^1 tu\,du\,dt = 6/11\ P2 = \frac{24}{11}\int_1^2\int_0^{2-t} tu\,du\,dt = 5/11$$

$$P3 = \frac{24}{11}\int_0^1\int_t^1 tu\,du\,dt = 3/11$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density  (24/11)*t.*u.*(u<=2-t)
Use array operations on X, Y, PX, PY, t, u, and P
M1 = (t<=1)&(u<=1);
P1 = total(M1.*P)
P1 = 0.5447             % Theoretical = 6/11 = 0.5455
P2 = total((t>1).*P)
P2 =  0.4553            % Theoretical = 5/11 = 0.4545
P3 = total((t<u).*P)
P3 =  0.2705            % Theoretical = 3/11 = 0.2727
```

Exercise 8.3.18

Region is bounded by $t = 0, t = 2, u = 0, u = 2 - t\ (0 \le t \le 1),\ u = t (1 < t \le 2)$

$$f_X(t) = I_{[0,1]}(t)\frac{3}{23}\int_0^{2-t}(t+2u)du + I_{(1,2]}(t)\frac{3}{23}\int_0^{t}(t+2u)du = I_{[0,1]}(t)\frac{6}{23}(2-t) + I_{(1,2]}(t)\frac{6}{23}t^2$$

$$f_Y(u) = I_{[0,1]}(u)\frac{3}{23}\int_0^{2}(t+2u)du + I_{(1,2]}(u)[\frac{3}{23}\int_0^{2-u}(t+2u)dt + \frac{3}{23}\int_u^{2}(t+2u)dt] =$$

$$I_{[0,1]}(u)\frac{6}{23}(2u+1) + I_{(1,2]}(u)\frac{3}{23}(4+6u-4u^2)$$

$$P1 = \frac{3}{23}\int_1^2\int_1^t(t+2u)dudt = 13/46,\ P2 = \frac{3}{23}\int_0^2\int_0^1(t+2u)dudt = 12/23$$

$$P3 = \frac{3}{23}\int_0^2\int_0^t(t+2u)dudt = 16/23$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (3/23)*(t+2*u).*(u<=max(2-t,t))
Use array operations on X, Y, PX, PY, t, u, and P
M1 = (t>=1)&(u>=1);
P1 = total(M1.*P)
P1 =  0.2841
13/46                  % Theoretical = 13/46 = 0.2826
P2 = total((u<=1).*P)
P2 =  0.5190             % Theoretical = 12/23 = 0.5217
P3 = total((u<=t).*P)
P3 =  0.6959             % Theoretical = 16/23 = 0.6957
```

Exercise 8.3.19

Region has two parts: (1) $0 \le t \le 1, 0 \le u \le 2$ (2) $1 < t \le 2, 0 \le u \le 3 - t$

$$f_X(t) = I_{[0,1]}(t)\frac{12}{179}\int_0^{2}(3t^2+u)du + I_{(1,2]}(t)\frac{12}{179}\int_0^{3-t}(3t^2+u)du =$$

$$I_{[0,1]}(t)\frac{24}{179}(3t^2+1) + I_{(1,2]}(t)\frac{6}{179}(9-6t+19t^2-6t^3)$$

$$f_Y(u) = I_{[0,1]}(u)\frac{12}{179}\int_0^{2}(3t^2+u)dt + I_{(1,2]}(u)\frac{12}{179}\int_0^{3-u}(3t^2+u)dt =$$

$$I_{[0,1]}(u)\frac{24}{179}(4+u) + I_{(1,2]}(u)\frac{12}{179}(27-24u+8u^2-u^3)$$

$$P1 = \frac{12}{179}\int 1^2\int_1^{3-t}(3t^2+u)dudt = 41/179\ P2 = \frac{12}{179}\int_0^1\int_0^1(3t^2+u)dudt = 18/179$$

$$P3 = \frac{12}{179}\int_0^{3/2}\int_0^t(3t^2+u)dudt + \frac{12}{179}\int_{3/2}^2\int_0^{3-t}(3t^2+u)dudt = 1001/1432$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (12/179)*(3*t.^2+u).* ...
     (u<=min(2,3-t))
Use array operations on X, Y, PX, PY, t, u, and P
fx = PX/dx;
FX = cumsum(PX);
plot(X,fx,X,FX)
M1 = (t>=1)&(u>=1);
P1 = total(M1.*P)
P1 =  2312             % Theoretical = 41/179 = 0.2291
M2 = (t<=1)&(u<=1);
P2 = total(M2.*P)
P2 =  0.1003            % Theoretical = 18/179 = 0.1006
M3 = u<=min(t,3-t);
P3 = total(M3.*P)
P3 =  0.7003             % Theoretical = 1001/1432 = 0.6990
```

Exercise 8.3.20

Region is in two parts:

1. $0 \le t \le 1, 0 \le u \le 1+t$
2. $1 < t \le 2, 0 \le u \le 2$

$$f_X(t) = I_{[0,1]}(t)\int_0^{1+t} f_{XY}(t,u)du + I_{(1,2]}(t)\int_0^2 f_{XY}(t,u)du =$$
$$I_{[0,1]}(t)\frac{12}{227}(t^3+5t^2+4t) + I_{(1,2]}(t)\frac{120}{227}t$$
$$f_Y(u) = I_{[0,1]}(u)\int_0^2 f_{XY}(t,u)dt + I_{(1,2]}(u)\int_{u-1}^2 f_{XY}(t,u)dt =$$
$$I_{[0,1]}(u)\frac{24}{227}(2u+3) + I_{(1,2]}(u)\frac{6}{227}(2u+3)(3+2u-u^2)$$
$$= I_{[0,1]}(u)\frac{24}{227}(2u+3) + I_{(1,2]}(u)\frac{6}{227}(9+12u+u^2-2u^3)$$
$$P1 = \frac{12}{227}\int_0^{1/2}\int_0^{1+t}(3t+2tu)dudt = 139/3632$$
$$P2 = \frac{12}{227}\int_0^1\int_1^{1+t}(3t+2tu)dudt + \frac{12}{227}\int_1^{3/2}\int_1^2(3t+2tu)dudt = 68/227$$
$$P3 = \frac{12}{227}\int_0^2\int_1^t(3t+2tu)dudt = 144/227$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (12/227)*(3*t+2*t.*u).* ...
(u<=min(1+t,2))
Use array operations on X, Y, PX, PY, t, u, and P
M1 = (t<=1/2)&(u<=3/2);
P1 = total(M1.*P)
P1 =  0.0384             % Theoretical = 139/3632 = 0.0383
M2 = (t<=3/2)&(u>1);
P2 = total(M2.*P)
P2 =  0.3001             % Theoretical = 68/227 = 0.2996
M3 = u<t;
P3 = total(M3.*P)
P3 =  0.6308             % Theoretical = 144/227 = 0.6344
```

Exercise 8.3.21

Region bounded by $t = 2, u = 2t\ (0 \le t \le 1), 3-t\ (1 \le t \le 2)$

$$f_X(t) = I_{[0,1]}(t)\frac{2}{13}\int_0^{2t}(t+2u)du + I_{(1,2]}(t)\frac{2}{13}\int_0^{3-t}(t+2u)du = I_{[0,1]}(t)\frac{12}{13}t^2 + I_{(1,2]}(t)\frac{6}{13}(3-t)$$
$$f_Y(u) = I_{[0,1]}(u)\frac{2}{13}\int_{u/2}^2(t+2u)dt + I_{(1,2]}(u)\frac{2}{13}\int_{u/2}^{3-u}(t+2u)dt =$$
$$I_{[0,1]}(u)(\frac{4}{13}+\frac{8}{13}u-\frac{9}{52}u^2) + I_{(1,2]}(u)(\frac{9}{13}+\frac{6}{13}u-\frac{21}{52}u^2)$$
$$P1 = \int_0^1\int_0^{2t}(t+2u)dudt = 4/13\ P2 = \int_1^2\int_0^1(t+2u)dudt = 5/13$$
$$P3 = \int_0^2\int_0^{u/2}(t+2u)dudt = 4/13$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  400
Enter number of Y approximation points  400
Enter expression for joint density  (2/13)*(t+2*u).*(u<=min(2*t,3-t))
Use array operations on X, Y, PX, PY, t, u, and P
P1 = total((t<1).*P)
P1 = 0.3076              % Theoretical = 4/13 = 0.3077
M2 = (t>=1)&(u<=1);
P2 = total(M2.*P)
P2 =  0.3844             % Theoretical = 5/13 = 0.3846
P3 = total((u<=t/2).*P)
P3 =  0.3076              % Theoretical = 4/13 = 0.3077
```

Exercise 8.3.22

Region is rectangle bounded by $t = 0, t = 2, u = 0, u = 1$

$$f_{XY}(t,u) = I_{[0,1]}(t)\frac{3}{8}(t^2+2u) + I_{(1,2]}(t)\frac{9}{14}t^2u^2, 0 \le u \le 1$$

$$f_X(t) = I_{[0,1]}(t)\frac{3}{8}\int_0^1 (t^2+2u)du + I_{(1,2]}(t)\frac{9}{14}\int_0^1 t^2u^2du = I_{[0,1]}(t)\frac{3}{8}(t^2+1) + I_{(1,2]}(t)\frac{3}{14}t^2$$

$$f_Y(u) = \frac{3}{8}\int_0^1 (t^2+2u0dt + \frac{9}{14}\int_1^2 t^2u^2dt = \frac{1}{8} + \frac{3}{4}u + \frac{3}{2}u^2 \; 0 \le u \le 1$$

$$P1 = \frac{3}{8}\int_{1/2}^1\int_0^{1/2}(t^2+2u)dudt + \frac{9}{14}\int_1^{3/2}\int_0^{1/2} t^2u^2dudt = 55/448$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density  (3/8)*(t.^2+2*u).*(t<=1) ...
       + (9/14)*(t.^2.*u.^2).*(t > 1)
Use array operations on X, Y, PX, PY, t, u, and P
M = (1/2<=t)&(t<=3/2)&(u<=1/2);
P = total(M.*P)
P =  0.1228          % Theoretical = 55/448 = 0.1228
```

This page titled 8.3: Problems on Random Vectors and Joint Distributions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

9: INDEPENDENT CLASSES OF RANDOM VARIABLES

9.1: Independent Classes of Random Variables
9.2: Problems on Independent Classes of Random Variables

This page titled 9: Independent Classes of Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Independent Classes of Random Variables

The concept of independence for classes of events is developed in terms of a product rule. In this unit, we extend the concept to classes of random variables.

Independent pairs

Recall that for a random variable X, the inverse image $X^{-1}(M)$ (i.e., the set of all outcomes $\omega \in \Omega$ which are mapped into M by X) is an event for each reasonable subset M on the real line. Similarly, the inverse image $Y^{-1}(N)$ is an event determined by random variable Y for each reasonable set N. We extend the notion of independence to a pair of random variables by requiring independence of the events they determine. More precisely,

Definition

A pair $\{X, Y\}$ of random variables is (stochastically) *independent* iff each pair of events $\{X^{-1}(M), Y^{-1}(N)\}$ is independent.

This condition may be stated in terms of the *product rule*

$$P(X \in M, Y \leq N) = P(X \in M)P(Y \in N) \text{ for all (Borel) sets } M, N$$

Independence implies

$$\begin{aligned} F_{XY}(t,u) &= P(X \in (-\infty, t], Y \in (-\infty, u]) \\ &= P(X \in (-\infty, t])P(Y \in (-\infty, u]) \\ &= F_X(t)F_Y(u) \quad \forall t, u \end{aligned}$$

Note that the product rule on the distribution function is equivalent to the condition the product rule holds for the inverse images of a special class of sets $\{M, N\}$ of the form $M = (-\infty, t]$ and $N = (-\infty, u]$. An important theorem from measure theory ensures that if the product rule holds for this special class it holds for the general class of $\{M, N\}$. Thus we may assert

The pair $\{X, Y\}$ is independent iff the following product rule holds

$$F_{XY}(t,u) = F_X(t)F_Y(u) \quad \forall t, u \tag{9.1.1}$$

Example 9.1.1: an independent pair

Suppose $F_{XY}(t,u) = (1 - e^{-\infty})(1 - e^{-\beta u})$ $0 \leq t$, $0 \leq u$. Taking limits shows

$$F_X(t) = \lim_{u \to \infty} F_{XY}(t,u) = 1 - e^{-\alpha t}$$

and

$$F_Y(u) = \lim_{t \to \infty} F_{XY}(t,u) = 1 - e^{-\beta u}$$

so that the product rule $F_{XY}(t,u) = F_X(t)F_Y(u)$ holds. The pair $\{X, Y\}$ is therefore independent.

If there is a joint density function, then the relationship to the joint distribution function makes it clear that the pair is independent iff the product rule holds for the density. That is, the pair is independent iff

$$f_{XY}(t,u) = f_X(t)f_Y(u) \ \forall t, u$$

example 9.1.2: joint uniform distributin on a rectangle

suppose the joint probability mass distributions induced by the pair $\{X, Y\}$ is uniform on a rectangle with sides $I_1 = [a, b]$ and $I_2 = [c, d]$. Since the area is $(b - a)(d - c)$, the constant value of f_{XY} is $1/(b - a)(d - c)$. Simple integration gives

$$f_X(t) = \frac{1}{(b-a)(d-c)} \int_c^d du = \frac{1}{b-a} \quad a \le t \le b$$

and

$$f_Y(u) = \frac{1}{(b-a)(d-c)} \int_a^b dt = \frac{1}{d-c} \quad c \le u \le d$$

Thus it follows that X is uniform on $[a, b]$. Y is uniform on $[c, d]$, and $f_{XY}(t, u) = f_X(t) f_Y(u)$ for all t, u, so that the pair $\{X, Y\}$ is independent. The converse is also true: if the pair is independent with X uniform on $[a, b]$ and Y is uniform on $[c, d]$, the pair has uniform joint distribution on $I_1 \times I_2$.

The Joint Mass Distribution

It should be apparent that the independence condition puts restrictions on the character of the joint mass distribution on the plane. In order to describe this more succinctly, we employ the following terminology.

Definition

If M is a subset of the horizontal axis and N is a subset of the vertical axis, then the *cartesian product* $M \times N$ is the (generalized) rectangle consisting of those points (t, u) on the plane such that $t \in M$ and $u \in N$.

example 9.1.3: Rectangle with interval sides

The rectangle in Example 9.1.2 is the artesian product $I_1 \times I_2$, consisting of all those points (t, u) such that $a \le t \le b$ and $c \le u \le d$ (i.e. $t \in I_1$ and $u \in I_2$).

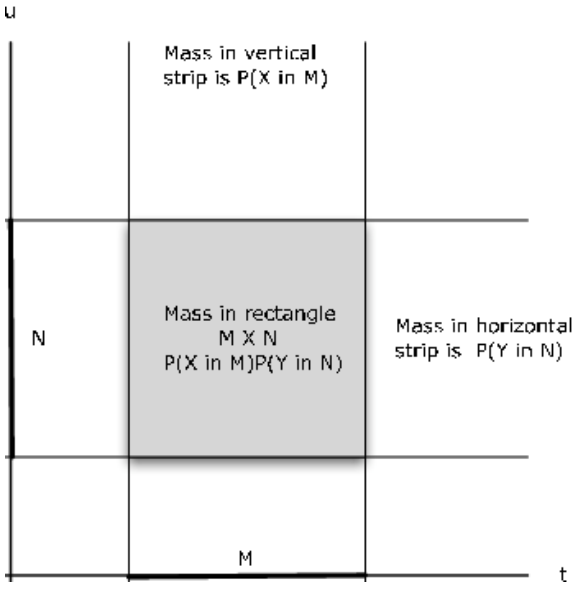


Figure 9.1.1. Joint distribution for an independent pair of random variables.

We restate the product rule for independence in terms of cartesian product sets.

$$P(X \in M, Y \in N) = P((X,Y) \in M \times N) = P(X \in M)P(Y \in N) \tag{9.1.2}$$

Reference to Figure 9.1.1 illustrates the basic pattern. If M, N are intervals on the horizontal and vertical axes, respectively, then the rectangle $M \times N$ is the intersection of the vertical strip meeting the horizontal axis in M with the horizontal strip meeting the vertical axis in N. The probability $X \in M$ is the portion of the joint probability mass in the vertical strip; the probability $Y \in N$ is the part of the joint probability in the horizontal strip. The probability in the rectangle is the product of these marginal probabilities.

This suggests a useful test for nonindependence which we call the *rectangle test*. We illustrate with a simple example.

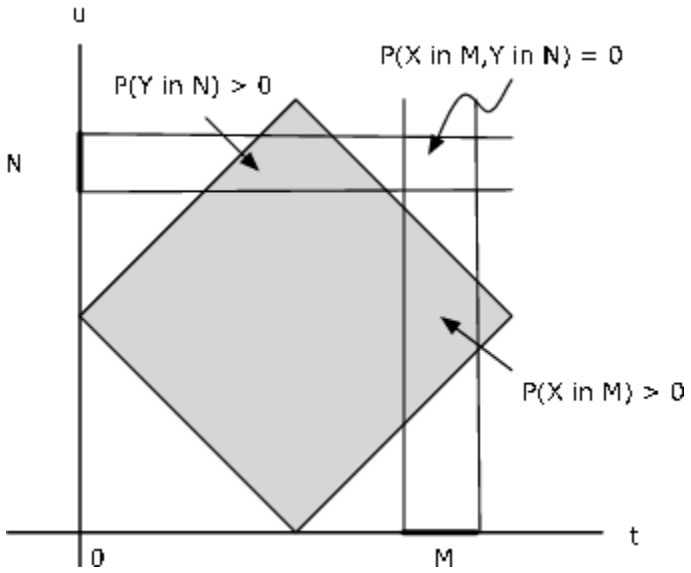


Figure 9.1.2. Rectangle test for nonindependence of a pair of random variables.

example 9.1.4: The rectangle test for nonindependence

Supose probability mass is uniformly distributed over the square with vertices at (1,0), (2,1), (1,2), (0,1). It is evident from Figure 9.1.2 that a value of X determines the possible values of Y and vice versa, so that we would not expect independence of the pair. To establish this, consider the small rectangle $M \times N$ shown on the figure. There is no probability mass in the region. Yet $P(X \in M) > 0$ and $P(Y \in N) > 0$, so that

$P(X \in M)P(Y \in N) > 0$, but $P((X,Y) \in M \times N) = 0$. The product rule fails; hence the pair cannot be stochastically independent.

Remark. There are nonindependent cases for which this test does not work. And it does not provide a test *for* independence. In spite of these limitations, it is frequently useful. Because of the information contained in the independence condition, in many cases the complete joint and marginal distributions may be obtained with appropriate partial information. The following is a simple example.

example 9.1.5: Joint and marginal probabilities from partial information

Suppose the pair $\{X, Y\}$ is independent and each has three possible values. The following four items of information are available.

$P(X = t_1) = 0.2$, $P(Y = u_1) = 0.3$, $P(X = t_1, Y = u_2) = 0.08$

$P(X = t_2, Y = u_1) = 0.15$

These values are shown in bold type on Figure 9.1.3. A combination of the product rule and the fact that the total probability mass is one are used to calculate each of the marginal and joint probabilities. For example $P(X = t_1) = 0.2$ and $P(X = t_1, Y = u_2) = P(X = t_1)P(Y = u_2) = 0.8$ implies $P(Y = u_2) = 0.4$. Then P(Y = u_3) = 1 - P(Y = u_1) - P(Y = u_2) = 0.3\). Others are calculated similarly. There is no unique procedure for solution. And it has not seemed useful to develop MATLAB procedures to accomplish this.

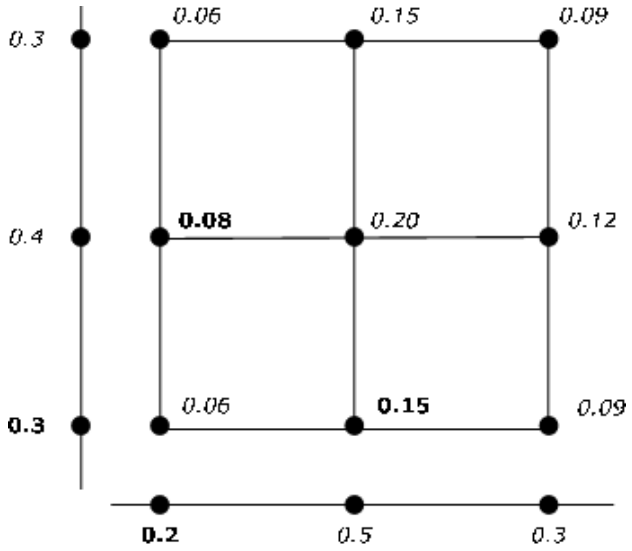


Figure 9.1.3. Joint and marginal probabilities from partial information.

example 9.1.6: The joint normal distribution

A pair $\{X, Y\}$ has the joint normal distribution iff the joint density is

$$f_{XY}(t,u) = \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho^2)^{1/2}} e^{-Q(t,u)/2}$$

where

$$Q(t,u) = \frac{1}{1-\rho^2}[(\frac{t-\mu_X}{\sigma_X})^2 - 2\rho(\frac{t-\mu_X}{\sigma_X})(\frac{t-\mu_Y}{\sigma_Y}) + (\frac{t-\mu_Y}{\sigma_Y})^2]$$

The marginal densities are obtained with the aid of some algebraic tricks to integrate the joint density. The result is that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. If the parameter ρ is set to zero, the result is

$$f_{XY}(t,u) = f_X(t)f_Y(u)$$

so that the pair is independent iff $\rho = 0$. The details are left as an exercise for the interested reader.

Remark. While it is true that every independent pair of normally distributed random variables is joint normal, not every pair of normally distributed random variables has the joint normal distribution.

Example 9.1.7: a normal pair not joint normally distributed

We start with the distribution for a joint normal pair and derive a joint distribution for a normal pair which is *not* joint normal. The function

$$\varphi(t,u) = \frac{1}{2\pi}\exp\,(-\frac{t^2}{2} - \frac{u^2}{2})$$

is the joint normal density for an independent pair ($\rho = 0$) of standardized normal random variables. Now define the joint density for a pair $\{X, Y\}$ by

$$f_{XY}(t, u) = 2\varphi(t, u) \text{ in the first and third quadrants, and zero elsewhere}$$

Both $X \sim N(0, 1)$ and $Y \sim N(0, 1)$. However, they cannot be joint normal, since the joint normal distribution is positive for all (t, u).

Independent classes

Since independence of random variables is independence of the events determined by the random variables, extension to general classes is simple and immediate.

Definition

A class $\{X_i : i \in J\}$ of random variables is (stochastically) *independent* iff the product rule holds for every finite subclass of two or more.

Remark. The index set J in the definition may be finite or infinite.

For a finite class $\{X_i : 1 \leq i \leq n\}$, independence is equivalent to the product rule

$$F_{X_1X_2\cdots X_n}(t_1, t_2, \cdots, t_n) = \prod_{i=1}^{n} F_{X_i}(t_i) \text{ for all } (t_1, t_2, \cdots, t_n)$$

Since we may obtain the joint distribution function for any finite subclass by letting the arguments for the others be ∞ (i.e., by taking the limits as the appropriate t_i increase without bound), the single product rule suffices to account for all finite subclasses.

Absolutely continuous random variables

If a class $\{X_i : i \in J\}$ is independent and the individual variables are absolutely continuous (i.e., have densities), then any finite subclass is jointly absolutely continuous and the product rule holds for the densities of such subclasses

$$f_{X_{i1}X_{i2}\cdots X_{im}}(t_{i1}, t_{i2}, \cdots, t_{im}) = \prod_{k=1}^{m} f_{X_{ik}}(t_{ik}) \text{ for all } (t_1, t_2, \cdots, t_n)$$

Similarly, if each finite subclass is jointly absolutely continuous, then each individual variable is absolutely continuous and the product rule holds for the densities. Frequently we deal with independent classes in which each random variable has the same marginal distribution. Such classes are referred to as *iid* classes (an acronym for **i**ndependent,**i**dentically **d**istributed). Examples are simple random samples from a given population, or the results of repetitive trials with the same distribution on the outcome of each component trial. A Bernoulli sequence is a simple example.

Simple random variables

Consider a pair $\{X, Y\}$ of simple random variables in canonical form

$$X = \sum_{i=1}^{n} t_i I_{A_i} \; Y = \sum_{j=1}^{m} u_j I_{B_j}$$

Since $A_i = \{X = t_i\}$ and $B_j = \{Y = u_j\}$ the pair $\{X, Y\}$ is independent iff each of the pairs $\{A_i, B_j\}$ is independent. The joint distribution has probability mass at each point (t_i, u_j) in the range of $W = (X, Y)$. Thus at every point on the grid,

$$P(X = t_i, Y = u_j) = P(X = t_i)P(Y = u_j)$$

According to the rectangle test, no gridpoint having one of the t_i or u_j as a coordinate has zero probability mass . The marginal distributions determine the joint distributions. If X has n distinct values and Y has m distinct values, then the *n+m* marginal probabilities suffice to determine the *m·n* joint probabilities. Since the marginal probabilities for each variable must add to one, only $n - 1) + (m - 1) = m + n - 2$ values are needed.

Suppose X and Y are in affine form. That is,

$$X = a_0 + \sum_{i=1}^{n} a_i I_{E_i}\ Y = b_0 + \sum_{j=1}^{m} b_j I_{E_j}$$

Since $A_r = \{X = t_r\}$ is the union of minterms generated by the E_i and $B_j = \{Y = u_s\}$ is the union of minterms generated by the F_j, the pair $\{X, Y\}$ is independent iff each pair of minterms $\{M_a, N_b\}$ generated by the two classes, respectivly, is independent. Independence of the minterm pairs is implied by independence of the combined class

$$\{E_i, F_j : 1 \le i \le n, 1 \le j \le m\}$$

Calculations in the joint simple case are readily handled by appropriate m-functions and m-procedures.

MATLAB and independent simple random variables

In the general case of pairs of joint simple random variables we have the m-procedure jcalc, which uses information in matrices X, Y and P to determine the marginal probabilities and the calculation matrices t and u. In the independent case, we need only the marginal distributions in matrices X, PX, Y and PY to determine the joint probability matrix (hence the joint distribution) and the calculation matrices t and u. If the random variables are given in canonical form, we have the marginal distributions. If they are in affine form, we may use canonic (or the function form canonicf) to obtain the marginal distributions.

Once we have both marginal distributions, we use an m-procedure we call *icalc*. Formation of the joint probability matrix is simply a matter of determining all the joint probabilities

$$p(i, j) = P(X = t_i, Y = u_j) = P(X = t_i)P(Y = u_j)$$

Once these are calculated, formation of the calculation matrices t and u is achieved exactly as in jcalc.

Example 9.1.8: Use of icalc to set up for joint calculations

```
X = [-4 -2 0 1 3];
Y = [0 1 2 4];
PX = 0.01*[12 18 27 19 24];
PY = 0.01*[15 43 31 11];
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
disp(P)                        % Optional display of the joint matrix
    0.0132    0.0198    0.0297    0.0209    0.0264
    0.0372    0.0558    0.0837    0.0589    0.0744
    0.0516    0.0774    0.1161    0.0817    0.1032
    0.0180    0.0270    0.0405    0.0285    0.0360
disp(t)                        % Calculation matrix t
    -4    -2     0     1     3
    -4    -2     0     1     3
    -4    -2     0     1     3
    -4    -2     0     1     3
disp(u)                        % Calculation matrix u
     4     4     4     4     4
     2     2     2     2     2
     1     1     1     1     1
     0     0     0     0     0
M = (t>=-3)&(t<=2);            % M = [-3, 2]
PM = total(M.*P)               % P(X in M)
PM =   0.6400
N = (u>0)&(u.^2<=15);          % N = {u: u > 0, u^2 <= 15}
PN = total(N.*P)               % P(Y in N)
```

```
PN =   0.7400
Q = M&N;                        % Rectangle MxN
PQ = total(Q.*P)                % P((X,Y) in MxN)
PQ =   0.4736
p = PM*PN
p  =   0.4736                   % P((X,Y) in MxN) = P(X in M)P(Y in N)
```

As an example, consider again the problem of joint Bernoulli trials described in the treatment of 4.3 Composite trials.

Example 9.1.9: The joint Bernoulli trial of Example 4.9

1 Bill and Mary take ten basketball free throws each. We assume the two seqences of trials are independent of each other, and each is a Bernoulli sequence.

Mary: Has probability 0.80 of success on each trial.

Bill: Has probability 0.85 of success on each trial.

What is the probability Mary makes more free throws than Bill?

Solution

Let X be the number of goals that Mary makes and Y be the number that Bill makes. Then X ~ binomial (10, 0.8) and Y ~ binomial (10, 0.85).

```
X = 0:10;
Y = 0:10;
PX = ibinom(10,0.8,X);
PY = ibinom(10,0.85,Y);
icalc
Enter row matrix of X-values  X  % Could enter 0:10
Enter row matrix of Y-values  Y  % Could enter 0:10
Enter X probabilities  PX        % Could enter ibinom(10,0.8,X)
Enter Y probabilities  PY        % Could enter ibinom(10,0.85,Y)
 Use array operations on matrices X, Y, PX, PY, t, u, and P
PM = total((t>u).*P)
PM =  0.2738                     % Agrees with solution in Example 9 from "Composite
Trials".
Pe = total((u==t).*P)            % Additional information is more easily
Pe =  0.2276                     % obtained than in the event formulation
Pm = total((t>=u).*P)            % of Example 9 from "Composite Trials".
Pm =  0.5014
```

Example 9.1.10: Sprinters time trials

Twelve world class sprinters in a meet are running in two heats of six persons each. Each runner has a reasonable chance of breaking the track record. We suppose results for individuals are independent.

First heat probabilities: 0.61 0.73 0.55 0.81 0.66 0.43

Second heat probabilities: 0.75 0.48 0.62 0.58 0.77 0.51

Compare the two heats for numbers who break the track record.

Solution

Let X be the number of successes in the first heat and Y be the number who are successful in the second heat. Then the pair $\{X, Y\}$ is independent. We use the m-function canonicf to determine the distributions for X and for Y, then icalc to get the joint distribution.

```
c1 = [ones(1,6) 0];
c2 = [ones(1,6) 0];
P1 = [0.61 0.73 0.55 0.81 0.66 0.43];
P2 = [0.75 0.48 0.62 0.58 0.77 0.51];
[X,PX] = canonicf(c1,minprob(P1));
[Y,PY] = canonicf(c2,minprob(P2));
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
Pm1 = total((t>u).*P)   % Prob first heat has most
Pm1 =  0.3986
Pm2 = total((u>t).*P)   % Prob second heat has most
Pm2 =  0.3606
Peq = total((t==u).*P)  % Prob both have the same
Peq =  0.2408
Px3 = (X>=3)*PX'        % Prob first has 3 or more
Px3 =  0.8708
Py3 = (Y>=3)*PY'        % Prob second has 3 or more
Py3 =  0.8525
```

As in the case of jcalc, we have an m-function version *icalcf*

[x, y, t, u, px, py, p] = icalcf(X, Y, PX, PY)\)

We have a related m-function *idbn* for obtaining the joint probability matrix from the marginal probabilities. Its formation of the joint matrix utilizes the same operations as icalc.

Example 9.1.11: A numerical example

```
PX = 0.1*[3 5 2];
PY = 0.01*[20 15 40 25];
P  = idbn(PX,PY)
P =
    0.0750    0.1250    0.0500
    0.1200    0.2000    0.0800
    0.0450    0.0750    0.0300
    0.0600    0.1000    0.0400
```

An m- procedure *itest* checks a joint distribution for independence. It does this by calculating the marginals, then forming an independent joint test matrix, which is compared with the original. We do not ordinarily exhibit the matrix P to be tested. However, this is a case in which the product rule holds for most of the minterms, and it would be very difficult to pick out those for which it fails. The m-procedure simply checks all of them.

```
idemo1                          % Joint matrix in datafile idemo1
P =  0.0091  0.0147  0.0035  0.0049  0.0105  0.0161  0.0112
     0.0117  0.0189  0.0045  0.0063  0.0135  0.0207  0.0144
```

```
     0.0104  0.0168  0.0040  0.0056  0.0120  0.0184  0.0128
     0.0169  0.0273  0.0065  0.0091  0.0095  0.0299  0.0208
     0.0052  0.0084  0.0020  0.0028  0.0060  0.0092  0.0064
     0.0169  0.0273  0.0065  0.0091  0.0195  0.0299  0.0208
     0.0104  0.0168  0.0040  0.0056  0.0120  0.0184  0.0128
     0.0078  0.0126  0.0030  0.0042  0.0190  0.0138  0.0096
     0.0117  0.0189  0.0045  0.0063  0.0135  0.0207  0.0144
     0.0091  0.0147  0.0035  0.0049  0.0105  0.0161  0.0112
     0.0065  0.0105  0.0025  0.0035  0.0075  0.0115  0.0080
     0.0143  0.0231  0.0055  0.0077  0.0165  0.0253  0.0176
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent   % Result of test
```

```
To see where the product rule fails, call for D
disp(D)                          % Optional call for D
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
     1     1     1     1     1     1     1
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
     1     1     1     1     1     1     1
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
     0     0     0     0     0     0     0
```

Next, we consider an example in which the pair is known to be independent.

```
jdemo3      % call for data in m-file
disp(P)     % call to display P
     0.0132    0.0198    0.0297    0.0209    0.0264
     0.0372    0.0558    0.0837    0.0589    0.0744
     0.0516    0.0774    0.1161    0.0817    0.1032
     0.0180    0.0270    0.0405    0.0285    0.0360

itest
Enter matrix of joint probabilities  P
The pair {X,Y} is independent       % Result of test
```

The procedure icalc can be extended to deal with an independent class of three random variables. We call the m-procedure *icalc3*. The following is a simple example of its use.

Example 9.1.14: Calculations for three independent random variables

```
X = 0:4;
Y = 1:2:7;
Z = 0:3:12;
PX = 0.1*[1 3 2 3 1];
PY = 0.1*[2 2 3 3];
PZ = 0.1*[2 2 1 3 2];
```

```
icalc3
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter row matrix of Z-values  Z
Enter X probabilities  PX
Enter Y probabilities  PY
Enter Z probabilities  PZ
Use array operations on matrices X, Y, Z,
PX, PY, PZ, t, u, v, and P
G = 3*t + 2*u - 4*v;        % W = 3X + 2Y -4Z
[W,PW] = csort(G,P);        % Distribution for W
PG = total((G>0).*P)        % P(g(X,Y,Z) > 0)
PG =  0.3370
Pg = (W>0)*PW'            % P(Z > 0)
Pg =  0.3370
```

An m-procedure *icalc4* to handle an independent class of four variables is also available. Also several variations of the m-function *mgsum* and the m-function *diidsum* are used for obtaining distributions for sums of independent random variables. We consider them in various contexts in other units.

Approximation for the absolutely continuous case

In the study of functions of random variables, we show that an approximating simple random variable X_s of the type we use is a function of the random variable X which is approximated. Also, we show that if $\{X, Y\}$ is an independent pair, so is $\{g(X), h(Y)\}$ for any reasonable functions g and h. Thus if $\{X, Y\}$ is an independent pair, so is any pair of approximating simple functions $\{X_s, Y_s\}$ of the type considered. Now it is theoretically possible for the approximating pair $\{X_s, Y_s\}$ to be independent, yet have the approximated pair $\{X, Y\}$ not independent. But *this is highly unlikely*. For all practical purposes, we may consider $\{X, Y\}$ to be independent iff $\{X_s, Y_s\}$ is independent. When in doubt, consider a second pair of approximating simple functions with more subdivision points. This decreases even further the likelihood of a false indication of independence by the approximating random variables.

Example 9.1.15: An independent pair

Suppose X ~ exponential (3) and Y ~ exponential (2) with

$$f_{XY}(t, u) = 6e^{-3t}e^{-2u} = 6e^{-(3t+2u)} \; t \geq 0, u \geq 0$$

Since $e^{-12} \approx 6 \times 10^{-6}$, we approximate X for values up to 4 and Y for values up to 6.

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 4]
Enter matrix [c d] of Y-range endpoints  [0 6]
Enter number of X approximation points  200
Enter number of Y approximation points  300
Enter expression for joint density  6*exp(-(3*t + 2*u))
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is independent
```

Example 9.1.16: Test for independence

The pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = 4tu \ 0 \le t \le 1, \ 0 \le u \le 1$. It is easy enough to determine the marginals in this case. By symmetry, they are the same.

$$f_X(t) = 4t \int_0^1 u du = 2t, \ 0 \le t \le 1$$

so that $f_{XY} = f_X f_Y$ which ensures the pair is independent. Consider the solution using tuappr and itest.

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  100
Enter number of Y approximation points  100
Enter expression for joint density  4*t.*u
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is independent
```

This page titled 9.1: Independent Classes of Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

9.2: Problems on Independent Classes of Random Variables

Exercise 9.2.1

The pair $\{X, Y\}$ has the joint distribution (in m-file npr08_06.m):

$$X = [-2.3\ -0.7\ 1.1\ 3.9\ 5.1]\ Y = [1.3\ 2.5\ 4.1\ 5.3]$$

$$P = \begin{bmatrix} 0.0483 & 0.0357 & 0.0420 & 0.0399 & 0.0441 \\ 0.0437 & 0.0323 & 0.0380 & 0.0361 & 0.0399 \\ 0.0713 & 0.0527 & 0.0620 & 0.0609 & 0.0551 \\ 0.0667 & 0.0493 & 0.0580 & 0.0651 & 0.0589 \end{bmatrix} \tag{9.2.1}$$

Determine whether or not the pair $\{X, Y\}$ is independent.

Answer available at the end of this section.

Exercise 9.2.2

The pair $\{X, Y\}$ has the joint distribution (in m-file npr09_02.m):

$$X = [-3.9\ -1.7\ 1.5\ 2.8\ 4.1]\ Y = [-2\ 1\ 2.6\ 5.1]$$

$$P = \begin{bmatrix} 0.0589 & 0.0342 & 0.0304 & 0.0456 & 0.0209 \\ 0.0961 & 0.0556 & 0.0498 & 0.0744 & 0.0341 \\ 0.0682 & 0.0398 & 0.0350 & 0.0528 & 0.0242 \\ 0.0868 & 0.0504 & 0.0448 & 0.0672 & 0.0308 \end{bmatrix} \tag{9.2.2}$$

Determine whether or not the pair $\{X, Y\}$ is independent.

Answer available at the end of this section.

Exercise 9.2.3

The pair $\{X, Y\}$ has the joint distribution (in m-file npr08_07.m):

$$P(X = t, Y = u)$$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189
-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

Determine whether or not the pair $\{X, Y\}$ is independent.

Answer available at the end of this section.

For the distributions in Exercises 4-10 below

a. Determine whether or not the pair is independent.
b. Use a discrete approximation and an independence test to verify results in part (a).

Exercise 9.2.4

$f_{XY}(t,u) = 1/\pi$ on the circle with radius one, center at (0,0).

Answer available at the end of this section.

Exercise 9.2.5

$f_{XY}(t,u) = 1/2$ on the square with vertices at (1, 0), (2, 1), (1, 2), (0, 1) (see Exercise 11 from "Problems on Random Vectors and Joint Distributions").

Answer available at the end of this section.

Exercise 9.2.6

$f_{XY}(t,u) = 4t(1-u)$ for $0 \le t \le 1$, $0 \le u \le 1$ (see Exercise 12 from "Problems on Random Vectors and Joint Distributions").

From the solution for Exercise 12 from "Problems on Random Vectors and Joint Distributions" we have

$$f_X(t) = 2t,\ 0 \le t \le 1,\ f_Y(u) = 2(1-u),\ 0 \le u \le 1,\ f_{XY} = f_X f_Y$$

so the pair is independent.

Answer available at the end of this section.

Exercise 9.2.7

$f_{XY} = \frac{1}{8}(t+u)$ for $0 \le t \le 2$, $0 \le u \le 2$ (see Exercise 13 from "Problems on Random Vectors and Joint Distributions").

From the solution of Exercise 13 from "Problems on Random Vectors and Joint Distributions" we have

$$f_X(t) = f_Y(t) = \frac{1}{4}(t+1),\ 0 \le t \le 2$$

so $f_{XY} \neq f_X f_Y$ which implies the pair is not independent.

Answer available at the end of this section.

Exercise 9.2.8

$f_{XY}(t,u) = 4ue^{-2t}$ for $0 \le t, 0 \le u \le 1$ (see Exercise 14 from "Problems on Random Vectors and Joint Distributions").

From the solution for Exercise 14 from "Problems on Random Vectors and Joint Distribution" we have

$$f_X(t) = 2e^{-2t},\ 0 \le t,\ f_Y(u) = 2u,\ 0 \le u \le 1$$

so that $f_{XY} = f_X f_Y$ and the pair is independent.

Answer available at the end of this section.

Exercise 9.2.9

$f_{XY}(t,u) = 12t^2u$ on the parallelogram with vertices (-1, 0), (0, 0), (1, 1), (0, 1)

(see Exercise 16 from "Problems on Random Vectors and Joint Distributions").

Answer available at the end of this section.

Exercise 9.2.10

$f_{XY} = \frac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2-t\}$ (see Exercise 17 from "Problems on Random Vectors and Joint Distributions").

Answer available at the end of this section.

Exercise 9.2.11

Two software companies, MicroWare and BusiCorp, are preparing a new business package in time for a computer trade show 180 days in the future. They work independently. MicroWare has anticipated completion time, in days, exponential (1/150). BusiCorp has time to completion, in days, exponential (1/130). What is the probability both will complete on time; that at least one will complete on time; that neither will complete on time?

Answer available at the end of this section.

Exercise 9.2.12

Eight similar units are put into operation at a given time. The time to failure (in hours) of each unit is exponential (1/750). If the units fail independently, what is the probability that five or more units will be operating at the end of 500 hours?

Answer available at the end of this section.

Exercise 9.2.13

The location of ten points along a line may be considered iid random variables with symmytric triangular distribution on [1,3]. What is the probability that three or more will lie within distance 1/2 of the point $t = 2$?

Answer available at the end of this section.

Exercise 9.2.14

A Christmas display has 200 lights. The times to failure are iid, exponential (1/10000). The display is on continuously for 750 hours (approximately one month). Determine the probability the number of lights which survive the entire period is at least 175, 180, 185, 190.

Answer available at the end of this section.

Exercise 9.2.15

A critical module in a network server has time to failure (in hours of machine time) exponential (1/3000). The machine operates continuously, except for brief times for maintenance or repair. The module is replaced routinely every 30 days (720 hours), unless failure occurs. If successive units fail independently, what is the probability of no breakdown due to the module for one year?

Answer available at the end of this section.

Exercise 9.2.16

Joan is trying to decide which of two sales opportunities to take.

- In the first, she makes three independent calls. Payoffs are \$570, \$525, and \$465, with respective probabilities of 0.57, 0.41, and 0.35.
- In the second, she makes eight independent calls, with probability of success on each call $p = 0.57$. She realizes \$150 profit on each successful sale.

Let X be the net profit on the first alternative and Y be the net gain on the second. Assume the pair $\{X, Y\}$ is independent.

a. Which alternative offers the maximum possible gain?
b. Compare probabilities in the two schemes that total sales are at least \$600, \$900, \$1000, \$1100.
c. What is the probability the second exceeds the first— i.e., what is $P(Y > X)$?

Answer

$X = 570I_A + 525I_B + 465I_C$ with $[P(A)P(B)P(C)]$ = [0.57 0.41 0.35]. $Y = 150S$. where S binomial (8, 0.57).

```
c = [570 525 465 0];
pm = minprob([0.57 0.41 0.35]);
canonic                              % Distribution for X
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
Y = 150*[0:8];                       % Distribution for Y
PY = ibinom(8,0.57,0:8);
icalc                                % Joint distribution
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
xmax = max(X)
xmax =   1560
ymax = max(Y)
ymax =   1200
k = [600 900 1000 1100];
px = zeros(1,4);


for i = 1:4
    px(i) = (X>=k(i))*PX';
end
```

```
py = zeros(1,4);
for i = 1:4
  py(i) = (Y>=k(i))*PY';
end
disp([px;py]')
    0.4131    0.7765
    0.4131    0.2560
    0.3514    0.0784
    0.0818    0.0111
M = u > t;
PM = total(M.*P)
PM = 0.5081          % P(Y>X)
```

Exercise 9.2.17

Margaret considers five purchases in the amounts 5, 17, 21, 8, 15 dollars with respective probabilities 0.37, 0.22, 0.38, 0.81, 0.63. Anne contemplates six purchases in the amounts 8, 15, 12, 18, 15, 12 dollars. with respective probabilities 0.77, 0.52, 0.23, 0.41, 0.83, 0.58. Assume that all eleven possible purchases form an independent class.

a. What is the probability Anne spends at least twice as much as Margaret?
b. What is the probability Anne spends at least $30 more than Margaret?

Answer

```
cx = [5 17 21 8 15 0];
pmx = minprob(0.01*[37 22 38 81 63]);
cy = [8 15 12 18 15 12 0];
pmy = minprob(0.01*[77 52 23 41 83 58]);
[X,PX] = canonicf(cx,pmx);
[Y,PY] = canonicf(cy,pmy);
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
M1 = u >= 2*t;
PM1 = total(M1.*P)
PM1 =  0.3448
M2 = u - t >=30;
PM2 = total(M2.*P)
PM2 =  0.2431
```

Exercise 9.2.18

James is trying to decide which of two sales opportunities to take.

- In the first, he makes three independent calls. Payoffs are $310, $380, and $350, with respective probabilities of 0.35, 0.41, and 0.57.
- In the second, he makes eight independent calls, with probability of success on each call p=0.57. He realizes $100 profit on each successful sale.

Let X be the net profit on the first alternative and Y be the net gain on the second. Assume the pair $\{X, Y\}$ is independent.

- Which alternative offers the maximum possible gain?

- What is the probability the second exceeds the first— i.e., what is $P(Y > X)$?
- Compare probabilities in the two schemes that total sales are at least $600, $700, $750.

Answer available at the end of this section.

Exercise 9.2.19

A residential College plans to raise money by selling "chances" on a board. There are two games:

Game 1: Pay $5 to play; win $20 with probability p_1 =0.05 (one in twenty)
Game 2: Pay $10 to play; win $30 with probability p_2 =0.2 (one in five)

Thirty chances are sold on Game 1 and fifty chances are sold on Game 2. If X and Y are the profits on the respective games, then

$$X = 30 \cdot 5 - 20N_1 \text{ and } Y = 50 \cdot 10 - 30N_2$$

where N_1, N_2 are the numbers of winners on the respective games. It is reasonable to suppose N_1 binomial (30, 0.05) and N_2 binomial (50, 0.2). It is reasonable to suppose the pair $\{N_1, N_2\}$ is independent, so that $\{X, Y\}$ is independent. Determine the marginal distributions for X and Y then use icalc to obtain the joint distribution and the calculating matrices. The total profit for the College is $Z = X + Y$. What is the probability the College will lose money? What is the probability the profit will be $400 or more, less than $200, between $200 and $450?

Answer

```
N1 = 0:30;
PN1 = ibinom(30,0.05,0:30);
x  = 150 - 20*N1;
[X,PX] = csort(x,PN1);
N2 = 0:50;
PN2 = ibinom(50,0.2,0:50);
y  = 500 - 30*N2;
[Y,PY] = csort(y,PN2);
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = t + u;
Mlose = G < 0;
Mm400  = G >= 400;
Ml200  = G < 200;
M200_450  = (G>=200)&(G<=450);
Plose = total(Mlose.*P)
Plose =   3.5249e-04
Pm400 = total(Mm400.*P)
Pm400 =   0.1957
Pl200 = total(Ml200.*P)
Pl200 =
    0.0828
P200_450 = total(M200_450.*P)
P200_450 =  0.8636
```

Exercise 9.2.20

The class $\{X, Y, Z\}$ of random variables is iid (independent, identically distributed) with common distribution

$$X = \text{[-5 -1 3 4 7]}\ PX = \text{0.01 * [15 20 30 25 10]}$$

Let $W = 3X - 4Y + 2Z$. Determine the distribution for W and from this determine $P(W > 0)$ and $P(-20 \leq W \leq 10)$. Do this with icalc, then repeat with icalc3 and compare results.

Answer

Since icalc uses X and PX in its output, we avoid a renaming problem by using x and px for data vectors X and PX.

```
x = [-5 -1 3 4 7];
px = 0.01*[15 20 30 25 10];
icalc
Enter row matrix of X-values  3*x
Enter row matrix of Y-values  -4*x
Enter X probabilities  px
Enter Y probabilities  px
 Use array operations on matrices X, Y, PX, PY, t, u, and P
a = t + u;
[V,PV] = csort(a,P);
icalc
Enter row matrix of X-values  V
Enter row matrix of Y-values  2*x
Enter X probabilities  PV
Enter Y probabilities  px
 Use array operations on matrices X, Y, PX, PY, t, u, and P
b = t + u;
[W,PW] = csort(b,P);
P1 = (W>0)*PW'
P1 =  0.5300
P2 = ((-20<=W)&(W<=10))*PW'
P2 =  0.5514
icalc3                          % Alternate using icalc3
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter row matrix of Z-values  x
Enter X probabilities  px
Enter Y probabilities  px
Enter Z probabilities  px
Use array operations on matrices X, Y, Z,
PX, PY, PZ, t, u, v, and P
a = 3*t - 4*u + 2*v;
[W,PW] = csort(a,P);
P1 = (W>0)*PW'
P1 = 0.5300
P2 = ((-20<=W)&(W<=10))*PW'
P2 = 0.5514
```

Exercise 9.2.21

The class $\{A, B, C, D, E, F\}$ is independent; the respective probabilities for these events are $\{0.46, 0.27, 0.33, 0.47, 0.37, 0.41\}$. Consider the simple random variables

$$X = 3I_A - 9I_B + 4I_C,\ Y = -2I_D + 6I_E + 2I_F - 3, \text{ and } Z = 2X - 3Y$$

Determine $P(Y > X)$, $P(Z > 0)$, $P(5 \leq Z \leq 25)$.

Answer available at the end of this section.

Exercise 9.2.22

Two players, Ronald and Mike, throw a pair of dice 30 times each. What is the probability Mike throws more "sevens" than does Ronald?

Answer available at the end of this section.

Exercise 9.2.23

A class has fifteen boys and fifteen girls. They pair up and each tosses a coin 20 times. What is the probability that at least eight girls throw more heads than their partners?

Answer available at the end of this section.

Exercise 9.2.24

Glenn makes five sales calls, with probabilities 0.37, 0.52, 0.48, 0.71, 0.63, of success on the respective calls. Margaret makes four sales calls with probabilities 0.77, 0.82, 0.75, 0.91, of success on the respective calls. Assume that all nine events form an independent class. If Glenn realizes a profit of \$18.00 on each sale and Margaret earns \$20.00 on each sale, what is the probability Margaret's gain is at least \$10.00 more than Glenn's?

Answer available at the end of this section.

Exercise 9.2.25

Mike and Harry have a basketball shooting contest.

- Mike shoots 10 ordinary free throws, worth two points each, with probability 0.75 of success on each shot.
- Harry shoots 12 "three point" shots, with probability 0.40 of success on each shot.

Let X, Y be the number of points scored by Mike and Harry, respectively. Determine $P(X \geq 15)$, and $P(Y \geq 15)$, $P(X \geq Y)$.

Answer available at the end of this section.

Exercise 9.2.26

Martha has the choice of two games.

Game 1: Pay ten dollars for each “play.” If she wins, she receives $20, for a net gain of $10 on the play; otherwise, she loses her $10. The probability of a win is 1/2, so the game is “fair.”
Game 2: Pay five dollars to play; receive $15 for a win. The probability of a win on any play is 1/3.

Martha has $100 to bet. She is trying to decide whether to play Game 1 ten times or Game 2 twenty times. Let $W1$ and $W2$ be the respective net winnings (payoff minus fee to play).

- Determine $P(W2 \geq W1)$
- Compare the two games further by calculating $P(W1 > 0)$ and $P(W2 > 0)$

Which game seems preferable?

Answer available at the end of this section.

Exercise 9.2.27

Jim and Bill of the men's basketball team challenge women players Mary and Ellen to a free throw contest. Each takes five free throws. Make the usual independence assumptions. Jim, Bill, Mary, and Ellen have respective probabilities $p = 0.82$, 0.87, 0.80, and 0.85 of making each shot tried. What is the probability Mary and Ellen make a total number of free throws at least as great as the total made by the guys?

Answer available at the end of this section.

Answers

Exercise 9.2.1

```
npr08_06
Data are in X, Y, P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
disp(D)
     0     0     0     1     1
     0     0     0     1     1
     1     1     1     1     1
     1     1     1     1     1
```

Exercise 9.2.2

```
npr09_02
Data are in X, Y, P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
disp(D)
     0     0     0     0     0
     0     1     1     0     0
     0     1     1     0     0
     0     0     0     0     0
```

Exercise 9.2.3

```
npr08_07
```

```
Data are in X, Y, P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
disp(D)
     1     1     1     1     1     1
     1     1     1     1     1     1
     1     1     1     1     1     1
     1     1     1     1     1     1
```

Exercise 9.2.4

Not independent by the rectangle test.

```
tuappr
Enter matrix [a b] of X-range endpoints  [-1 1]
Enter matrix [c d] of Y-range endpoints  [-1 1]
Enter number of X approximation points  100
Enter number of Y approximation points  100
Enter expression for joint density  (1/pi)*(t.^2 + u.^2<=1)
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D  % Not practical-- too large
```

Exercise 9.2.5

Not independent, by the rectangle test.

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (1/2)*(u<=min(1+t,3-t)).* ...
   (u>=max(1-t,t-1))
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
```

Exercise 9.2.6

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  100
Enter number of Y approximation points  100
Enter expression for joint density  4*t.*(1-u)
Use array operations on X, Y, PX, PY, t, u, and P



itest
Enter matrix of joint probabilities  P
The pair {X,Y} is independent
```

Exercise 9.2.7

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  100
Enter number of Y approximation points  100
Enter expression for joint density  (1/8)*(t+u)
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
```

Exercise 9.2.8

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 5]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  500
Enter number of Y approximation points  100
Enter expression for joint density  4*u.*exp(-2*t)
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is independent       % Product rule holds to within 10^{-9}
```

Exercise 9.2.9

Not independent by the rectangle test.

```
tuappr
Enter matrix [a b] of X-range endpoints  [-1 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  100
Enter expression for joint density  12*t.^2.*u.*(u<=min(t+1,1)).* ...
          (u>=max(0,t))
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
```

Exercise 9.2.10

By the rectangle test, the pair is not independent.

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  100
Enter expression for joint density  (24/11)*t.*u.*(u<=min(1,2-t))
Use array operations on X, Y, PX, PY, t, u, and P
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is NOT independent
To see where the product rule fails, call for D
```

Exercise 9.2.11

```
p1 = 1 - exp(-180/150)
p1 =  0.6988
```

```
p2 = 1 - exp(-180/130)
p2 =  0.7496
Pboth = p1*p2
Pboth =  0.5238
Poneormore = 1 - (1 - p1)*(1 - p2) % 1 - Pneither
Poneormore =  0.9246
Pneither = (1 - p1)*(1 - p2)
Pneither =    0.0754
```

Exercise 9.2.12

```
p = exp(-500/750);  % Probability any one will survive
P = cbinom(8,p,5)   % Probability five or more will survive
P =  0.3930
```

Exercise 9.2.13

Geometrically, $p = 3/4$, so that `P = cbinom(10,p,3) = 0.9996`.

Exercise 9.2.14

```
p = exp(-750/10000)
p =  0.9277
k = 175:5:190;
P = cbinom(200,p,k);
disp([k;P]')
  175.0000    0.9973
  180.0000    0.9449
  185.0000    0.6263
  190.0000    0.1381
```

Exercise 9.2.15

```
p = exp(-720/3000)
p = 0.7866     % Probability any unit survives
P = p^12    % Probability all twelve survive (assuming 12 periods)
P = 0.056
```

Exercise 9.2.18

```
cx = [310 380 350 0];
pmx = minprob(0.01*[35 41 57]);
Y  = 100*[0:8];
PY = ibinom(8,0.57,0:8);
canonic
 Enter row vector of coefficients  cx
 Enter row vector of minterm probabilities  pmx
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
xmax = max(X)
xmax =  1040
ymax = max(Y)
ymax =   800
PYgX = total((u>t).*P)
PYgX =  0.5081
k = [600 700 750];
```

```
px = zeros(1,3);
py = zeros(1,3);
for i = 1:3
    px(i) = (X>=k(i))*PX';
end
for i = 1:3
    py(i) = (Y>=k(i))*PY';
end
disp([px;py]')
    0.4131    0.2560
    0.2337    0.0784
    0.0818    0.0111
```

Exercise 9.2.21

```
cx = [3 -9 4 0];
pmx = minprob(0.01*[42 27 33]);
cy = [-2 6 2 -3];
pmy = minprob(0.01*[47 37 41]);
[X,PX] = canonicf(cx,pmx);
[Y,PY] = canonicf(cy,pmy);
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = 2*t - 3*u;
[Z,PZ] = csort(G,P);
PYgX = total((u>t).*P)
PYgX =  0.3752
PZpos = (Z>0)*PZ'
PZpos = 0.5654
P5Z25 = ((5<=Z)&(Z<=25))*PZ'
P5Z25 = 0.4745
```

Exercise 9.2.22

```
P = (ibinom(30,1/6,0:29))*(cbinom(30,1/6,1:30))' = 0.4307
```

Exercise 9.2.23

```
pg = (ibinom(20,1/2,0:19))*(cbinom(20,1/2,1:20))'
pg =  0.4373            % Probability each girl throws more
P = cbinom(15,pg,8)
P =   0.3100            % Probability eight or more girls throw more
```

Exercise 9.2.24

```
cg = [18*ones(1,5) 0];
cm = [20*ones(1,4) 0];
pmg = minprob(0.01*[37 52 48 71 63]);
pmm = minprob(0.01*[77 82 75 91]);
[G,PG] = canonicf(cg,pmg);
[M,PM] = canonicf(cm,pmm);
icalc
Enter row matrix of X-values  G
Enter row matrix of Y-values  M
Enter X probabilities  PG
Enter Y probabilities  PM
 Use array operations on matrices X, Y, PX, PY, t, u, and P
```

```
H = u-t>=10;
p1 = total(H.*P)
p1 =  0.5197
```

Exercise 9.2.25

```
X = 2*[0:10];
PX = ibinom(10,0.75,0:10);
Y = 3*[0:12];
PY = ibinom(12,0.40,0:12);
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
PX15 = (X>=15)*PX'
PX15 = 0.5256
PY15 = (Y>=15)*PY'
PY15 = 0.5618
G = t>=u;
PG = total(G.*P)
PG =   0.5811
```

Exercise 9.2.26

```
W1 = 20*[0:10] - 100;
PW1 = ibinom(10,1/2,0:10);
W2 = 15*[0:20] - 100;
PW2 = ibinom(20,1/3,0:20);
P1pos = (W1>0)*PW1'
P1pos = 0.3770
P2pos = (W2>0)*PW2'
P2pos = 0.5207
icalc
Enter row matrix of X-values  W1
Enter row matrix of Y-values  W2
Enter X probabilities  PW1
Enter Y probabilities  PW2
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = u >= t;
PG = total(G.*P)
PG =  0.5182
```

Exercise 9.2.27

```
x = 0:5;
PJ = ibinom(5,0.82,x);
PB = ibinom(5,0.87,x);
PM = ibinom(5,0.80,x);
PE = ibinom(5,0.85,x);

icalc
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter X probabilities  PJ
Enter Y probabilities  PB
 Use array operations on matrices X, Y, PX, PY, t, u, and P
H = t+u;
```

```
[Tm,Pm] = csort(H,P);
icalc
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter X probabilities  PM
Enter Y probabilities  PE
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = t+u;
[Tw,Pw] = csort(G,P);
icalc
Enter row matrix of X-values  Tm
Enter row matrix of Y-values  Tw
Enter X probabilities  Pm
Enter Y probabilities  Pw
 Use array operations on matrices X, Y, PX, PY, t, u, and P
Gw = u>=t;
PGw = total(Gw.*P)
PGw = 0.5746

icalc4               % Alternate using icalc4
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter row matrix of Z-values  x
Enter row matrix of W-values  x
Enter X probabilities  PJ
Enter Y probabilities  PB
Enter Z probabilities  PM
Enter W probabilities  PE
Use array operations on matrices X, Y, Z,W
PX, PY, PZ, PW t, u, v, w, and P
H = v+w >= t+u;
PH = total(H.*P)
PH =  0.5746
```

This page titled 9.2: Problems on Independent Classes of Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

10: FUNCTIONS OF RANDOM VARIABLES

Chapter 10. Functions of Random Variables

10.1: Functions of a Random Variable

10.2: Function of Random Vectors

10.3: The Quantile Function

10.4: Problems on Functions of Random Variables

This page titled 10: Functions of Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Functions of a Random Variable

Introduction

Frequently, we observe a value of some random variable, but are really interested in a value derived from this by a function rule. If X is a random variable and g is a reasonable function (technically, a *Borel function*), then $Z = g(X)$ is a new random variable which has the value $g(t)$ for any ω such that $X(\omega) = t$. Thus $Z(\omega) = g(X(\omega))$.

The problem; an approach

We consider, first, functions of a single random variable. A wide variety of functions are utilized in practice.

Example 10.1.1: A quality control problem

In a quality control check on a production line for ball bearings it may be easier to weigh the balls than measure the diameters. If we can assume true spherical shape and w is the weight, then diameter is $kw^{1/3}$, where k is a factor depending upon the formula for the volume of a sphere, the units of measurement, and the density of the steel. Thus, if X is the weight of the sampled ball, the desired random variable is $D = kX^{1/3}$.

Example 10.1.2: Price breaks

The cultural committee of a student organization has arranged a special deal for tickets to a concert. The agreement is that the organization will purchase ten tickets at \$20 each (regardless of the number of individual buyers). Additional tickets are available according to the following schedule:

- 11-20, \$18 each
- 21-30, \$16 each
- 31-50, \$15 each
- 51-100, \$13 each

If the number of purchasers is a random variable X, the total cost (in dollars) is a random quantity $Z = g(X)$ described by

$$g(X) = 200 + 18I_{M1}(X)(X - 10) + (16 - 18)I_{M2}(X)(X - 20)$$

$$+(15 - 16)I_{M3}(X)(X - 30) + (13 - 15)I_{M4}(X)(X - 50)$$

where $M1 = [10, \infty)$, $M2 = [20, \infty)$, $M3 = [30, \infty)$, $M4 = [50, \infty)$

The function rule is more complicated than in Example 10.1.1, but the essential problem is the same.

The problem

If X is a random variable, then $Z = g(X)$ is a new random variable. Suppose we have the distribution for X. How can we determine $P(Z \in M)$, the probability Z takes a value in the set M?

An approach to a solution

We consider two equivalent approaches

To find $P(X \in M)$.

1. *Mapping approach*. Simply find the amount of probability mass mapped into the set M by the random variable X.
   - In the absolutely continuous case, calculate $\int_M f_X$.

- In the discrete case, identify those values t_i of X which are in the set M and add the associated probabilities.

2. *Discrete alternative*. Consider each value t_i of X. Select those which meet the defining conditions for M and add the associated probabilities. This is the approach we use in the MATLAB calculations. Note that it is not necessary to describe geometrically the set M; merely use the defining conditions.

To find $P(g(X) \in M)$.

1. *Mapping approach*. Determine the set N of all those *t* which are mapped into M by the function g. Now if $X(\omega) \in N$, then $g(X(\omega)) \in M$, and if $g(X(\omega)) \in M$, then $X(\omega) \in N$. Hence

$$\{\omega : g(X(\omega)) \in M\} = \{\omega : X(\omega) \in N\}$$

Since these are the same event, they must have the same probability. Once N is identified, determine $P(X \in N)$ in the usual manner (see part a, above).

- *Discrete alternative*. For each possible value t_i of X, determine whether $g(t_i)$ meets the defining condition for M. Select those t_i which do and add the associated probabilities.

— □

Remark. The set N in the mapping approach is called the *inverse image* $N = g^{-1}(M)$

Example 10.1.3: A discrete example

Suppose X has values -2, 0, 1, 3, 6, with respective probabilities 0.2, 0.1, 0.2, 0.3 0.2.

Consider $Z = g(X) = (X + 1)(X - 4)$. Determine $P(Z > 0)$.

Solution

First solution. The mapping approach

$g(t) = (t + 1)(t - 4)$. $N = \{t : g(t) > 0\}$ is the set of points to the left of –1 or to the right of 4. The X-values –2 and 6 lie in this set. Hence

$$P(g(X) > 0) = P(X = -2) + P(X = 6) = 0.2 + 0.2 = 0.4$$

Second solution. The discrete alternative

X =	-2	0	1	3	6
P X =	0.2	0.1	0.2	0.3	0.2
Z =	6	-4	-6	-4	14
Z > 0	1	0	0	0	1

Picking out and adding the indicated probabilities, we have

$$P(Z > 0) = 0.2 + 0.2 = 0.4$$

In this case (and often for “hand calculations”) the mapping approach requires less calculation. However, for MATLAB calculations (as we show below), the discrete alternative is more readily implemented.

Example 10.1.4. An absolutely continuous example

Suppose X ~ uniform [–3,7]. Then $f_X(t) = 0.1$, $-3 \le t \le 7$ (and zero elsewhere). Let

$$Z = g(X) = (X + 1)(X - 4)$$

Determine $P(Z > 0)$.

Solution

First we determine $N = \{t : g(t) > 0\}$. As in Example 10.1.3, $g(t) = (t + 1)(t - 4) > 0$ for $t < -1$ or $t > 4$). Because of the uniform distribution, the integral of the density over any subinterval of $\{X, Y\}$ is 0.1 times the length of that subinterval. Thus, the desired probability is

$$P(g(X) > 0) = 0.1[(-1 - (-3)) + (7 - 4)] = 0.5$$

We consider, next, some important examples.

Example 10.1.5: The normal distribution and standardized normal distribution

To show that if $X \sim N(\mu, \sigma^2)$ then

$$Z = g(X) = \frac{X - \mu}{\sigma} \; N(0, 1)$$

VERIFICATION

We wish to show the denity function for Z is

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

Now

$$g(t) = \frac{t - \mu}{\sigma} \leq v \text{ iff } t \leq \sigma v + \mu$$

Hence, for given $M = (-\infty, v]$ the inverse image is $N = (-\infty, \sigma v + \mu]$, so that

$$F_Z(v) = P(Z \leq v) = P(Z \in M) = P(X \in N) = P(X \leq \sigma v + \mu) = F_X(\sigma v + \mu)$$

Since the density is the derivative of the distribution function,

$$f_Z(v) = F_Z'(v) = F_X'(v) = F_X'(\sigma v + \mu)\sigma = \sigma f_X(\sigma v + \mu)$$

Thus

$$f_Z(v) = \frac{\sigma}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{\sigma v + \mu - \mu}{\sigma})^2] = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} = \varphi(v)$$

We conclude that $Z \sim N(0, 1)$.

Example 10.1.1

Suppose X has distribution function F_X. If it is absolutely continuous, the corresponding density is f_X. Consider $Z = aX + b$. Here $g(t) = at + b$, an affine function (linear plus a constant). Determine the distribution function for Z (and the density in the absolutely continuous case).

Solution

$$F_Z(v) = P(Z \leq v) = P(aX + b \leq v)$$

There are two cases

- $a > 0$:

$$F_Z(v) = P(X \leq \frac{v-b}{a}) = F_X(\frac{v-b}{a})$$

- $a < 0$

$$F_Z(v) = P(X \geq \frac{v-b}{a}) = P(X > \frac{v-b}{a}) + P(X = \frac{v-b}{a})$$

So that

$$F_Z(v) = 1 - F_X(\frac{v-b}{a}) + P(X = \frac{v-b}{a})$$

For the absolutely continuous case, $P(X = \frac{v-b}{a}) = 0$, and by differentiation

- for $a > 0$ $f_Z(v) = \frac{1}{a} f_X(\frac{v-b}{a})$
- for $a < 0$ $f_Z(v) = -\frac{1}{a} f_X(\frac{v-b}{a})$

Since for $a < 0$, $-a = |a|$, the two cases may be combined into one formula.

$$f_Z(v) = \frac{1}{|a|} f_X(\frac{v-b}{a})$$

Example 10.1.7: Completion of normal and standardized normal relationship

Suppose $Z \sim N(0,1)$. show that $X = \sigma Z + \mu$ $(\sigma > 0)$ is $N(\mu, \sigma^2)$.

VERIFICATION

Use of the result of Example 10.1.6 on affine functions shows that

$$f_X(t) = \frac{1}{\sigma} \varphi(\frac{t-\mu}{\sigma}) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{t-\mu}{\sigma})^2]$$

Example 10.1.8: Fractional power of a nonnegative random variable

Suppose $X \geq 0$ and $Z = g(X) = X^{1/a}$ for $a > 1$. Since for $t \geq 0$, $t^{1/a}$ is increasing, we have $0 \leq t^{1/a} \leq v$ iff $0 \leq t \leq v^a$. Thus

$$F_Z(v) = P(Z \leq v) = P(X \leq v^a) = F_X(v^a)$$

In the absolutely continuous case

$$f_Z(v) = F_Z'(v) = f_X(v^a) a v^{a-1}$$

Example 10.1.9: Fractional power of an exponentially distributed random variable

Suppose $X \sim$ exponential (λ). Then $Z = X^{1/a} \sim$ Weibull $(a, \lambda, 0)$.

According to the result of Example 10.1.8,

$$F_Z(t) = F_X(t^a) = 1 - e^{-\lambda t^a}$$

which is the distribution function for Z ~ Weibull $(a, \lambda, 0)$.

Example 10.1.10: A simple approximation as a function of *X*

If X is a random variable, a simple function approximation may be constructed (see Distribution Approximations). We limit our discussion to the bounded case, in which the range of X is limited to a bounded interval $I = [a, b]$. Suppose I is partitioned into n subintervals by points t_i, $1 \le i \le n - 1$, with $a = t_0$ and $b = t_n$. Let $M_i = [t_{i-1}, t_i)$ be the ith subinterval, $1 \le i \le n - 1$ and $M_n = [t_{n-1}, t_n]$. Let $E_i = X^{-1}(M_i)$ be the set of points mapped into M_i by X. Then the E_i form a partition of the basic space Ω. For the given subdivision, we form a simple random variable X_s as follows. In each subinterval, pick a point $s_i, t_{i-1} \le s_i < t_i$. The simple random variable

$$X_s = \sum_{i=1}^{n} s_i I_{E_i}$$

approximates X to within the length of the largest subinterval M_i. Now $I_{E_i} = I_{M_i}(X)$, since $I_{E_i}(\omega) = 1$ iff $X(\omega) \in M_i$ iff $I_{M_i}(X(\omega)) = 1$. We may thus write

$$X_s = \sum_{i=1}^{n} s_i I_{M_i}(X), \text{ a function of } X$$

Use of MATLAB on simple random variables

For simple random variables, we use the discrete alternative approach, since this may be implemented easily with MATLAB. Suppose the distribution for X is expressed in the row vectors X and PX.

- We perform *array operations* on vector X to obtain

$$G = [g(t_1)g(t_2) \cdots g(t_n)]$$

- We use *relational* and *logical* operations on G to obtain a matrix M which has ones for those t_i (values of X) such that $g(t_i)$ satisfies the desired condition (and zeros elsewhere).
- The zero-one matrix M is used to select the the corresponding $p_i = P(X = t_i)$ and sum them by the taking the dot product of M and PX.

Example 10.1.11: Basic calculations for a function of a simple random variable

```
X = -5:10;                      % Values of X
PX = ibinom(15,0.6,0:15);       % Probabilities for X
G = (X + 6).*(X - 1).*(X - 8); % Array operations on X matrix to get G = g(X)
M = (G > - 100)&(G < 130);      % Relational and logical operations on G
PM = M*PX'                      % Sum of probabilities for selected values
PM =  0.4800
disp([X;G;M;PX]')               % Display of various matrices (as columns)
   -5.0000   78.0000    1.0000    0.0000
   -4.0000  120.0000    1.0000    0.0000
   -3.0000  132.0000         0    0.0003
   -2.0000  120.0000    1.0000    0.0016
   -1.0000   90.0000    1.0000    0.0074
         0   48.0000    1.0000    0.0245
    1.0000         0    1.0000    0.0612
    2.0000  -48.0000    1.0000    0.1181
    3.0000  -90.0000    1.0000    0.1771
    4.0000 -120.0000         0    0.2066
    5.0000 -132.0000         0    0.1859
```

```
    6.0000 -120.0000         0    0.1268
    7.0000  -78.0000    1.0000    0.0634
    8.0000         0    1.0000    0.0219
    9.0000  120.0000    1.0000    0.0047
   10.0000  288.0000         0    0.0005
[Z,PZ] = csort(G,PX);          % Sorting and consolidating to obtain
disp([Z;PZ]')                  % the distribution for Z = g(X)
 -132.0000    0.1859
 -120.0000    0.3334
  -90.0000    0.1771
  -78.0000    0.0634
  -48.0000    0.1181
         0    0.0832
   48.0000    0.0245
   78.0000    0.0000
   90.0000    0.0074
  120.0000    0.0064
  132.0000    0.0003
  288.0000    0.0005
P1 = (G<-120)*PX '             % Further calculation using G, PX
P1 =  0.1859
p1 = (Z<-120)*PZ'              % Alternate using Z, PZ
p1 =  0.1859
```

Example 10.1.12

$X = 10I_A + 18I_B + 10I_C$ with $\{A, B, C\}$ independent and $P = [0.6\ 0.3\ 0.5]$.

We calculate the distribution for X, then determine the distribution for

$$Z = X^{1/2} - X + 50$$

```
c = [10 18 10 0];
pm = minprob(0.1*[6 3 5]);
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
disp(XDBN)
         0    0.1400
   10.0000    0.3500
   18.0000    0.0600
   20.0000    0.2100
   28.0000    0.1500
   38.0000    0.0900
G = sqrt(X) - X + 50;       % Formation of G matrix
[Z,PZ] = csort(G,PX);       % Sorts distinct values of g(X)
disp([Z;PZ]')               % consolidates probabilities
   18.1644    0.0900
   27.2915    0.1500
   34.4721    0.2100
   36.2426    0.0600
   43.1623    0.3500
   50.0000    0.1400
```

```
M = (Z < 20)|(Z >= 40)      % Direct use of Z distribution
M =    1     0     0     0     1     1
PZM = M*PZ'
PZM =  0.5800
```

Remark. Note that with the m-function csort, we may name the output as desired.

Example 10.1.13: Continuation of example 10.1.12, above.

```
H = 2*X.^2 - 3*X + 1;
[W,PW] = csort(H,PX)
W  =     1      171     595     741    1485    2775
PW =  0.1400  0.3500  0.0600  0.2100  0.1500  0.0900
```

Example 10.1.14: A discrete approximation

Suppose X has density function $f_X(t) = \frac{1}{2}(3t^2 + 2t)$ for $0 \le t \le 1$. Then $F_X(t) = \frac{1}{2}(t^3 + t^2)$. Let $Z = X^{1/2}$. We may use the approximation m-procedure tappr to obtain an approximate discrete distribution. Then we work with the approximating random variable as a simple random variable. Suppose we want $P(Z \le 0.8)$. Now $Z \le 0.8$ iff $X \le 0.8^2 = 0.64$. The desired probability may be calculated to be

$$P(Z \le 0.8) = F_X(0.64) = (0.64^3 + 0.64^2)/2 = 0.3359$$

Using the approximation procedure, we have

```
tappr
Enter matrix [a b] of x-range endpoints  [0 1]
Enter number of x approximation points  200
Enter density as a function of t  (3*t.^2 + 2*t)/2
Use row matrices X and PX as in the simple case
G = X.^(1/2);
M = G <= 0.8;
PM = M*PX'
PM =   0.3359       % Agrees quite closely with the theoretical
```

This page titled 10.1: Functions of a Random Variable is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Function of Random Vectors

Introduction

The general mapping approach for a single random variable and the discrete alternative extends to functions of more than one variable. It is convenient to consider the case of two random variables, considered jointly. Extensions to more than two random variables are made similarly, although the details are more complicated.

The general approach extended to a pair

Consider a pair $\{X, Y\}$ having joint distribution on the plane. The approach is analogous to that for a single random variable with distribution on the line.

To find $P((X, Y) \in Q)$.

1. *Mapping approach*. Simply find the amount of probability mass mapped into the set Q on the plane by the random vector $W = (X, Y)$.
   - In the absolutely continuous case, calculate $\int \int_Q f_{XY}$.
   - In the discrete case, identify those vector values (t_i, u_j) of (X, Y) which are in the set Q and add the associated probabilities.
2. *Discrete alternative*. Consider each vector value (t_i, u_j) of (X, Y). Select those which meet the defining conditions for Q and add the associated probabilities. This is the approach we use in the MATLAB calculations. It does not require that we describe geometrically the region Q.

To find $P(g(X, Y) \in M)$. g is real valued and M is a subset the real line.

1. *Mapping approach*. Determine the set Q of all those (t, u) which are mapped into M by the function g. Now

$$W(\omega) = (X(\omega), Y(\omega)) \in Q \text{ iff } g((X(\omega), Y(\omega)) \in M \text{ Hence}$$

$$\{\omega : g(X(\omega), Y(\omega)) \in M\} = \{\omega : (X(\omega), Y(\omega)) \in Q\}$$

Since these are the same event, they must have the same probability. Once Q is identified on the plane, determine $P((X, Y) \in Q)$ in the usual manner (see part a, above).

- *Discrete alternative*. For each possible vector value (t_i, u_j) of (X, Y), determine whether $g(t_i, u_j)$ meets the defining condition for M. Select those (t_i, u_j) which do and add the associated probabilities.

We illustrate the mapping approach in the absolutely continuous case. A key element in the approach is finding the set Q on the plane such that $g(X, Y) \in M$ iff $(X, Y) \in Q$. The desired probability is obtained by integrating f_{XY} over Q.

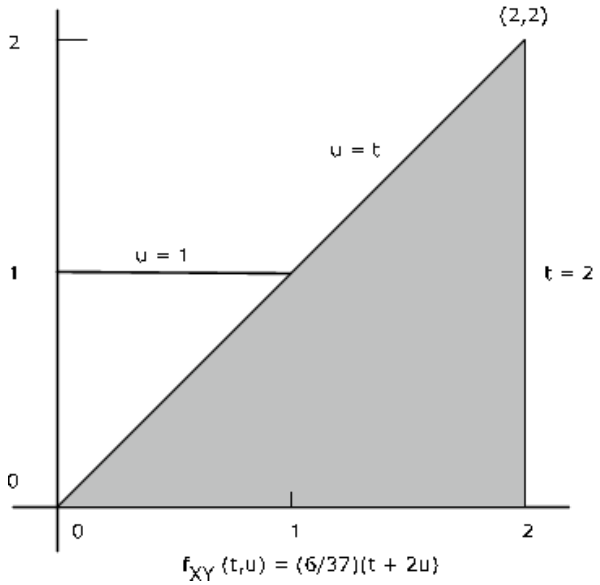


Figure 10.2.1. Distribution for Example 10.2.15.

Example 10.2.15. A numerical example

The pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = \dfrac{6}{37}(t + 2u)$ on the region bounded by $t = 0$, $t = 2$, $u = 0$, $u = \max\{1, t\}$ (see Figure 1). Determine $P(Y \leq X) = P(X - Y \geq 0)$. Here $g(t,u) = t - u$ and $M = [0, \infty)$. Now $Q = \{(t,u) : t - u \geq 0\} = \{(t,u) : u \leq t\}$ which is the region on the plane on or below the line $u = t$. Examination of the figure shows that for this region, f_{XY} is different from zero on the triangle bounded by $t = 2$, $u = 0$, and $u = t$. The desired probability is

$$P(Y \leq X) = \int_0^2 \int_0^t \frac{6}{37}(t + 2u)du\ dt = 32/37 \approx 0.8649$$

Example 10.2.16 *X Y*. The density for the sum +

Suppose the pair $\{X, Y\}$ has joint density f_{XY}. Determine the density for

$$Z = X + Y$$

Solution

$$F_Z(v) = P(X + Y \leq v) = P((X,Y) \in Q_v) \text{ where } Q_v = \{(t,u) : t + u \leq v\} = \{(t,u) : u \leq v - t\}$$

For any fixed v, the region Q_v is the portion of the plane on or below the line $u = v - t$ (see Figure 10.2.2). Thus

$$F_Z(v) = \int\int_Q f_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{v-t} f_{XY}(t,u)du\ dt$$

Differentiating with the aid of the fundamental theorem of calculus, we get

$$f_Z(v) = \int_{\infty}^{\infty} f_{XY}(t, v - t)\ dt$$

This integral expresssion is known as a *convolution integral*.

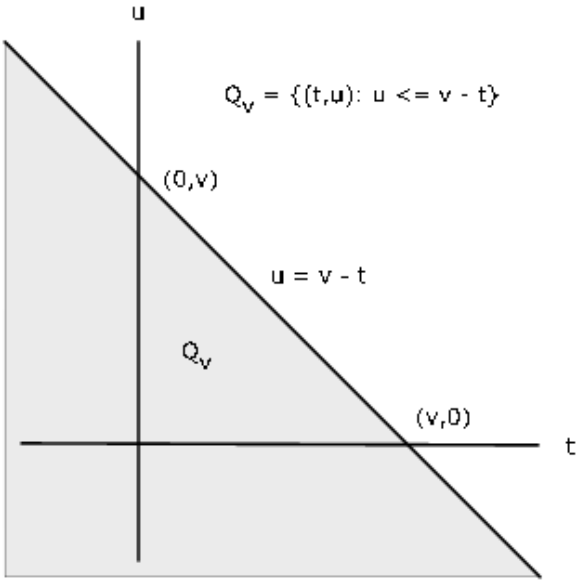


Figure 10.2.2. Region Q_v for $X + Y \leq v$.

Example 10.2.17. Sum of joint uniform random variables

Suppose the pair $\{X, Y\}$ has joint uniform density on the unit square $0 \leq t \leq 1, 0 \leq u \leq 1$.. Determine the density for $Z = X + Y$.

Solution

$F_Z(v)$ is the probability in the region $Q_v : u \leq v - t$. Now $P_{XY}(Q_v) = 1 - P_{XY}(Q_v^c)$, where the complementary set Q_v^c is the set of points above the line. As Figure 3 shows, for $v \leq 1$, the part of Q_v which has probability mass is the lower shaded triangular region on the figure, which has area (and hence probability) $v^2/2$. For $v > 1$, the complementary region Q_v^c is the upper shaded region. It has area $(2 - v)^2/2$. so that in this case, $P_{XY}(Q_v) = 1 - (2 - v)^2/2$. Thus,

$$F_Z(v) = \frac{v^2}{2} \text{ for } 0 \leq v \leq 1 \text{ and } F_Z(v) = 1 - \frac{(2 - v)^2}{2} \text{ for } 1 \leq v \leq 2$$

Differentiation shows that Z has the symmetric triangular distribution on [0, 2], since

$$f_Z(v) = v \text{ for } 0 \leq v \leq 1 \text{ and } f_Z(v) = (2 - v) \text{ for } 1 \leq v \leq 2$$

With the use of indicator functions, these may be combined into a single expression

$$f_Z(v) = I_{[0,1]}(v)v + I_{(1,2]}(2 - v)$$

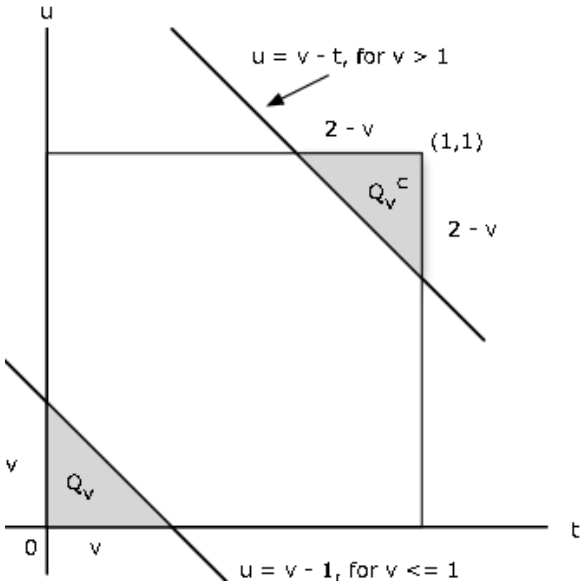


Figure 10.2.3. Geometry for sum of joint uniform random variables.

ALTERNATE Solution

Since $f_{XY}(t,u) = I_{[0,1]}(t)I_{[0,1]}(u)$, we have $f_{XY}(t,v-t) = I_{[0,1]}(t)I_{[0,1]}(v-t)$. Now $0 \le v-t \le 1$ iff $v-1 \le t \le v$, so that

$$f_{XY}(t,v-t) = I_{[0,1]}(v)I_{[0,v]}(t) + I_{(1,2]}(v)I_{[v-1,1]}(t)$$

Integration with respect to t gives the result above.

Independence of functions of independent random variables

Suppose $\{X,Y\}$ is an independent pair. Let $Z = g(X), W = h(Y)$. Since

$$Z^{-1}(M) = X^{-1}[g^{-1}(M)] \text{ and } W^{-1}(N) = Y^{-1}[h^{-1}(N)]$$

the pair $\{Z^{-1}(M), W^{-1}(N)\}$ is independent for each pair $\{M,N\}$. Thus, the pair $\{Z,W\}$ is independent.

If $\{X,Y\}$ is an independent pair and $Z = g(X)$, $W = g(X)$, then the pair $\{Z,W\}$ is independent. However, if $Z = g(X,Y)$ and $W = h(X,Y)$, then in general $\{Z,W\}$ is not independent. This is illustrated for simple random variables with the aid of the m-procedure *jointzw* at the end of the next section.

Example 10.2.18. Independence of simple approximations to an independent pair

Suppose $\{X,Y\}$ is an independent pair with simple approximations X_s and Y_s as described in Distribution Approximations.

$$X_s = \sum_{i=1}^{n} t_i I_{E_i} = \sum_{i=1}^{n} t_i I_{M_i}(X) \text{ and } Y_s = \sum_{j=1}^{m} u_j I_{F_j} = \sum_{j=1}^{m} u_j I_{N_j}(Y)$$

As functions of X and Y, respectively, the pair $\{X_s, Y_s\}$ is independent. Also each pair $\{I_{M_i}(X), I_{N_j}(Y)\}$ is independent.

Use of MATLAB on pairs of simple random variables

In the single-variable case, we use array operations on the values of X to determine a matrix of values of $g(X)$. In the two-variable case, we must use array operations on the calculating matrices t and u to obtain a matrix G whose elements are $g(t_i, u_j)$. To obtain the distribution for $Z = g(X,Y)$, we may use the m-function csort on G and the joint probability matrix P. A first step, then, is the use of jcalc or icalc to set up the joint distribution and the calculating matrices. This is illustrated in the following example.

Example 10.2.19

```
% file jdemo3.m
% data for joint simple distribution
X = [-4 -2 0 1 3];
Y = [0 1 2 4];
P = [0.0132    0.0198    0.0297    0.0209    0.0264;
     0.0372    0.0558    0.0837    0.0589    0.0744;
     0.0516    0.0774    0.1161    0.0817    0.1032;
     0.0180    0.0270    0.0405    0.0285    0.0360];
jdemo3                % Call for data
jcalc                 % Set up of calculating matrices t, u.
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = t.^2 -3*u;        % Formation of G = [g(ti,uj)]
M = G >= 1;           % Calculation using the XY distribution
PM = total(M.*P)      % Alternately, use total((G>=1).*P)
PM =  0.4665
[Z,PZ] = csort(G,P);
PM = (Z>=1)*PZ'     % Calculation using the Z distribution
PM =  0.4665
disp([Z;PZ]')         % Display of the Z distribution
  -12.0000    0.0297
  -11.0000    0.0209
   -8.0000    0.0198
   -6.0000    0.0837
   -5.0000    0.0589
   -3.0000    0.1425
   -2.0000    0.1375
         0    0.0405
    1.0000    0.1059
    3.0000    0.0744
    4.0000    0.0402
    6.0000    0.1032
    9.0000    0.0360
   10.0000    0.0372
   13.0000    0.0516
   16.0000    0.0180
```

We extend the example above by considering a function $W = h(X, Y)$ which has a composite definition.

Example 10.2.20 Continuation of example 10.2.19

Let

$$W = \begin{cases} X & \text{for } X + Y \geq 1 \\ X^2 + Y^2 & \text{for } X + Y < 1 \end{cases}$$ Determine the distribution for W

```
H = t.*(t+u>=1) + (t.^2 + u.^2).*(t+u<1);  % Specification of h(t,u)
                                           %
                                           %
[W,PW] = csort(H,P);                       % Distribution for W = h(X,Y)
```

```
disp([W;PW]')
   -2.0000    0.0198
         0    0.2700
    1.0000    0.1900
    3.0000    0.2400
    4.0000    0.0270
    5.0000    0.0774
    8.0000    0.0558
   16.0000    0.0180
   17.0000    0.0516
   20.0000    0.0372
   32.0000    0.0132
ddbn                                    % Plot of distribution function
Enter row matrix of values  W
Enter row matrix of probabilities  PW
print                                   % See Figure 10.2.4
```

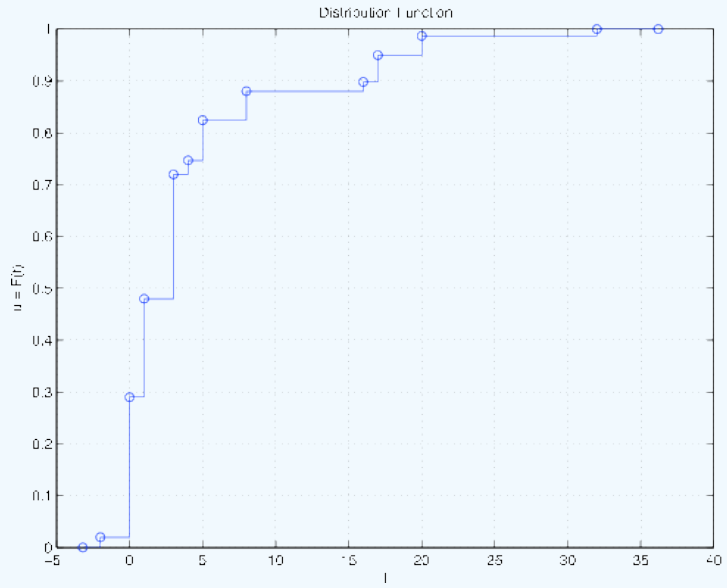


Figure 10.2.4. Distribution for random variable W in Example 10.2.20.

Joint distributions for two functions of (X, Y)

In previous treatments, we use csort to obtain the *marginal* distribution for a single function $Z = g(X, Y)$. It is often desirable to have the *joint* distribution for a pair $Z = g(X, Y)$ and $W = h(X, Y)$. As special cases, we may have $Z = X$ or $W = Y$. Suppose

$$Z \text{ has values } [z_1\ z_2 \cdots z_c] \text{ and } W \text{ has calues } [w_1\ w_2 \cdots w_c]$$

The joint distribution requires the probability of each pair, $P(W = w_i, Z = z_j)$. Each such pair of values corresponds to a set of pairs of X and Y values. To determine the joint probability matrix PZW for (Z, W) arranged as on the plane, we assign to each position (i, j) the probability $P(W = w_i, Z = z_j)$, with values of W increasing upward. Each pair of (W, Z) values corresponds to one or more pairs of (Y, X) values. If we select and add the probabilities corresponding to the latter pairs, we have $P(W = w_i, Z = z_j)$. This may be accomplished as follows:

Set up calculation matrices t and u as with jcalc.
Use array arithmetic to determine the matrices of values $G = [g(t, u)]$ and $H = [h(t, u)]$.
Use csort to determine the Z and W value matrices and the PZ and PW marginal probability matrices.
For each pair (w_i, z_j), use the MATLAB function *find* to determine the positions *a* for which

(H==W(i))&(G==Z(j))

Assign to the (i, j) position in the joint probability matrix PZW for (Z, W) the probability

PZW(i, j) = total (P(a))

We first examine the basic calculations, which are then implemented in the m-procedure *jointzw*.

Example 10.2.21. Illustration of the basic joint calculations

```
% file jdemo7.m
P = [0.061  0.030  0.060  0.027  0.009;
       0.015  0.001  0.048  0.058  0.013;
       0.040  0.054  0.012  0.004  0.013;
       0.032  0.029  0.026  0.023  0.039;
       0.058  0.040  0.061  0.053  0.018;
       0.050  0.052  0.060  0.001  0.013];
X = -2:2;
Y = -2:3;
jdemo7                      % Call for data in jdemo7.m
jcalc                       % Used to set up calculation matrices t, u
- - - - - - - - - - -
H = u.^2                    % Matrix of values for W = h(X,Y)
H =
     9     9     9     9     9
     4     4     4     4     4
     1     1     1     1     1
     0     0     0     0     0
     1     1     1     1     1
     4     4     4     4     4
G = abs(t)                  % Matrix of values for Z = g(X,Y)

G =
     2     1     0     1     2
     2     1     0     1     2
     2     1     0     1     2
     2     1     0     1     2
     2     1     0     1     2
     2     1     0     1     2
[W,PW] = csort(H,P)         % Determination of marginal for W
W =     0     1     4     9
PW =    0.1490    0.3530    0.3110    0.1870
[Z,PZ] = csort(G,P)         % Determination of marginal for Z
Z =     0     1     2
PZ =    0.2670    0.3720    0.3610
r = W(3)                    % Third value for W
r =   4
s = Z(2)                    % Second value for Z
s =   1
```

To determine $P(W = 4, Z = 1)$, we need to determine the (t, u) positions for which this pair of (W, Z) values is taken on. By inspection, we find these to be (2,2), (6,2), (2,4), and (6,4). Then $P(W = 4, Z = 1)$ is the total probability at these positions. This is 0.001 + 0.052 + 0.058 + 0.001 = 0.112. We put this probability in the joint probability matrix PZW at the $W = 4, Z = 1$ position. This may be achieved by MATLAB with the following operations.

```
[i,j] = find((H==W(3))&(G==Z(2)));  % Location of (t,u) positions
disp([i j])                         % Optional display of positions
     2     2
```

```
     6     2
     2     4
     6     4
a = find((H==W(3))&(G==Z(2)));      % Location in more convenient form
P0 = zeros(size(P));                % Setup of zero matrix
P0(a) = P(a)                        % Display of designated probabilities in P
P0 =
         0         0         0         0         0
         0    0.0010         0    0.0580         0
         0         0         0         0         0
         0         0         0         0         0
         0         0         0         0         0
         0    0.0520         0    0.0010         0
PZW = zeros(length(W),length(Z))    % Initialization of PZW matrix
PZW(3,2) = total(P(a))              % Assignment to PZW matrix with
PZW =    0         0         0      % W increasing downward
         0         0         0
         0    0.1120         0
         0         0         0
```

```
PZW = flipud(PZW)                   % Assignment with W increasing upward
PZW =
         0         0         0
         0    0.1120         0
         0         0         0
         0         0         0
```

The procedure *jointzw* carries out this operation for each possible pair of W and Z values (with the `flipud` operation coming only after all individual assignments are made).

example 10.2.22. joint distribution for z = g(x,y) = ||x| - y| and w = h(x, y) = |xy|

```
% file jdemo3.m   data for joint simple distribution
X = [-4 -2 0 1 3];
Y = [0 1 2 4];
P = [0.0132    0.0198    0.0297    0.0209    0.0264;
     0.0372    0.0558    0.0837    0.0589    0.0744;
     0.0516    0.0774    0.1161    0.0817    0.1032;
     0.0180    0.0270    0.0405    0.0285    0.0360];
jdemo3          % Call for data
jointzw         % Call for m-program
Enter joint prob for (X,Y): P
Enter values for X: X
Enter values for Y: Y
Enter expression for g(t,u): abs(abs(t)-u)
Enter expression for h(t,u): abs(t.*u)
Use array operations on Z, W, PZ, PW, v, w, PZW
disp(PZW)
    0.0132         0         0         0         0
         0    0.0264         0         0         0
         0         0    0.0570         0         0
         0    0.0744         0         0         0
    0.0558         0         0    0.0725         0
         0         0    0.1032         0         0
         0    0.1363         0         0         0
```

```
    0.0817         0         0         0         0
    0.0405    0.1446    0.1107    0.0360    0.0477
EZ = total(v.*PZW)
EZ =   1.4398

ez = Z*PZ'       % Alternate, using marginal dbn
ez =   1.4398
EW = total(w.*PZW)
EW =   2.6075
ew = W*PW'       % Alternate, using marginal dbn
ew =   2.6075
M = v > w;           % P(Z>W)
PM = total(M.*PZW)
PM =   0.3390
```

At noted in the previous section, if $\{X, Y\}$ is an independent pair and $Z = g(X)$,

$W = h(Y)$, then the pair $\{Z, W\}$ is independent. However, if $Z = g(X, Y)$ and

$W = h(X, Y)$, then in general the pair $\{Z, W\}$ is not independent. We may illustrate this with the aid of the m-procedure *jointzw*

Example 10.2.23 Functions of independent random variables

```
jdemo3
itest
Enter matrix of joint probabilities  P
The pair {X,Y} is independent          % The pair {X,Y} is independent
jointzw
Enter joint prob for (X,Y): P
Enter values for X: X
Enter values for Y: Y
Enter expression for g(t,u): t.^2 - 3*t  % Z = g(X)
Enter expression for h(t,u): abs(u) + 3  % W = h(Y)
Use array operations on Z, W, PZ, PW, v, w, PZW
itest
Enter matrix of joint probabilities  PZW
The pair {X,Y} is independent          % The pair {g(X),h(Y)} is independent
jdemo3                                 % Refresh data
jointzw
Enter joint prob for (X,Y): P
Enter values for X: X
Enter values for Y: Y
Enter expression for g(t,u): t+u         % Z = g(X,Y)
Enter expression for h(t,u): t.*u        % W = h(X,Y)
Use array operations on Z, W, PZ, PW, v, w, PZW
```

```
itest
Enter matrix of joint probabilities  PZW
The pair {X,Y} is NOT independent  % The pair {g(X,Y),h(X,Y)} is not indep
To see where the product rule fails, call for D  % Fails for all pairs
```

Absolutely continuous case: analysis and approximation

As in the analysis Joint Distributions, we may set up a simple approximation to the joint distribution and proceed as for simple

random variables. In this section, we solve several examples analytically, then obtain simple approximations.

Example 10.2.24 . Distribution for a product

Suppose the pair $\{X, Y\}$ has joint density f_{XY}. Let $Z = XY$. Determine Q_v such that $P(Z \leq v) = P((X, Y) \in Q_v)$.

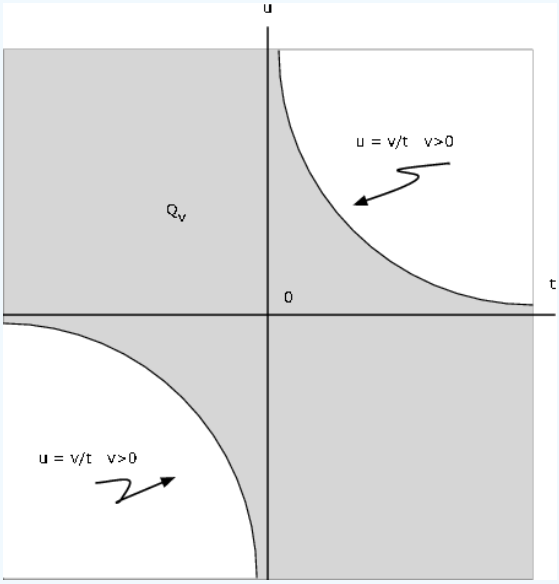


Figure 10.2.5

Solution

$Q_v = \{(t, u) : tu \leq v\} = \{(t, u) : t > 0, u \leq v/t\} \bigvee \{(t, u) : t < 0, u \geq v/t\}\}$

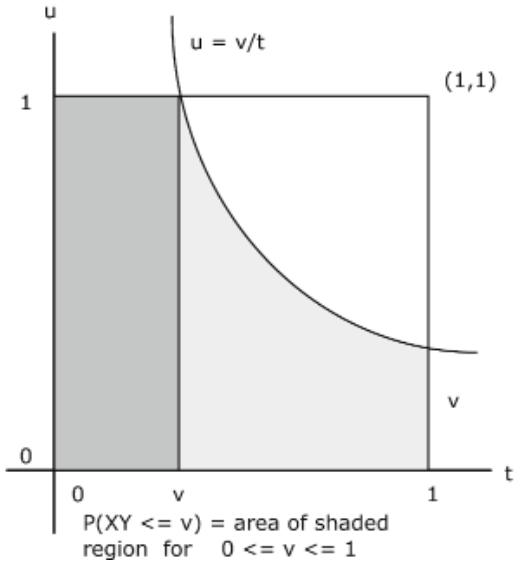


Figure 10.2.6. Product of X, Y with uniform joint distribution on the unit square.

Example 10.2.25 .

$\{X, Y\}$ ~ uniform on unit square

$f_{XY}(t, u) = 1$. Then (see Figure 10.2.6)

$$P(XY \le v) = \int\int_{Q_v} 1du\ dt \text{ where } Q_v = \{(t,u): 0 \le t \le 1, 0 \le u \le \min\{1, v/t\}\}$$

Integration shows

$F_Z(v) = P(XY \le v) = v(1 - \ln(v))$ so that $f_Z(v) = -\ln(v) = \ln(1/v)$, $0 < v \le 1$

For $v = 0.5$, $F_Z(0.5) = 0.8466$.

```
% Note that although f = 1, it must be expressed in terms of t, u.
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (u>=0)&(t>=0)
Use array operations on X, Y, PX, PY, t, u, and P
G = t.*u;
```

```
[Z,PZ] = csort(G,P);
p = (Z<=0.5)*PZ'
p =  0.8465                 % Theoretical value 0.8466, above
```

Example 10.2.26 . Continuation of example 5 from "Random Vectors and Joint Distributions"

The pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = \frac{6}{37}(t + 2u)$ on the region bounded by $t = 0$, $t = 2$ and $u = \max\{1, t\}$(see Figure 7). Let $Z = XY$. Determine $P(Z \le 1)$.

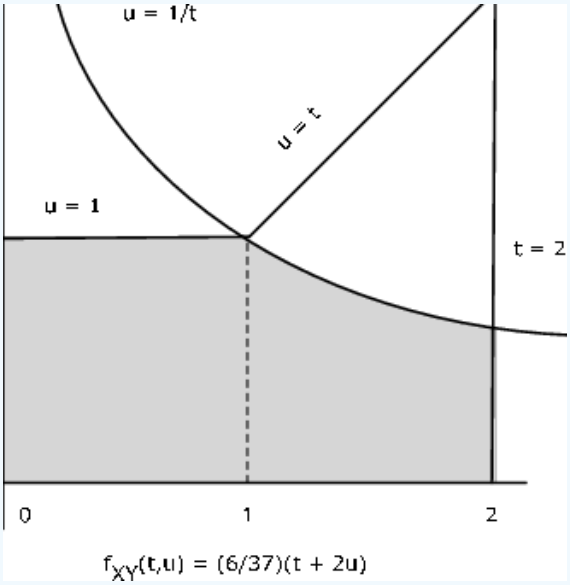


Figure 10.2.7. Area of integration for Example 10.2.26 .

Analytic Solution

$P(Z \le 1) = P((X,Y) \in Q)$ where $Q = \{(t,u): u \le 1/t\}$

Reference to Figure 10.2.7 shows that

$$P((X,Y) \in Q = \frac{6}{37}\int_0^1\int_0^1 (t + 2u)du\ dt + \frac{6}{37}\int_1^2\int_0^{1/t}(t + 2u)du\ dt = 9/37 + 9/37 = 18/37 \approx 0.4865$$

APPROXIMATE Solution

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  300
Enter number of Y approximation points  300
Enter expression for joint density  (6/37)*(t + 2*u).*(u<=max(t,1))
Use array operations on X, Y, PX, PY, t, u, and P
Q = t.*u<=1;
PQ = total(Q.*P)
PQ =  0.4853            % Theoretical value 0.4865, above
G = t.*u;               % Alternate, using the distribution for Z
[Z,PZ] = csort(G,P);
PZ1 = (Z<=1)*PZ'
PZ1 = 0.4853
```

In the following example, the function g has a compound definition. That is, it has a different rule for different parts of the plane.

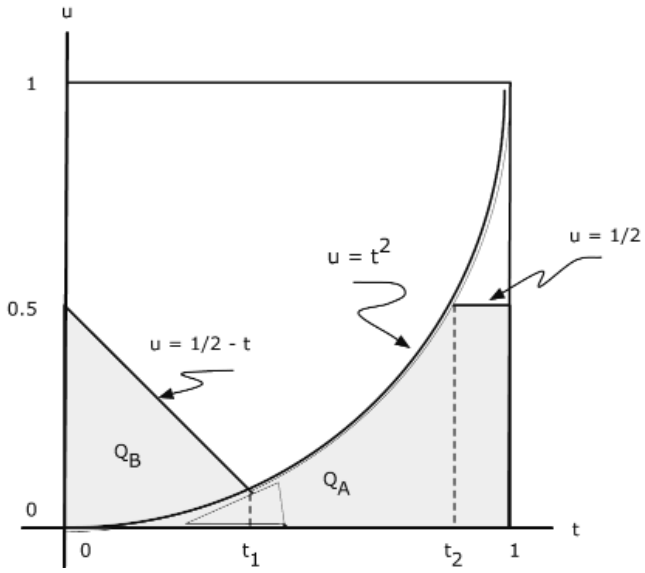


Figure 10.2.8. Regions for $P(Z \leq 1/2)$ in Example 10.2.27.

Example 10.2.27. A compound function

The pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = \dfrac{2}{3}(t+2u)$ on the unit square $0 \leq t \leq 1$, $0 \leq u \leq 1$.

$$Z = \begin{cases} X & \text{for } X^2 - Y \geq 0 \\ X+Y & \text{for } X^2 - Y < 0 \end{cases} = I_Q(X,Y)Y + I_{Q^c}(X,Y)(X+Y)$$

for $Q = \{(t,u) : u \leq t^2\}$. Determine $P(Z <= 0.5)$.

Analytical Solution

$$P(Z \leq 1/2) = P(Y \leq 1/2, Y \leq X^2) + P(X+Y \leq 1/2, Y > X^2) = P((X,Y) \in Q_A \bigvee Q_B)$$

where $Q_A = \{(t,u) : u \leq 1/2, u \leq t^2\}$ and $Q_B = \{(t,u) : t+u \leq 1/2, u > t^2\}$. Reference to Figure 10.2.8 shows that this is the part of the unit square for which $u \leq \min\,(\max\,(1/2-t, t^2), 1/2)$. We may break up the integral into three parts. Let $1/2 - t_1 = t_1^2$ and $t_2^2 = 1/2$. Then

$$P(Z \le 1/2) = \frac{2}{3}\int_0^{t_1}\int_0^{1/2-t}(t+2u)du\ dt + \frac{2}{3}\int_{t_1}^{t_2}\int_0^{t^2}(t+2u)du\ dt + \frac{2}{3}\int_{t_2}^{1}\int_0^{1/2}(t+2u)du\ dt = 0.2322$$

APPROXIMATE Solution

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (2/3)*(t + 2*u)
Use array operations on X, Y, PX, PY, t, u, and P
Q = u <= t.^2;
G = u.*Q + (t + u).*(1-Q);
prob = total((G<=1/2).*P)
prob =  0.2328          % Theoretical is 0.2322, above
```

The setup of the integrals involves careful attention to the geometry of the system. Once set up, the evaluation is elementary but tedious. On the other hand, the approximation proceeds in a straightforward manner from the normal description of the problem. The numerical result compares quite closely with the theoretical value and accuracy could be improved by taking more subdivision points.

This page titled 10.2: Function of Random Vectors is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

10.3: The Quantile Function

The Quantile Function

The quantile function for a probability distribution has many uses in both the theory and application of probability. If F is a probability distribution function, the quantile function may be used to "construct" a random variable having F as its distributions function. This fact serves as the basis of a method of simulating the "sampling" from an arbitrary distribution with the aid of a *random number generator*. Also, given any finite class

$\{X_i : 1 \leq i \leq n\}$ of random variables, an *independent class* $\{Y_i : 1 \leq i \leq n\}$ may be constructed, with each X_i and associated Y_i having the same (marginal) distribution. Quantile functions for simple random variables may be used to obtain an important Poisson approximation theorem (which we do not develop in this work). The quantile function is used to derive a number of useful special forms for mathematical expectation.

General concept—properties, and examples

If F is a probability distribution function, the associated quantile function Q is essentially an inverse of F. The quantile function is defined on the unit interval (0, 1). For F continuous and strictly increasing at t, then $Q(u) = t$ iff $F(t) = u$. Thus, if u is a probability value, $t = Q(u)$ is the value of t for which $P(X \leq t) = u$.

Example 10.3.28: The Weibull distribution (3, 2, 0)

$$u = F(t) = 1 - e^{-3t^2} \; t \geq 0 \Rightarrow t = Q(u) = \sqrt{-\ln (1-u)/3}$$

Example 10.3.29: The Normal Distribution

The m-function *norminv*, based on the MATLAB function *erfinv* (inverse error function), calculates values of Q for the normal distribution.

The restriction to the continuous case is not essential. We consider a general definition which applies to any probability distribution function.

Definition: If F is a function having the properties of a probability distribution function, then the *quantile function* for F is given by

$$Q(u) = \inf \{t : F(t) \geq u\} \; \forall u \in (0, 1)$$

We note

- If $F(t^*) \geq u^*$, then $t^* \geq \inf \{t : F(t) \geq u^*\} = Q(u^*)$
- If $F(t^*) < u^*$, then $t^* < \inf \{t : F(t) \geq u^*\} = Q(u^*)$

Hence, we have the important property:

(Q1) $Q(u) \leq t$ iff $u \leq F(t) \; \forall u \in (0, 1)$

The property (Q1) implies the following important property:

(Q2) If U~ uniform (0, 1), then $X = Q(U)$ has distribution function $F_X = F$. To see this, note that $F_X(t) = P(Q(U) \leq t] = P[U \leq F(t)]$ $= F(t)$.

Property (Q2) implies that if F is any distribution function, with quantile function Q, then the random variable $X = Q(U)$, with U uniformly distributed on (0, 1), has distribution function F.

Example 10.3.30: Independent classes with prescribed distributions

Suppose $\{X_i : 1 \le i \le n\}$ is an arbitrary class of random variables with corresponding distribution functions $\{F_i : 1 \le i \le n\}$. Let $\{Q_i : 1 \le i \le n\}$ be the respective quantile functions. There is always an independent class $\{U_i : 1 \le i \le n\}$ iid uniform (0, 1) (marginals for the joint uniform distribution on the unit hypercube with sides (0, 1)). Then the random variables $Y_i = Q_i(U_i)$, $1 \le i \le n$, form an independent class with the same marginals as the X_i.

Several other important properties of the quantile function may be established.

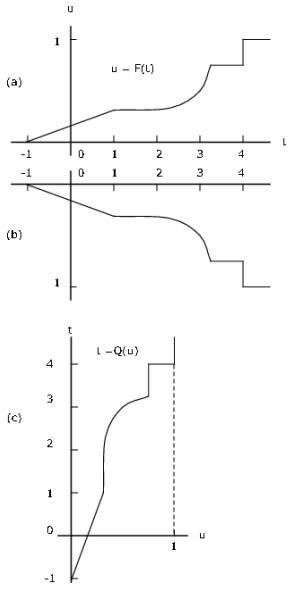


Figure 10.3.9. Graph of quantile function from graph of distribution function,

Q is left-continuous, whereas F is right-continuous.

If jumps are represented by vertical line segments, construction of the graph of $u = Q(t)$ may be obtained by the following two step procedure:

- Invert the entire figure (including axes), then
- Rotate the resulting figure 90 degrees counterclockwise

This is illustrated in Figure 10.3.9. If jumps are represented by vertical line segments, then jumps go into flat segments and flat segments go into vertical segments.

If X is discrete with probability p_i at t_i, $1 \le i \le n$, then F has jumps in the amount p_i at each t_i and is constant between. The quantile function is a left-continuous step function having value t_i on the interval $(b_{i-1}, b_i]$, where $b_0 = 0$ and $b_i = \sum_{j=1}^{i} p_j$. This may be stated

$$\text{If } F(t_i) = b_i \text{, then } Q(u) = t_i \text{ for } F(t_{i-1}) < u \le F(t_i)$$

Example 10.2.31: Quantile function for a simple random variable

Suppose simple random variable X has distribution

$$X = [-2\ 0\ 1\ 3] \backslash (PX = [0.2\ 0.1\ 0.3\ 0.4]$$

Figure 1 shows a plot of the distribution function F_X. It is reflected in the horizontal axis then rotated counterclockwise to give the graph of $Q(u$ versus u.

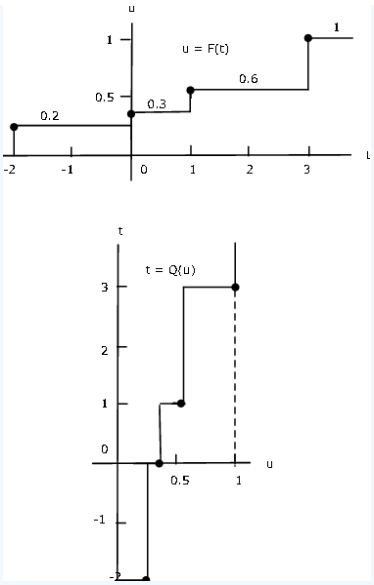


Figure 10.3.10. Distribution and quantile functions for Example 10.3.31.

We use the analytic characterization above in developing a number of m-functions and m-procedures.

m-procedures for a simple random variable

The basis for quantile function calculations for a simple random variable is the formula above. This is implemented in the m-function *dquant*, which is used as an element of several simulation procedures. To plot the quantile function, we use *dquanplot* which employs the stairs function and plots X vs the distribution function FX. The procedure *dsample* employs dquant to obtain a "sample" from a population with simple distribution and to calculate relative frequencies of the various values.

Example 10.3.32: Simple Random Variable

```
X =  [-2.3 -1.1 3.3 5.4 7.1 9.8];
PX = 0.01*[18 15 23 19 13 12];
dquanplot
Enter VALUES for X  X
Enter PROBABILITIES for X  PX     % See Figure 10.3.11 for plot of results
rand('seed',0)                 % Reset random number generator for reference
dsample
Enter row matrix of values  X
Enter row matrix of probabilities  PX
Sample size n  10000
```

```
    Value      Prob    Rel freq
   -2.3000    0.1800    0.1805
   -1.1000    0.1500    0.1466
    3.3000    0.2300    0.2320
    5.4000    0.1900    0.1875
    7.1000    0.1300    0.1333
    9.8000    0.1200    0.1201
Sample average ex = 3.325
Population mean E[X] = 3.305
Sample variance = 16.32
Population variance Var[X] = 16.33
```

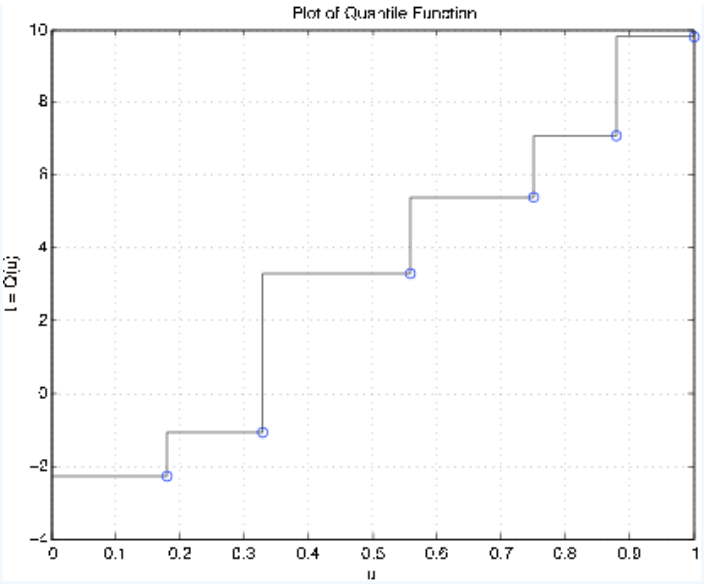


Figure 10.3.11. Quantile function for Example 10.3.32.

Sometimes it is desirable to know how many trials are required to reach a certain value, or one of a set of values. A pair of m-procedures are available for simulation of that problem. The first is called *targetset*. It calls for the population distribution and then for the designation of a "target set" of possible values. The second procedure, *targetrun*, calls for the number of repetitions of the experiment, and asks for the number of members of the target set to be reached. After the runs are made, various statistics on the runs are calculated and displayed.

Example10.3.33

```
X = [-1.3 0.2 3.7 5.5 7.3];     % Population values
PX = [0.2 0.1 0.3 0.3 0.1];     % Population probabilities
E = [-1.3 3.7];                 % Set of target states
targetset
Enter population VALUES  X
Enter population PROBABILITIES  PX
The set of population values is
   -1.3000    0.2000    3.7000    5.5000    7.3000
Enter the set of target values  E
Call for targetrun
```

```
rand('seed',0)                  % Seed set for possible comparison
targetrun
Enter the number of repetitions  1000
The target set is
   -1.3000    3.7000
Enter the number of target values to visit  2
The average completion time is 6.32
The standard deviation is 4.089
The minimum completion time is 2
The maximum completion time is 30
To view a detailed count, call for D.
The first column shows the various completion times;
the second column shows the numbers of trials yielding those times
% Figure 10.6.4 shows the fraction of runs requiring t steps or less
```

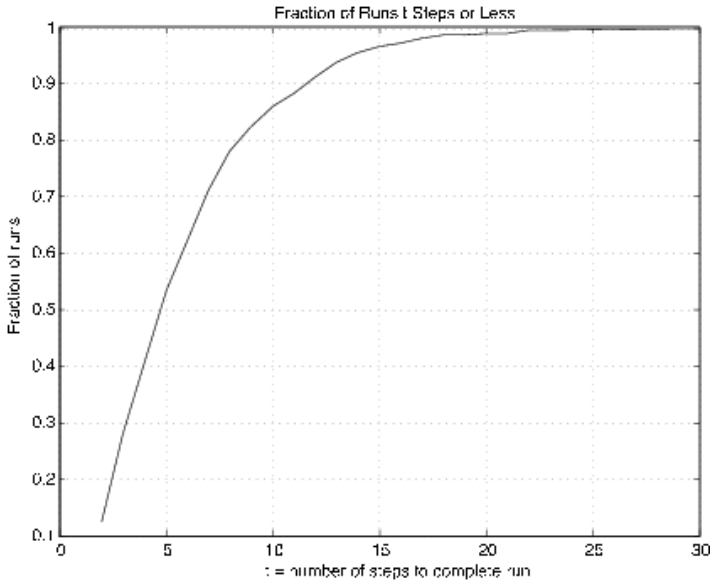


Figure 10.3.12. Fraction of runs requiring t steps or less.

m-procedures for distribution functions

A procedure *dfsetup* utilizes the distribution function to set up an approximate simple distribution. The m-procedure *quanplot* is used to plot the quantile function. This procedure is essentially the same as dquanplot, except the ordinary *plot* function is used in the continuous case whereas the plotting function *stairs* is used in the discrete case. The m-procedure *qsample* is used to obtain a sample from the population. Since there are so many possible values, these are not displayed as in the discrete case.

Example 10.3.34: Quantile function associated with a distribution function

```
F = '0.4*(t + 1).*(t < 0) + (0.6 + 0.4*t).*(t >= 0)';  % String
dfsetup
Distribution function F is entered as a string
variable, either defined previously or upon call
Enter matrix [a b] of X-range endpoints  [-1 1]
Enter number of X approximation points  1000
Enter distribution function F as function of t  F
Distribution is in row matrices X and PX
quanplot
Enter row matrix of values  X
Enter row matrix of probabilities  PX
Probability increment h  0.01          % See Figure 10.3.13 for plot
qsample
Enter row matrix of X values  X
Enter row matrix of X probabilities  PX
Sample size n  1000
Sample average ex = -0.004146
Approximate population mean E(X) = -0.0004002     % Theoretical = 0
Sample variance vx = 0.25
Approximate population variance V(X) = 0.2664
```

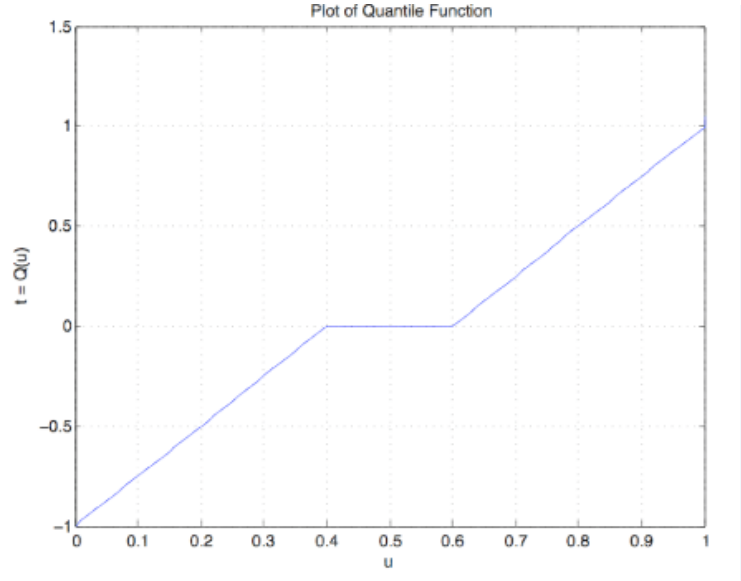


Figure 10.3.13. Quantile function for Example 10.3.34.

m-procedures for density functions

An m- procedure *acsetup* is used to obtain the simple approximate distribution. This is essentially the same as the procedure tuappr, except that the density function is entered as a string variable. Then the procedures quanplot and qsample are used as in the case of distribution functions.

Example 10.3.35: Quantile function associated with a density function

```
acsetup
Density f is entered as a string variable.
either defined previously or upon call.
Enter matrix [a b] of x-range endpoints  [0 3]
Enter number of x approximation points  1000
Enter density as a function of t  '(t.^2).*(t<1) + (1- t/3).*(1<=t)'
Distribution is in row matrices X and PX
quanplot
Enter row matrix of values  X
Enter row matrix of probabilities  PX
Probability increment h  0.01              % See Figure 10.3.14 for plot
rand('seed',0)
qsample
Enter row matrix of values  X
Enter row matrix of probabilities  PX
Sample size n  1000
Sample average ex = 1.352
Approximate population mean E(X) = 1.361  % Theoretical = 49/36 = 1.3622
Sample variance vx = 0.3242
Approximate population variance V(X) = 0.3474    % Theoretical = 0.3474
```

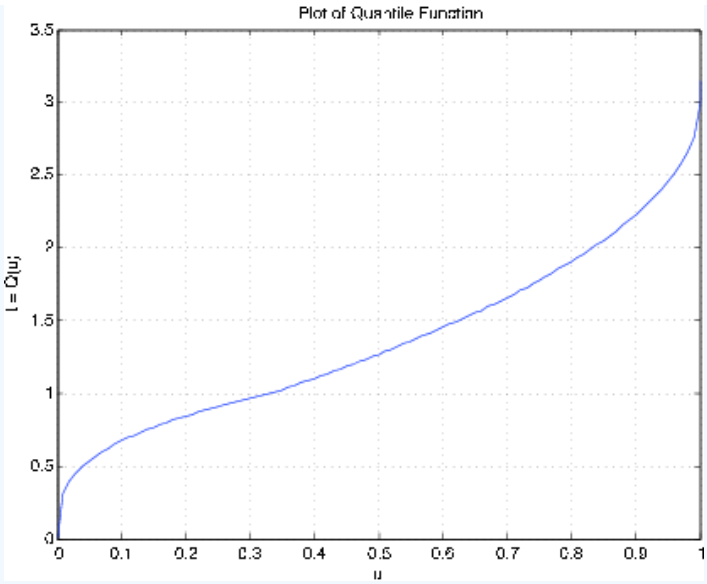


Figure 10.3.14. Quantile function for Example 10.3.35.

This page titled 10.3: The Quantile Function is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

10.4: Problems on Functions of Random Variables

Exercise 10.4.1

Suppose X is a nonnegative, absolutely continuous random variable. Let $Z = g(X) = Ce^{-aX}$, where $a > 0$, $C > 0$. Then $0 < Z \leq C$. Use properties of the exponential and natural log function to show that

$$F_Z(v) = 1 - F_X(-\frac{\ln (v/C)}{a}) \text{ for } 0 < v \leq C$$

Answer available at the end of this section.

Exercise 10.4.2

Use the result of Exercise 10.4.1 to show that if X ~ exponential (λ), then

$$F_Z(v) = (\frac{v}{C})^{\lambda/a} \; 0 < v \leq C$$

Answer available at the end of this section.

Exercise 10.4.3

Present value of future costs. Suppose money may be invested at an annual rate *a*, compounded continually. Then one dollar in hand now, has a value e^{ax} at the end of x years. Hence, one dollar spent x years in the future has a *present value* e^{-ax}. Suppose a device put into operation has time to failure (in years) X ~ exponential (λ). If the cost of replacement at failure is C dollars, then the present value of the replacement is $Z = Ce^{-aX}$. Suppose $\lambda = 1/10$, $a = 0.07$, and $C =$ \$1000.

a. Use the result of Exercise 10.4.2. to determine the probability $Z \leq 700, 500, 200$.
b. Use a discrete approximation for the exponential density to approximate the probabilities in part (a). Truncate X at 1000 and use 10,000 approximation points.

Answer

$$P(Z \leq v) = (\frac{v}{1000})^{10/7}$$

```
v = [700 500 200];
P = (v/1000).^(10/7)
P =  0.6008    0.3715    0.1003
tappr
Enter matrix [a b] of x-range endpoints  [0 1000]
Enter number of x approximation points  10000
Enter density as a function of t  0.1*exp(-t/10)
Use row matrices X and PX as in the simple case
G = 1000*exp(-0.07*t);
PM1 = (G<=700)*PX'
PM1 =  0.6005
PM2 = (G<=500)*PX'
PM2 =  0.3716
PM3 = (G<=200)*PX'
PM3 =  0.1003
```

Exercise 10.4.4

Optimal stocking of merchandise. A merchant is planning for the Christmas season. He intends to stock *m* units of a certain item at a cost of *c* per unit. Experience indicates demand can be represented by a random variable D ~ Poisson (μ). If units remain in stock at the end of the season, they may be returned with recovery of r per unit. If demand exceeds the number originally ordered, extra units may be ordered at a cost of *s* each. Units are sold at a price p per unit. If $Z = g(D)$ is the gain from the sales, then

- For $t \leq m$, $g(t) = (p - c)t - (c - r)(m - t) = (p - r)t + (r - c)m$
- For $t > m$, $g(t) = (p - c)m + (t - m)(p - s) = (p - s)t + (s - c)m$

Let $M = (-\infty, m]$. Then

$g(t) = I_M(t)[(p - r)t + (r - c)m] + I_M(t)[(p - s)t + (s - c)m]$

Suppose $\mu = 50$ $m = 50$ $c = 30$ $p = 50$ $r = 20$ $s = 40$.

Approximate the Poisson random variable D by truncating at 100. Determine $P(500 \leq Z \leq 1100)$.

Answer available at the end of this section.

Exercise 10.4.5

(See Example 2 from "Functions of a Random Variable") The cultural committee of a student organization has arranged a special deal for tickets to a concert. The agreement is that the organization will purchase ten tickets at $20 each (regardless of the number of individual buyers). Additional tickets are available according to the following schedule:

- 11-20, $18 each
- 21-30, $16 each
- 31-50, $15 each
- 51-100, $13 each

If the number of purchasers is a random variable X, the total cost (in dollars) is a random quantity $Z = g(X)$ described by

$$g(X) = 200 + 18I_{M1}(X)(X - 10) + (16 - 18)I_{M2}(X)(X - 20) +$$

$$(15 - 16)I_{M_3}(X)(X - 30) + (13 - 15)I_{M4}(X)(X - 50)$$

where $M1 = [10, \infty)$, $M2 = [20, \infty)$, $M3 = [30, \infty)$, $M4 = [50, \infty)$

Suppose X~ Poisson (75). Approximate the Poisson distribution by truncating at 150. Determine $P(Z \geq 1000)$, $P(Z \geq 1300)$ and $P(900 \leq Z \leq 1400)$.

Answer available at the end of this section.

Exercise 10.4.6

(See Exercise 6 from "Problems on Random Vectors and Joint Distributions", and Exercise 1 from "Problems on Independent Classes of Random Variables")) The pair $\{X, Y\}$ has the joint distribution

(in m-file npr08_06.m):

X = [-2.3 -0.7 1.1 3.9 5.1] Y = [1.3 2.5 4.1 5.3]

$$P = \begin{bmatrix} 0.0483 & 0.0357 & 0.0420 & 0.0399 & 0.0441 \\ 0.0437 & 0.0323 & 0.0380 & 0.0361 & 0.0399 \\ 0.0713 & 0.0527 & 0.0620 & 0.0609 & 0.0551 \\ 0.0667 & 0.0493 & 0.0580 & 0.0651 & 0.0589 \end{bmatrix}$$

Determine $P(\max\{X, Y\} \leq 4)$. Let $Z = 3X^3 + 3X^2Y - Y^3$.

Determine $P(Z < 0)$ and $P(-5 < Z \leq 300)$.

Answer available at the end of this section.

Exercise 10.4.7

(See Exercise 2 from "Problems on Independent Classes of Random Variables") The pair $\{X, Y\}$ has the joint distribution (in m-file npr09_02.m):

$$X = [-3.9\ -1.7\ 1.5\ 2\ 8\ 4.1]\ Y = [-2\ 1\ 2.6\ 5.1]$$

$$P = \begin{bmatrix} 0.0589 & 0.0342 & 0.0304 & 0.0456 & 0.0209 \\ 0.0962 & 0.056 & 0.0498 & 0.0744 & 0.0341 \\ 0.0682 & 0.0398 & 0.0350 & 0.0528 & 0.0242 \\ 0.0868 & 0.0504 & 0.0448 & 0.0672 & 0.0308 \end{bmatrix}$$

Determine $P(\{X + Y \geq 5\} \cup \{Y \leq 2\})$, $P(X^2 + Y^2 \leq 10)$.

Answer available at the end of this section.

Exercise 10.4.8

(See Exercsie 7 from "Problems on Random Vectors and Joint Distributions", and Exercise 3 from "Problems on Independent Classes of Random Variables") The pair has the joint distribution

(in m-file npr08_07.m):

$$P(X = t, Y = u)$$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189
-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

Determine $P(X^2 - 3X \leq 0)$, $P(X^3 - 3|Y| < 3)$.

Answer available at the end of this section.

Exercise 10.4.9

For the pair $\{X, Y\}$ in Exercise 10.4.8, let $Z = g(X, Y) = 3X^2 + 2XY - Y^2$. Determine and plot the distribution function for Z.

Answer available at the end of this section.

Exercise 10.4.10

For the pair $\{X, Y\}$ in Exercise 8, let

$$W = g(X,Y) = \begin{cases} X & \text{for } X+Y \le 4 \\ 2Y & \text{for } X+Y > 4 \end{cases} = I_M(X,Y)X + I_{M^c}(X,Y)2Y$$

Determine and plot the distribution function for W.

Answer available at the end of this section.

For the distributions in Exercises 10-15 below

a. Determine analytically the indicated probabilities.
b. Use a discrete approximation to calculate the same probablities.'

Exercise 10.4.11

$f_{XY}(t,u) = \frac{3}{88}(2t + 3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1 + t$ (see Exercise 15 from "Problems on Random Vectors and Joint Distributions").

$$Z = I_{[0,1]}(X)4X + I_{(1,2]}(X)(X+Y)$$

Determine $P(Z \le 2)$

Answer available at the end of this section.

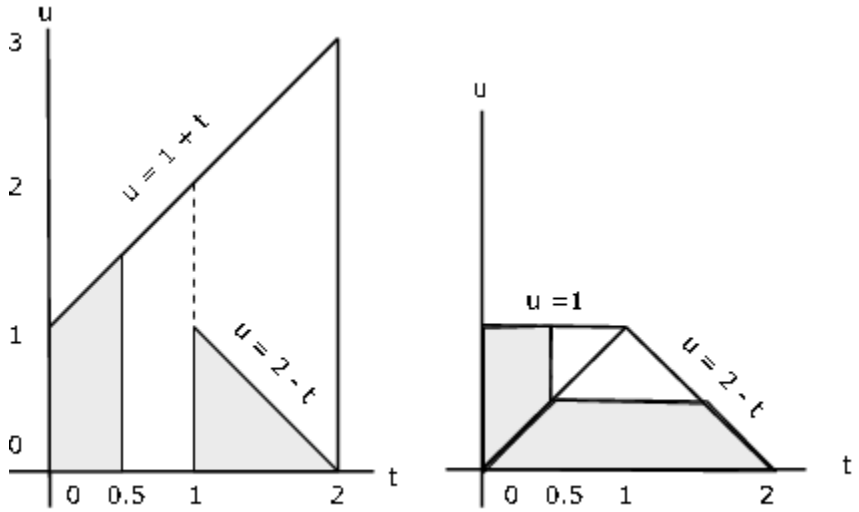


Figure 10.4.1

Exercise 10.4.12

$f_{XY}(t,u) = \frac{24}{11}$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2-t\}$(see Exercise 17 from "Problems on Random Vectors and Joint Distributions").

$$Z = I_M(X,Y)\frac{1}{2}X + I_{M^c}(X,Y)Y^2,\ M = \{(t,u) : u > t\}$$

Determine $P(Z \leq 1/4)$.

Answer available at the end of this section.

Exercise 10.4.13

$f_{XY}(t,u) = \frac{3}{23}(t+2u)$ for $0 \leq t \leq 2$, $0 \leq u \leq \max\{2-t, t\}$ (see Exercise 18 from "Problems on Random Vectors and Joint Distributions").

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y,\ M = \{(t,u) : \max(t,u) \leq 1\}$$

Determine $P(Z \leq 1)$

Answer available at the end of this section.

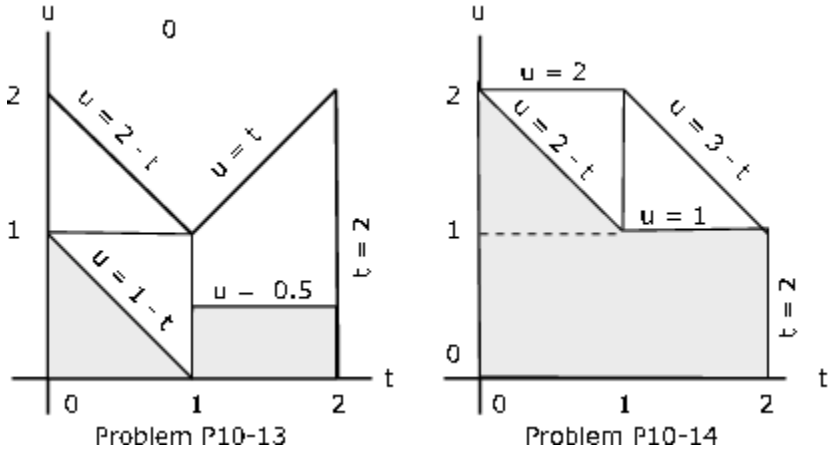


Figure 10.4.2

Exercise 10.4.14

$f_{XY}(t,u) = \frac{12}{179}(3t^2+u)$, for $0 \leq t \leq 2$, $0 \leq u \leq \min\{2, 3-t\}$ (see Exercise 19 from "Problems on Random Vectors and Joint Distributions").

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y^2,\ M = \{(t,u) : t \leq 1, u \geq 1\}$$

Determine $P(Z \leq 2)$.

Answer available at the end of this section.

Exercise 10.4.15

$f_{XY}(t,u) = \frac{12}{227}(3t+2tu)$, for $0 \leq t \leq 2$, $0 \leq u \leq \min\{1+t, 2\}$ (see Exercise 20 from "Problems on Random Variables and joint Distributions")

$$Z = I_M(X,Y)X + I_{M^c}(X,Y)\frac{Y}{X},\ M = \{(t,u) : u \leq \min(1, 2-t)\}$$

Determine $P(Z \leq 1)$.

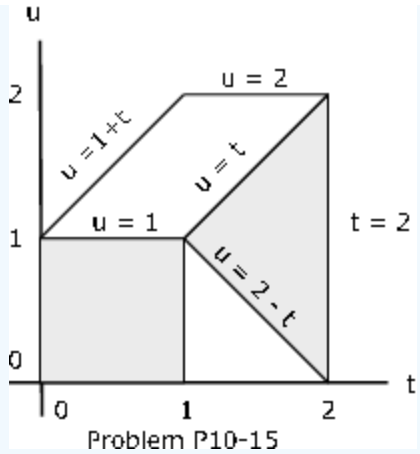


Figure 10.4.3

Answer

$$P(Z \le 1) = P((X,Y) \in M_1Q_1 \bigvee V_2Q_2),\ M_1 = M,\ M_2 = M^c$$

$$Q_1 = \{(t,u) : 0 \le t \le\}\ Q_2 = \{(t,u) : u \le t\} \text{ (see figure)}$$

$$P = \frac{12}{227}\int_0^1\int_0^1 (3t + 2tu)du\ dt + \frac{12}{227}\int_1^2\int_{2-t}^t (3t + 2tu)du\ dt = \frac{124}{227}$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  400
Enter number of Y approximation points  400
Enter expression for joint density  (12/227)*(3*t+2*t.*u).*(u<=min(1+t,2))
Use array operations on X, Y, PX, PY, t, u, and P
Q = (u<=1).*(t<=1) + (t>1).*(u>=2-t).*(u<=t);
P = total(Q.*P)
P =  0.5478                          % Theoretical = 124/227 = 0.5463
```

Exercise 10.4.16

The class $\{X, Y, Z\}$ is independent.

$X = -2I_A + I_B + 3I_C$. Minterm probabilities are (in the usual order)

0.255 0.025 0.375 0.045 0.108 0.012 0.162 0.018

$Y = I_D + 3I_E + I_F - 3$. The class $\{D, E, F\}$ is independent with

$$P(D) = 0.32\ P(E) = 0.56\ P(F) = 0.40$$

Z has distribution

Value	-1.3	1.2	2.7	3.4	5.8
Probability	0.12	0.24	0.43	0.13	0.08

Determine $P(X^2 + 3XY^2 > 3Z)$.

Answer available at the end of this section.

Exercise 10.4.17

The simple random variable *X* has distribution

$$X = [-3.1\ -0.5\ 1.2\ 2.4\ 3.7\ 4.9]\ PX = [0.15\ 0.22\ 0.33\ 0.12\ 0.11\ 0.07]$$

a. Plot the distribution function F_X and the quantile function Q_X.
b. Take a random sample of size $n = 10{,}000$. Compare the relative frequency for each value with the probability that value is taken on.

Answer

```
X = [-3.1 -0.5 1.2 2.4 3.7 4.9];
PX = 0.01*[15 22 33 12 11  7];
ddbn
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX  % Plot not reproduced here
dquanplot
Enter VALUES for X  X
Enter PROBABILITIES for X  PX          % Plot not reproduced here
rand('seed',0)                        % Reset random number generator
dsample                               % for comparison purposes
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX
Sample size n  10000
    Value      Prob    Rel freq
   -3.1000    0.1500    0.1490
   -0.5000    0.2200    0.2164
    1.2000    0.3300    0.3340
    2.4000    0.1200    0.1184
    3.7000    0.1100    0.1070
    4.9000    0.0700    0.0752
Sample average ex = 0.8792
Population mean E[X] = 0.859
Sample variance vx = 5.146
Population variance Var[X] = 5.112
```

Answers

Exercise 10.4.1

$Z = Ce^{-aX} \le v$ iff $e^{-aX} \le v/C$ iff $-aX \le \ln(v/C)$ iff $X \ge -\ln(v/C)/a$, so that

$$F_Z(v) = P(Z \le v) = P(X \ge -\ln(v/C)/a) = 1 - F_X(-\frac{\ln(v/C)}{a})$$

Exercise 10.4.2

$$F_Z(v) = 1 - [1 - exp(-\frac{\lambda}{a} \cdot \ln(v/C))] = (\frac{v}{C})^{\lambda/a}$$

Exercise 10.4.4

```
mu = 50;
D = 0:100;
c = 30;
p = 50;
r = 20;
s = 40;
```

```
m = 50;
PD = ipoisson(mu,D);
G = (p - s)*D + (s - c)*m +(s - r)*(D - m).*(D <= m);
M = (500<=G)&(G<=1100);
PM = M*PD'
PM =  0.9209

[Z,PZ] = csort(G,PD);         % Alternate: use dbn for Z
m = (500<=Z)&(Z<=1100);
pm = m*PZ'
pm =  0.9209
```

Exercise 10.4.5

```
X = 0:150;
PX = ipoisson(75,X);
G = 200 + 18*(X - 10).*(X>=10) + (16 - 18)*(X - 20).*(X>=20) + ...
     (15 - 16)*(X- 30).*(X>=30) + (13 - 15)*(X - 50).*(X>=50);
P1 = (G>=1000)*PX'
P1 =  0.9288
P2 = (G>=1300)*PX'
P2 =  0.1142
P3 = ((900<=G)&(G<=1400))*PX'
P3 =  0.9742
[Z,PZ] = csort(G,PX);         % Alternate: use dbn for Z
p1 = (Z>=1000)*PZ'
p1 =  0.9288
```

Exercise 10.4.6

```
npr08_06
Data are in X, Y, P
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
P1 = total((max(t,u)<=4).*P)
P1 =  0.4860
P2 = total((abs(t-u)>3).*P)
P2 =  0.4516
G = 3*t.^3 + 3*t.^2.*u - u.^3;
P3 = total((G<0).*P)
P3 =  0.5420
P4 = total(((-5<G)&(G<=300)).*P)
P4 =  0.3713
[Z,PZ] = csort(G,P);          % Alternate: use dbn for Z
p4 = ((-5<Z)&(Z<=300))*PZ'
p4 =  0.3713
```

Exercise 10.4.7

```
npr09_02
Data are in X, Y, P
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
```

```
M1 = (t+u>=5)|(u<=2);
P1 = total(M1.*P)
P1 =  0.7054
M2 = t.^2 + u.^2 <= 10;
P2 = total(M2.*P)
P2 =  0.3282
```

Exercise 10.4.8

```
npr08_07
Data are in X, Y, P
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
M1 = t.^2 - 3*t <=0;
P1 = total(M1.*P)
P1 =  0.4500
M2 = t.^3 - 3*abs(u) < 3;
P2 = total(M2.*P)
P2 =  0.7876
```

Exercise 10.4.9

```
G = 3*t.^2 + 2*t.*u - u.^2;  % Determine g(X,Y)
[Z,PZ] = csort(G,P);         % Obtain dbn for Z = g(X,Y)
ddbn                         % Call for plotting m-procedure
Enter row matrix of VALUES  Z
Enter row matrix of PROBABILITIES  PZ   % Plot not reproduced here
```

Exercise 10.4.10

```
H = t.*(t+u<=4) + 2*u.*(t+u>4);
[W,PW] = csort(H,P);
ddbn
Enter row matrix of VALUES  W
Enter row matrix of PROBABILITIES  PW   % Plot not reproduced here
```

Exercise 10.4.11

$$P(Z \le 2) = P(Z \in Q = Q1M1 \bigvee Q2M2), \text{ where } M1 = \{(t,u): 0 \le t \le 1, 0 \le u \le 1+t\}$$
$$M2 = \{(t,u): 1 < t \le 2, 0 \le u \le 1+t\}$$
$$Q1 = \{(t,u): 0 \le t \le 1/2\}, Q2 = \{(t,u): u \le 2-t\} \text{ (see figure)}$$
$$P = \frac{3}{88}\int_0^{1/2}\int_0^{1+t}(2t+3u^2)du\,dt + \frac{3}{88}\int_1^2\int_0^{2-t}(2t+3u^2)du\,dt = \frac{563}{5632}$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 3]
Enter number of X approximation points  200
Enter number of Y approximation points  300
Enter expression for joint density  (3/88)*(2*t + 3*u.^2).*(u<=1+t)
Use array operations on X, Y, PX, PY, t, u, and P
G = 4*t.*(t<=1) + (t+u).*(t>1);
[Z,PZ] = csort(G,P);
PZ2 = (Z<=2)*PZ'
PZ2 =  0.1010                        % Theoretical = 563/5632 = 0.1000
```

Exercise 10.4.12

$$P(Z \le 1/4) = P((X,Y) \in M_1Q_1 \bigvee M_2Q_2),\ M_1 = \{(t,u): 0 \le t \le u \le 1\}$$
$$M_2 = \{(t,u): 0 \le t \le 2, 0 \le t \le \min(t, 2-t)\}$$
$$Q_1 = \{(t,u): t \le 1/2\}\ Q_2 = \{(t,u): u \le 1/2\} \text{ (see figure)}$$

$$P = \frac{24}{11}\int_0^{1/2}\int_0^1 tu\ du\ dt + \frac{24}{11}\int_{1/2}^{3/2}\int_0^{1/2} tu\ du\ dt + \frac{24}{11}\int_{3/2}^{2}\int_0^{2-t} tu\ du\ dt = \frac{85}{176}$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density  (24/11)*t.*u.*(u<=min(1,2-t))
Use array operations on X, Y, PX, PY, t, u, and P
G = 0.5*t.*(u>t) + u.^2.*(u<t);
[Z,PZ] = csort(G,P);
pp = (Z<=1/4)*PZ'
pp =  0.4844                          % Theoretical = 85/176 = 0.4830
```

Exercise 10.4.13

$$P(Z \le 1) = P((X,Y) \in M_1Q_1 \bigvee M_2Q_2),\ M_1 = \{(t,u): 0 \le t \le 1, 0 \le u \le 1-t\}$$
$$M_2 = \{(t,u): 1 \le t \le 2, 0 \le u \le t\}$$
$$Q_1 = \{(t,u): u \le 1-t\}\ Q_2 = \{(t,u): u \le 1/2\}\ \text{(see figure)}$$
$$P = \frac{3}{23}\int_0^1\int_0^{1-t}(t+2u)\ du\ dt + \frac{3}{23}\int_1^2\int_0^{1/2}(t+2u)\ du\ dt = \frac{9}{46}$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  300
Enter number of Y approximation points  300
Enter expression for joint density  (3/23)*(t + 2*u).*(u<=max(2-t,t))
Use array operations on X, Y, PX, PY, t, u, and P
M = max(t,u) <= 1;
G = M.*(t + u) + (1 - M)*2.*u;
p = total((G<=1).*P)
p =  0.1960                           % Theoretical = 9/46 = 0.1957
```

Exercise 10.4.14

$$P(Z \le 2) = P((X,Y) \in M_1Q_1 \bigvee (M_2 \bigvee M_3)Q_2),\ M_1 = \{(t,u): 0 \le t \le 1, 1 \le u \le 2\}$$
$$M_2 = \{(t,u): 0 \le t \le 1, 0 \le u \le 1\}\ M_3 = \{(t,u): 1 \le t \le 2, 0 \le u \le 3-t\}$$
$$Q_1 = \{(t,u): u \le 1-t\}\ Q_2 = \{(t,u): u \le 1/2\}\ \text{(see figure)}$$
$$P = \frac{12}{179}\int_0^1\int_0^{2-t}(3t^2+u)du\ dt + \frac{12}{179}\int_1^2\int_0^1(3t^2+u)du\ dt = \frac{119}{179}$$

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  300
Enter number of Y approximation points  300
Enter expression for joint density  (12/179)*(3*t.^2 + u).*(u<=min(2,3-t))
Use array operations on X, Y, PX, PY, t, u, and P
M = (t<=1)&(u>=1);
Z = M.*(t + u) + (1 - M)*2.*u.^2;
G = M.*(t + u) + (1 - M)*2.*u.^2;
p = total((G<=2).*P)
p =  0.6662                           % Theoretical = 119/179 = 0.6648
```

Exercise 10.4.16

```
% file npr10_16.m  Data for Exercise 16.
cx = [-2 1 3 0];
pmx = 0.001*[255  25 375  45 108  12 162  18];
cy = [1 3 1 -3];
pmy = minprob(0.01*[32 56 40]);
Z = [-1.3 1.2 2.7 3.4 5.8];
```

```
PZ = 0.01*[12 24 43 13  8];
disp('Data are in cx, pmx, cy, pmy, Z, PZ')
npr10_16                % Call for data
Data are in cx, pmx, cy, pmy, Z, PZ
[X,PX] = canonicf(cx,pmx);
[Y,PY] = canonicf(cy,pmy);
icalc3
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter row matrix of Z-values  Z
Enter X probabilities  PX
Enter Y probabilities  PY
Enter Z probabilities  PZ
Use array operations on matrices X, Y, Z,
PX, PY, PZ, t, u, v, and P
M = t.^2 + 3*t.*u.^2 > 3*v;
PM = total(M.*P)
PM =  0.3587
```

This page titled 10.4: Problems on Functions of Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

11: MATHEMATICAL EXPECTATION

Page created for new attachment

11.1: Mathematical Expectation- Simple Random Variables

11.2: Mathematical Expectation and General Random Variables

11.3: Problems on Mathematical Expectation

This page titled 11: Mathematical Expectation is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Mathematical Expectation- Simple Random Variables

Introduction

The probability that real random variable X takes a value in a set M of real numbers is interpreted as the likelihood that the observed value $X(\omega)$ on any trial will lie in M. Historically, this idea of likelihood is rooted in the intuitive notion that if the experiment is repeated enough times the probability is approximately the fraction of times the value of X will fall in M. Associated with this interpretation is the notion of the *average* of the values taken on. We incorporate the concept of *mathematical expectation* into the mathematical model as an appropriate form of such averages. We begin by studying the mathematical expectation of simple random variables, then extend the definition and properties to the general case. In the process, we note the relationship of mathematical expectation to the Lebesque integral, which is developed in abstract measure theory. Although we do not develop this theory, which lies beyond the scope of this study, identification of this relationship provides access to a rich and powerful set of properties which have far reaching consequences in both application and theory.

Expectation for simple random variables

The notion of mathematical expectation is closely related to the idea of a weighted mean, used extensively in the handling of numerical data. Consider the arithmetic average $\bar{x}$ of the following ten numbers: 1, 2, 2, 2, 4, 5, 5, 8, 8, 8, which is given by

$$\bar{x} = \frac{1}{10}(1 + 2 + 2 + 2 + 4 + 5 + 5 + 8 + 8 + 8)$$

Examination of the ten numbers to be added shows that five distinct values are included. One of the ten, or the fraction 1/10 of them, has the value 1, three of the ten, or the fraction 3/10 of them, have the value 2, 1/10 has the value 4, 2/10 have the value 5, and 3/10 have the value 8. Thus, we could write

$$\bar{x} = (0.1 \cdot 1 + 0.3 \cdot 2 + 0.1 \cdot 4 + 0.2 \cdot 5 + 0.3 \cdot 8)$$

The pattern in this last expression can be stated in words: Multiply each possible value by the fraction of the numbers having that value and then sum these products. The fractions are often referred to as the *relative frequencies*. A sum of this sort is known as a *weighted average*.

In general, suppose there are n numbers $\{x_1, x_2, \cdots x_n\}$ to be averaged, with m≤nm≤n distinct values $\{t_1, t_2 \cdots t_m\}$. Suppose f_1 have value t_1, f_2 have value t_2, $\cdots$, f_m have value t_m. The f_i must add to n. If we set $p_i = f_i/n$, then the fraction p_i is called the relative frequency of those numbers in the set which have the value t_i, $1 \le i \le m$. The average $\bar{x}$ of the n numbers may be written

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \sum_{j=1}^{m} t_j p_j$$

In probability theory, we have a similar averaging process in which the relative frequencies of the various possible values of are replaced by the probabilities that those values are observed on any trial.

Definition. For a simple random variable X with values $\{t_1, t_2, \cdots t_n\}$ and corresponding probabilities $p_i = P(X = t_i)$ *mathematical expectation*, designated $E[X]$, is the probability weighted average of the values taken on by X. In symbols

$$E[X] = \sum_{i=1}^{n} t_i P(X = t_i) = \sum_{i=1}^{n} t_i p_i$$

Note that the expectation is determined by the distribution. Two quite different random variables may have the same distribution, hence the same expectation. Traditionally, this average has been called the *mean*, or the *mean value*, of the random variable X.

Example 11.1.1. Some special cases

1. Since $X = aI_E = 0I_{E^c} + aI_E$, we have $E[aI_E] = aP(E)$.
2. For X a constant c, $X = cI_\Omega$, so that $E[c] = cP(\Omega) = c$.

3. If $X = \sum_{i=1}^{n} t_i I_{A_i}$ then $aX = \sum_{i=1}^{n} at_i I_{A_i}$, so that

$$E[aX] = \sum_{i=1}^{n} at_i P(A_i) = a \sum_{i=1}^{n} t_i P(A_i) = aE[X]$$

Figure 1 is a drawing of the moment of a probability distribution about the origin. The expected value of X, E[X], is equal to the sum of the moments, which is equal to the center of mass. The drawing shows one major horizontal line split in half by one major vertical line. As a title, the top of the drawing reads Negative Moments to the left of the vertical line, and Positive Moments to the right, which are meant to distinguish the arrows and labels in the drawing. On the horizontal line are five black dots, two to the left of the vertical line and three to the right. Below the corresponding dots are the corresponding labels: t_1, t_2, t_3, t_4, and t_5. Above the black dots are the following labels: p_1, p_2, p_3, p_4, and p_5. Above the horizontal line is another smaller horizontal line with arrows pointing in both directions. The label for the arrow pointing to the left is t_2 p_2, and the label for the arrow on the left is t_3 p_3. A longer horizontal line sits further up on the drawing, which also has arrows pointing in both directions. and intersects the same vertical line. The arrows are approximately twice as long as the two arrows below. The label for the arrow pointing to the left is t_1 p_1, and the label for the arrow to the left is t_4 p_4. finally, there is one horizontal line extending only to the right of the vertical line, with an arrow pointing to the right. This line is longer in this direction than any of the arrows that sit below it pointing to the right. The arrow is labeled t_5 p_5.

Figure 1. Moment of a probability distribution about the origin.

Mechanical interpretation

In order to aid in visualizing an essentially abstract system, we have employed the notion of probability as mass. The distribution induced by a real random variable on the line is visualized as a unit of probability mass actually distributed along the line. We utilize the mass distribution to give an important and helpful mechanical interpretation of the expectation or mean value. In Example 6 in "Mathematical Expectation: General Random Variables", we give an alternate interpretation in terms of mean-square estimation.

Suppose the random variable X has values $\{t_i; 1 \leq i \leq n\}$, with $P(X = t_i) = p_i$. This produces a probability mass distribution, as shown in Figure 1, with point mass concentration in the amount of p_i at the point t_i. The expectation is

$$\sum_i t_i p_i$$

Now |ti||ti| is the distance of point mass p_i from the origin, with p_i to the left of the origin iff t_i is negative. Mechanically, the sum of the products tipitipi is the *moment* of the probability mass distribution about the origin on the real line. From physical theory, this moment is known to be the same as the product of the total mass times the number which locates the center of mass. Since the total mass is one, the *mean value* is the *location of the center of mass*. If the real line is viewed as a stiff, weightless rod with point mass p_i attached at each value t_i of X, then the mean value μ_X is the point of balance. Often there are symmetries in the distribution which make it possible to determine the expectation without detailed calculation.

Example 11.1.2. the number of spots on a die

Let X be the number of spots which turn up on a throw of a simple six-sided die. We suppose each number is equally likely. Thus the values are the integers one through six, and each probability is 1/6. By definition

$$E[X] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$$

Although the calculation is very simple in this case, it is really not necessary. The probability distribution places equal mass at each of the integer values one through six. The center of mass is at the midpoint.

Example 11.1.3. a simple choice

A child is told she may have one of four toys. The prices are \$2.50. \$3.00, \$2.00, and \$3.50, respectively. She choses one, with respective probabilities 0.2, 0.3, 0.2, and 0.3 of choosing the first, second, third or fourth. What is the expected cost of her selection?

$$E[X] = 2.00 \cdot 0.2 + 2.50 \cdot 0.2 + 3.00 \cdot 0.3 + 3.50 \cdot 0.3 + 2.85$$

For a simple random variable, the mathematical expectation is determined as the dot product of the value matrix with the probability matrix. This is easily calculated using MATLAB.

matlab calculation for example 11.1.3

```
X = [2 2.5 3 3.5];  % Matrix of values (ordered)
PX = 0.1*[2 2 3 3]; % Matrix of probabilities
EX = dot(X,PX)      % The usual MATLAB operation
EX = 2.8500
Ex = sum(X.*PX)     % An alternate calculation
Ex = 2.8500
ex = X*PX'          % Another alternate
ex = 2.8500
```

Expectation and primitive form

The definition and treatment above assumes X is in *canonical form*, in which case

$X = \sum_{i=1}^{n} t_i I_{A_i}$, where $A_i = \{X = t_i\}$, implies $E[X] = \sum_{i=1}^{n} t_i P(A_i)$

We wish to ease this restriction to canonical form.

Suppose simple random variable X is in a *primitive form*

$X = \sum_{j=1}^{m} c_j I_{C_j}$, where $\{C_j : 1 \le j \le m\}$ is a partition

We show that

$$E[X] = \sum_{j=1}^{m} c_j P(C_j)$$

Before a formal verification, we begin with an example which exhibits the essential pattern. Establishing the general case is simply a matter of appropriate use of notation.

Example 11.1.4. simple random variable x in primitive form

$X = I_{C_1} + 2I_{C_2} + I_{C_3} + 3I_{C_4} + 2I_{C_5} + 2I_{C_6}$, with $\{C_1, C_2, C_3, C_4, C_5, C_6\}$ a partition

inspection shows the distinct possible values of X to be 1, 2, or 3. Also

$A_1 = \{X = 1\} = C_1 \bigvee C_3$, $A_2 = \{X = 2\} = C_2 \bigvee C_5 \bigvee C_6$ and $A_3 = \{X = 3\} = C_4$

so that

$P(A-1) = P(C_1) + P(C_3)$, $P(A_2) = P(C_2) + P(C_5) + P(C_6)$, and $P(A_3) = P(C_4)$

Now

$$E[X] = P(A_1) + 2P(A_2) + 3P(A_3) = P(C_1) + P(C_3) + 2[P(C_2) + P(C_5) + P(C_6)] + 3P(C_4)$$

$$= P(C_1) + 2P(C_2) + P(C_3) + 3P(C_4) + 2P(C_5) + 2P(C_6)$$

To establish the general pattern, consider $X = \sum_{j=1}^{m} c_j I_{C_j}$. We identify the distinct set of values contained in the set $\{c_j : 1 \le j \le m\}$. Suppose these are $t_1 < t_2 < \cdots < t_n$. For any value t_i in the range, identify the index set J_i of those j such that $c_j = t_i$. Then the terms

$$\sum_{J_i} c_j I_{C_j} = t_i \sum_{J_i} I_{C_j} = t_i I_{A_i}, \text{ where } A_i = \bigvee_{j \in J_i} C_j$$

By the additivity of probability

$$P(A_i) = P(X = t_i) = \sum_{j \in J_i} P(C_j)$$

Since for each $j \in J_i$ we have $c_j = t_i$, we have

$$E[X] = \sum_{i=1}^{n} t_i P(A_i) = \sum_{i=1}^{n} t_i \sum_{j \in J_i} P(C_j) = \sum_{i=1}^{n} \sum_{j \in J_i} c_j P(C_j) = \sum_{j=1}^{m} c_j P(C_j)$$

— □

Thus, *the defining expression for expectation thus holds for X in a primitive form.*

An alternate approach to obtaining the expectation from a primitive form is to use the csort operation to determine the distribution of X from the coefficients and probabilities of the primitive form.

Example 11.1.5. Alternate determinations of E[X]

Suppose X in a primitive form is

$$X = I_{C_1} + 2I_{C_2} + I_{C_3} + 3I_{C_4} + 2I_{C_5} + 2I_{C_6} + I_{C_7} + 3I_{C_8} + 2I_{C_9} + I_{C_{10}}$$

with respective probabilities

$$P(C_i) = 0.08, 0.11, 0.06, 0.13, 0.05, 0.08, 0.12, 0.07, 0.14, 0.16$$

```
c = [1 2 1 3 2 2 1 3 2 1];             % Matrix of coefficients
pc = 0.01*[8 11 6 13 5 8 12 7 14 16];  % Matrix of probabilities
EX = c*pc'
EX = 1.7800                            % Direct solution
[X,PX] = csort(c,pc);                  % Determinatin of dbn for X
disp([X;PX]')
    1.0000    0.4200
    2.0000    0.3800
    3.0000    0.2000
Ex = X*PX'                             % E[X] from distribution
Ex = 1.7800
```

Linearity

The result on primitive forms may be used to establish the *linearity* of mathematical expectation for simple random variables. Because of its fundamental importance, we work through the verification in some detail.

Suppose $X = \sum_{i=1}^{n} t_i I_{A_i}$ and $Y = \sum_{j=1}^{m} u_j I_{B_j}$ (both in canonical form). Since

$$\sum_{i=1}^{n} I_{A_i} = \sum_{j=1}^{m} I_{B_j} = 1$$

we have

$$X + Y = \sum_{i=1}^{n} t_i I_{A_i} (\sum_{j=1}^{m} I_{B_j}) + \sum_{j=1}^{m} u_j I_{B_j} (\sum_{i=1}^{n} I_{A_i}) = \sum_{i=1}^{n} \sum_{j=1}^{m} (t_i + u_j) I_{A_i} I_{B_j}$$

Note that $I_{A_i} I_{B_j} = I_{A_i B_j}$ and $A_i B_j = \{X = t_i, Y = u_j\}$. The class of these sets for all possible pairs (i, j) forms a partition. Thus, the last summation expresses $Z = X + Y$ in a primitive form. Because of the result on primitive forms, above, we have

$$E[X + Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} (t_i + u_j) P(A_i B_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} t_i P(A_i B_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} u_j P(A_i B_j)$$

$$= \sum_{i=1}^{n} t_i \sum_{j=1}^{m} P(A_i B_j) + \sum_{j=1}^{m} u_j \sum_{i=1}^{n} P(A_i B_j)$$

We note that for each i and for each j

$$P(A_i) = \sum_{j=1}^{m} P(A_i B_j) \text{ and } P(B_j) = \sum_{i=1}^{n} P(A_i B_j)$$

Hence, we may write

$$E[X + Y] = \sum_{i=1}^{n} t_i P(A_i) + \sum_{j=1}^{m} u_j P(B_j) = E[X] + E[Y]$$

Now aX and bY are simple if X and Y are, so that with the aide of Example 11.1.1 we have

$$E[aX + bY] = E[aX] + E[bY] = aE[X] + bE[Y]$$

If X, Y, Z are simple, then so are $aX + bY$, and cZ. It follows that

$$E[aX + bY + cZ] = E[aX + bY] + cE[Z] = aE[X] + bE[Y] + cE[Z]$$

By an inductive argument, this pattern may be extended to a linear combination of any finite number of simple random variables. Thus we may assert

Linearity. The expectation of a linear combination of a finite number of simple random variables is that linear combination of the expectations of the individual random variables.

— □

Expectation of a simple random variable in affine form

As a direct consequence of linearity, whenever simple random variable X is in affine form, then

$$E[X] = E[c_0 + \sum_{i=1}^{n} c_i I_{E_i}] = c_0 + \sum_{i=1}^{n} c_i P(E_i)$$

Thus, the defining expression holds for any affine combination of indicator functions, whether in canonical form or not.

Example 11.1.6. binomial distribution (n,p)

This random variable appears as the number of successes in n Bernoulli trials with probability *p* of success on each component trial. It is naturally expressed in affine form

$$X = \sum_{i=1}^{n} I_{E_i} \text{ so that } E[X] = \sum_{i=1}^{n} p = np$$

Alternately, in canonical form

$$X = \sum_{k=0}^{n} k I_{A_{kn}}, \text{ with } p_k = P(A_{kn}) = P(X = k) = C(n, k) p^k q^{n-k}, \; q = 1 - p$$

so that

$$E[X] = \sum_{k=0}^{n} k C(n, k) p^k q^{n-k}, \; q = 1 - p$$

Some algebraic tricks may be used to show that the second form sums to np, but there is no need of that. The computation for the affine form is much simpler.

Example 11.1.7. Expected winnings

A bettor places three bets at \$2.00 each. The first bet pays \$10.00 with probability 0.15, the second pays \$8.00 with probability 0.20, and the third pays \$20.00 with probability 0.10. What is the expected gain?

Solution

The net gain may be expressed

$$X = 10I_A + 8I_B + 20I_C - 6, \text{ with } P(A) = 0.15,\ P(B) = 0.20,\ P(C) = 0.10$$

Then

$$E[X] = 10 \cdot 0.15 + 8 \cdot 0.20 + 20 \cdot 0.10 - 6 = -0.90$$

These calculations may be done in MATLAB as follows:

```
c = [10 8 20 -6];
p = [0.15 0.20 0.10 1.00]; % Constant a = aI_(Omega), with P(Omega) = 1
E = c*p'
E = -0.9000
```

Functions of simple random variables

If X is in a primitive form (including canonical form) and g is a real function defined on the range of X, then

$$Z = g(X) = \sum_{j=1}^{m} g(c_j) I_{C_j} \text{ a primitive form}$$

so that

$$E[Z] = E[g(X)] = \sum_{j=1}^{m} g(c_j) P(C_j)$$

Alternately, we may use csort to determine the distribution for Z and work with that distribution.

Caution. If X is in affine form (but not a primitive form)

$$X = c_0 + \sum_{j=1}^{m} c_j I_{E_j} \text{ then } g(X) \neq g(c_0) + \sum_{j=1}^{m} g(c_j) I_{E_j}$$

so that

$$E[g(X)] \neq g(c_0) + \sum_{j=1}^{m} g(c_j) P(E_j)$$

Example 11.1.8. expectation of a function of x

Suppose X in a primitive form is

$$X = -3I_{C_1} - I_{C_2} + 2I_{C_3} - 3I_{C_4} + 4I_{C_5} - I_{C_6} + I_{C_7} + 2I_{C_8} + 3I_{C_9} + 2I_{C_{10}}$$

with probabilities $P(C_i) = 0.08, 0.11, 0.06, 0.13, 0.05, 0.08, 0.12, 0.07, 0.14, 0.16$.

Let $g(t) = t^2 + 2t$. Determine $E(g(X))$.

```
c = [-3 -1 2 -3 4 -1 1 2 3 2];            % Original coefficients
pc = 0.01*[0 11 6 13 5 8 12 7 14 16];     % Probabilities for C_j
G = c.^2 + 2*c                            % g(c_j)
G = 3  -1  8  3  24  -1  3  8  15  8
EG = G*pc'                                % Direct computation
EG = 6.4200
[Z,PZ] = csort(G,pc);                     % Distribution for Z = g(X)
disp([Z; PZ]')
    -1.0000    0.1900
     3.0000    0.3300
     8.0000    0.2900
    15.0000    0.1400
    24.0000    0.0500
EZ = Z*PZ'                                % E[Z] from distribution for Z
EZ = 6.4200
```

A similar approach can be made to a function of a pair of simple random variables, provided the joint distribution is available. Suppose $X = \sum_{i=1}^{n} t_i I_{A_i}$ and $Y = \sum_{j=1}^{m} u_j I_{B_j}$ (both in canonical form). Then

$$Z = g(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i, u_j) I_{A_i B_j}$$

The $A_i B_j$ form a partition, so Z is in a primitive form. We have the same two alternative possibilities: (1) direct calculation from values of $g(t_i, u_j)$ and corresponding probabilities $P(A_i B_j) = P(X = t_i, Y = u_j)$, or (2) use of csort to obtain the distribution for Z.

Example 11.1.9. expectation for z = g(x,y)

We use the joint distribution in file jdemo1.m and let $g(t,u) = t^2 + 2tu - 3u$. To set up for calculations, we use jcalc.

```
% file jdemo1.m
X = [-2.37 -1.93 -0.47 -0.11 0 0.57 1.22 2.15 2.97 3.74];
Y = [-3.06 -1.44 -1.21 0.07 0.88 1.77 2.01 2.84];
P = 0.0001*[ 53   8 167 170 184  18  67 122  18  12;
             11  13 143 221 241 153  87 125 122 185;
            165 129 226 185  89 215  40  77  93 187;
            165 163 205  64  60  66 118 239  67 201;
            227   2 128  12 238 106 218 120 222  30;
             93  93  22 179 175 186 221  65 129   4;
            126  16 159  80 183 116  15  22 113 167;
            198 101 101 154 158  58 220 230 228 211];
```

```
jdemo1                    % Call for data
jcalc                     % Set up
Enter JOINT PROBABILITIES (as on the plane)   P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = t.^2 + 2*t.*u - 3*u; % Calculation of matrix of [g(t_i, u_j)]
EG = total(G.*P)          % Direct claculation of expectation
EG = 3.2529
[Z,PZ] = csort(G,P);      % Determination of distribution for Z
EZ = Z*PZ'                % E[Z] from distribution
EZ = 3.2529
```

This page titled 11.1: Mathematical Expectation- Simple Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Mathematical Expectation and General Random Variables

In this unit, we extend the definition and properties of mathematical expectation to the general case. In the process, we note the relationship of mathematical expectation to the Lebesque integral, which is developed in abstract measure theory. Although we do not develop this theory, which lies beyond the scope of this study, identification of this relationship provides access to a rich and powerful set of properties which have far reaching consequences in both application and theory.

Extension to the General Case

In the unit on Distribution Approximations, we show that a bounded random variable X can be represented as the limit of a nondecreasing sequence of simple random variables. Also, a real random variable can be expressed as the difference $X = X^+ - X^-$ of two nonnegative random variables. The extension of mathematical expectation to the general case is based on these facts and certain basic properties of simple random variables, some of which are established in the unit on expectation for simple random variables. We list these properties and sketch how the extension is accomplished.

Definition: almost surely

A condition on a random variable or on a relationship between random variables is said to hold *almost surely*, abbreviated "a.s." iff the condition or relationship holds for all ω except possibly a set with probability zero.

Basic properties of simple random variables

(E0) : If $X = Y$ a.s. then $E[X] = E[Y]$.
(E1): $E(aI_E) = aP(E)$.
(E2): *Linearity*. $X = \sum_{i=1}^{n} a_i X_i$ implies $E[X] = \sum_{i=1}^{n} a_i E[X_i]$|
(E3): *Positivity: monotonicity*
a. If $X \geq 0$ a.s. , then $E[X] \geq 0$, with equality iff $X = 0$ a.s. .
b. If $X \geq Y$ a.s. , then $E[X] \geq E[Y]$, with equality iff $X = Y$ a.s. .
(E4): *Fundamental lemma* If $X \geq 0$ is bounded and $\{X_n : 1 \leq n\}$ is an a.s. nonnegative, nondecreasing sequence with $\lim_n X_n(\omega) \geq X(\omega)$ for almost every ω, then $\lim_n E[X_n] \geq E[X]$.
(E4a): If for all n, $0 \leq X_n \leq X_{n+1}$ a.s. and $X_n \to X$ a.s. , then $E[X_n] \to E[X]$ (i.e. , the expectation of the limit is the limit of the expectations).

Ideas of the proofs of the fundamental properties

- Modifying the random variable X on a set of probability zero simply modifies one or more of the A_i without changing $P(A_i)$
- Properties (E1) and (E2) are established in the unit on expectation of simple random variables..
- Positivity (E3a) is a simple property of sums of real numbers. Modification of sets of probability zero cannot affect the expectation.
- Monotonicity (E3b) is a consequence of positivity and linearity.

$$X \geq Y \text{ iff } X - Y \geq 0 \text{ a.s. and } E[X] \geq E[Y] \text{ iff } E[X] - E[Y] = E[X - Y] \geq 0$$

- The fundamental lemma (E4) plays an essential role in extending the concept of expectation. It involves elementary, but somewhat sophisticated, use of linearity and monotonicity, limited to nonnegative random variables and positive coefficients. We forgo a proof.
- Monotonicity and the fundamental lemma provide a very simple proof of the monotone convergence theoem, often designated MC. Its role is essential in the extension.

Nonnegative random variables

There is a nondecreasing sequence of nonnegative simple random variables converging to X. Monotonicity implies the integrals

of the nondecreasing sequence is a nondecreasing sequence of real numbers, which must have a limit or increase without bound (in which case we say the limit is infinite). We define $E[X] = \lim E[X_n]$.

Two questions arise.

Is the limit unique? The approximating sequences for a simple random variable are not unique, although their limit is the same.
Is the definition consistent? If the limit random variable X is simple, does the new definition coincide with the old?

The fundamental lemma and monotone convergence may be used to show that the answer to both questions is affirmative, so that the definition is reasonable. Also, the six fundamental properties survive the passage to the limit.

As a simple applications of these ideas, consider discrete random variables such as the geometric (p) or Poisson (μ), which are integer-valued but unbounded.

Example 11.2.1: Unbounded, nonnegative, integer-valued random variables

The random variable X may be expressed

$$X = \sum_{k=0}^{\infty} kI_{E_k}, \text{ where } E_k = \{X = k\} \text{ with } P(E_k) = p_k$$

Let

$$X_n = \sum_{k=0}^{n-1} kI_{E_k} + nI_{B_n}, \text{ where } B_n = \{X \geq n\}$$

Then each X_n is a simple random variable with $X_n \leq X_{n+1}$. If $X(\omega) = k$, then $X_n(\omega) = k = X(\omega)$ for all $n \geq k + 1$. Hence, $X_n(\omega) \to X(\omega)$ for all ω. By monotone convergence, $E[X_n] \to E[X]$. Now

$$E[X_n] = \sum_{k=1}^{n-1} kP(E_k) + nP(B_n)$$

If $\sum_{k=0}^{\infty} kP(E_k) < \infty$, then

$$0 \leq nP(B_n) = n\sum_{k=n}^{\infty} P(E_k) \leq \sum_{k=n}^{\infty} kP(E_k) \to 0 \text{ as } n \to \infty$$

Hence

$$E[X] = \lim_n E[X_n] = \sum_{k=0}^{\infty} kP(A_k)$$

We may use this result to establish the expectation for the geometric and Poisson distributions.

Example 11.2.2: X~geometric (p)

We have $p_k = P(X = k) = q^k p$. $0 \leq k$. By the result of Example 11.2.1.

$$E[X] = \sum_{k=0}^{\infty} kpq^k = pq\sum_{k=1}^{\infty} kq^{k-1} = \frac{pq}{(1-q)^2} = q/p$$

For $Y - 1$ ~ geometric (p), $p_k = pq^{k-1}$ so that $E[Y] = \frac{1}{q}E[X] = 1/p$

Example 11.2.3: X~poisson (μ)

We have $p_k = e^{-\mu}\frac{\mu^k}{k!}$. By the result of Example 11.2.1.

$$E[X] = e^{-\mu}\sum_{k=0}^{\infty} k\frac{\mu^k}{k!} = \mu e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu}e^{\mu} = \mu$$

The general case

We make use of the fact that $X = X^+ - X^-$, where both X^+ and X^- are nonnegative. Then

$$E[X] = E[X^+] - E[X^-] \text{ provided at least one of } E[X^+], E[X^-] \text{ is finite}$$

Definition. If both $E[X^+]$ and $E[X^-]$ are finite, X is said to be *integrable*.

The term integrable comes from the relation of expectation to the abstract Lebesgue integral of measure theory.

Again, the basic properties survive the extension. The property (E0) is subsumed in a more general uniqueness property noted in the list of properties discussed below.

Theoretical note

The development of expectation sketched above is exactly the development of the Lebesgue integral of the random variable X as a measurable function on the basic probability space (Ω, F, P), so that

$$E[X] = \int_{\Omega} X\, dP \tag{11.2.1}$$

As a consequence, we may utilize the properties of the general Lebesgue integral. In its abstract form, it is not particularly useful for actual calculations. A careful use of the mapping of probability mass to the real line by random variable X produces a corresponding mapping of the integral on the basic space to an integral on the real line. Although this integral is also a Lebesgue integral it agrees with the ordinary Riemann integral of calculus when the latter exists, so that ordinary integrals may be used to compute expectations.

Additional properties

The fundamental properties of simple random variables which survive the extension serve as the basis of an extensive and powerful list of properties of expectation of real random variables and real functions of random vectors. Some of the more important of these are listed in the table in Appendix E. We often refer to these properties by the numbers used in that table.

Some basic forms

The mapping theorems provide a number of basic integral (or summation) forms for computation.

In general, if $Z = g(X)$ with distribution functions F_X and F_Z, we have the expectation as a Stieltjes integral.

$$E[Z] = E[g(X)] = \int g(t)F_X(dt) = \int uF_Z(du)$$

If X and $g(X)$ are absolutely continuous, the Stieltjes integrals are replaced by

$$E[Z] = \int g(t)f_X(t)\, dt = \int uF_Z(du)$$

where limits of integration are determined by f_X or f_Y. Justification for use of the density function is provided by the Radon-Nikodym theorem—property (E19).

If X is simple, in a primitive form (including canonical form), then

$$E[Z] = E[g(X)] = \sum_{j=1}^{m} g(c_j)P(C_j)$$

If the distribution for $Z = g(X)$ is determined by a csort operation, then

$$E[Z] = \sum_{k=1}^{n} v_k P(Z = v_k)$$

The extension to unbounded, nonnegative, integer-valued random variables is shown in Example 11.2.1, above. The finite sums are replaced by infinite series (provided they converge).

For $Z = g(X, Y)$,

$$E[Z] = E[g(X, Y)] = \int \int g(t, u) F_{XY}(dtdu) = \int v F_Z(dv)$$

In the absolutely continuous case

$$E[Z] = E[g(X, Y)] = \int \int g(t, u) f_{XY}(t, u) dudt = \int v f_Z(v) dv$$

For joint simple X, Y (Section on Expectation for Simple Random Variables)

$$E[Z] = E[g(X, Y)] = \sum_{i=1}^{n} \sum_{j=1}^{m} g(t_i, u_j) P(X = t_i, Y = u_j)$$

Mechanical interpretation and approximation procedures

In elementary mechanics, since the total mass is one, the quantity $E[X] = \int t f_X(t)\ dt$ is the location of the center of mass. This theoretically rigorous fact may be derived heuristically from an examination of the expectation for a simple approximating random variable. Recall the discussion of the m-procedure for discrete approximation in the unit on Distribution Approximations The range of X is divided into equal subintervals. The values of the approximating random variable are at the midpoints of the subintervals. The associated probability is the probability mass in the subinterval, which is approximately $f_X(t_i)dx$, where dx is the length of the subinterval. This approximation improves with an increasing number of subdivisions, with corresponding decrease in dxdx X_s is

$$E[X_s] = \sum_i t_i f_X(t_i) dx \approx \int t f_X(t)\ dt$$

The approximation improves with increasingly fine subdivisions. The center of mass of the approximating distribution approaches the center of mass of the smooth distribution.

It should be clear that a similar argument for $g(X)$ leads to the integral expression

$$E[g(X)] = \int g(t) f_X(t)\ dt$$

This argument shows that we should be able to use tappr to set up for approximating the expectation $E[g(X)]$ as well as for approximating $P(g(X) \in M)$, etc. We return to this in Section.

Mean values for some absolutely continuous distributions

Uniform on $[a, b] f_X(t) = \dfrac{1}{b-a}$, $a \le t \le b$ The center of mass is at $(a + b)/2$. To calculate the value formally, we write

$$E[X] = \int t f_X(t) dt = \frac{1}{b-a} \int_a^b t dt = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

Symmetric triangular on$[a, b]$ The graph of the density is an isoceles triangle with base on the interval $[a, b]$. By symmetry, the center of mass, hence the expectation, is at the midpoint $(a + b)/2$.

Exponential(λ) $f_X(t) = \lambda e^{-\lambda t}$, $0 \le t$ Using a well known definite integral (see Appendix B), we have

$$E[X] = \int t f_X(t)\ dt = \int_0^\infty \lambda t e^{-\lambda t} dt = 1/\lambda$$

Gamma(α, λ) $f_X(t) = \dfrac{1}{\Gamma}(\alpha) t^{\alpha-1} \lambda^\alpha e^{-\lambda t}$, $0 \le t$ Again we use one of the integrals in Appendix B to obtain

$$E[X] = \int t f_X(t)\ dt = \frac{1}{\Gamma} \int_0^\infty \lambda^\alpha t^\alpha e^{-\lambda t} dt = \frac{\Gamma(\alpha + 1)}{\lambda \Gamma(\alpha)} = a/lambda$$

The last equality comes from the fact that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

Beta(r, s). $f_X(t) = \dfrac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} t^{r-1}(1-t)^{s-1}$, $0 < t < 1$ We use the fact that

$\int_0^1 u^{r-1}(1-u)^{s-1}\ du = \dfrac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$, $r > 0$, $s > 0$.

$$E[X] = \int t f_X(t)\ dt = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_0^1 t^r (1-t)^{s-1} dt = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \cdot \frac{\Gamma(r+1)\Gamma(s)}{\Gamma(r+s+1)} = \frac{r}{r+s}$$

Weibull(α, λ, v). $F_X(t) = 1 - e^{-\lambda(t-v)^\alpha}$ $\alpha > 0$, $\lambda > 0$, $v \geq 0$, $t \geq v$. Differentiation shows

$$f_X(t) = \alpha\lambda(t-v)^{\alpha-1} e^{-\lambda(t-v)^\alpha},\ t \geq v$$

First, consider Y ~ exponential (λ). For this random variable

$$E[Y^r] = \int_0^\infty t^r \lambda e^{-\lambda t}\ dt = \frac{\Gamma(r+1)}{\lambda^r}$$

If Y is exponential (1), then techniques for functions of random variables show that $[\frac{1}{\lambda} Y]^{1/\alpha} + v$ ~ Weibull $(\alpha, lambda, v)$. Hence,

$$E[X] = \frac{1}{\lambda^{1/\alpha}} E[Y^{1/\alpha}] + v = \frac{1}{\lambda^{1/\alpha}} \Gamma(\frac{1}{\alpha} + 1) + v$$

Normal(μ, σ^2) The symmetry of the distribution about $t = \mu$ shows that $E[X] = \mu$. This, of course, may be verified by integration. A standard trick simplifies the work.

$$E[X] = \int_{-\infty}^\infty t f_X(t)\ dt = \int_{-\infty}^\infty (t-\mu) f_X(t)\ dt + \mu$$

We have used the fact that $\int_{-\infty}^\infty f_X(t)\ dt = 1$. If we make the change of variable $x = t - \mu$ in the last integral, the integrand becomes an odd function, so that the integral is zero. Thus, $E[X] = \mu$.

Properties and Computation

The properties in the table in Appendix E constitute a powerful and convenient resource for the use of mathematical expectation. These are properties of the abstract Lebesgue integral, expressed in the notation for mathematical expectation.

$$E[g(X)] = \int g(X)\ dP \tag{11.2.2}$$

In the development of additional properties, the four basic properties: (E1) Expectation of indicator functions, (E2) Linearity, (E3) Positivity; monotonicity, and (E4a) Monotone convergence play a foundational role. We utilize the properties in the table, as needed, often referring to them by the numbers assigned in the table.

In this section, we include a number of examples which illustrate the use of various properties. Some are theoretical examples, deriving additional properties or displaying the basis and structure of some in the table. Others apply these properties to facilitate computation

Example 11.2.4: Probability as expectation

Probability may be expressed entirely in terms of expectation.

- By properties (E1) and positivity (E3a), $P(A) = E[I_A] \geq 0$.
- As a special cases of (E1), we have $P(\Omega) = E[I_\Omega] = 1$
- By the countable sums property (E8),

$$A = \bigvee_i A_i \text{ implies } P(A) = E[I_A] = E[\sum_i I_{A_i}] = \sum_i E[I_{A_i}] = \sum_i P(A_i)$$

Thus, the three defining properties for a probability measure are satisfied.

Remark. There are treatments of probability which characterize mathematical expectation with properties (E0) through (E4a), then define $P(A) = E[I_A]$. Although such a development is quite feasible, it has not been widely adopted.

Example 11.2.5: An indicator function pattern

Suppose X is a real random variable and $E = X^{-1}(M) = \{\omega : X(\omega) \in M\}$. Then

$$I_E = I_M(X)$$

To see this, note that $X(\omega) \in M$ iff $\omega \in E$, so that $I_E(\omega) = 1$ iff $I_M(X(\omega)) = 1$.

Similarly, if $E = X^{-1}(M) \cap Y^{-1}(N)$, then $I_E = I_M(X)I_N(Y)$. We thus have, by (E1).

$$P(X \in M) = E[I_M(X)] \text{ and } P(X \in M, Y \in N) = E[I_M(X)I_N(Y)]$$

Example 11.2.6: Alternate interpretation of the mean value

$E[(X - c)^2]$ is a minimum iff $c = E[X]$, in which case $E[(X - E[X])^2] = E[X^2] - E^2[X]$

INTERPRETATION. If we approximate the random variable X by a constant c, then for any ω the error of approximation is $X(\omega) - c$. The probability weighted average of the square of the error (often called the *mean squared error*) is $E[(X - c)^2]$. This average squared error is smallest iff the approximating constant c is the mean value.

verification

We expand $(X - c)^2$ and apply linearity to obtain

$$E[(X - c)^2 = E[X^2 - 2cX + c^2] = E[X^2] - 2E[X]c + c^2$$

The last expression is a quadratic in c (since $E[X^2]$ and $E[X]$ are constants). The usual calculus treatment shows the expression has a minimum for $c = E[X]$. Substitution of this value for c shows the expression reduces to $E[X^2] - E^2[X]$.

A number of inequalities are listed among the properties in the table. The basis for these inequalities is usually some standard analytical inequality on random variables to which the monotonicity property is applied. We illustrate with a derivation of the important Jensen's inequality.

Example 11.2.7: Jensen's inequality

If X is a real random variable and g is a convex function on an interval I which includes the range of X, then

verification

The function g is convex on I iff for each $t_0 \in [a, b]$ there is a number $\lambda(t_0)$ such that

$$g(t) \geq g(t_0) + \lambda(t_0)(t - t_0)$$

This means there is a line through $(t_0, g(t_0))$ such that the graph of g lies on or above it. If $a \leq X \leq b$, then by monotonicity $E(a) = a \leq E[X] \leq E[b] = b$ (this is the mean value property (E11)). We may choose $t_0 = E[X] \in I$. If we designate the constant $\lambda(E[X])$ by c, we have

$$g(X) \geq g(E[X]) + c(X - E[X])$$

Recalling that $E[X]$ is a constant, we take expectation of both sides, using linearity and monotonicity, to get

$$E[g(X)] \geq g(E[X]) + c(E[X] - E[X]) = g(E[X])$$

Remark. It is easy to show that the function $\lambda(\cdot)$ is nondecreasing. This fact is used in establishing Jensen's inequality for conditional expectation.

The product rule for expectations of independent random variables

Example 11.2.8: product rule for simple random variables

Consider an independent pair $\{X, Y\}$ of simple random variables

$$X = \sum_{i=1}^{n} t_i I_{A_i}\ Y = \sum_{j=1}^{m} u_j I_{B_j} \text{ (both in canonical form)}$$

We know that each pair $\{A_i, B_j\}$ is independent, so that $P(A_iB_j) = P(A_i)P(B_j)$. Consider the product XY. According to the pattern described after Example 9 from "Mathematical Expectation: Simple Random Variables."

$$XY = \sum_{i=1}^{n} t_i I_{A_i} \sum_{j=1}^{m} u_j I_{B_j} = \sum_{i=1}^{n} \sum_{j=1}^{m} t_i u_j I_{A_iB_j}$$

The latter double sum is a primitive form, so that

$$E[XY] = \sum_{i=1}^{n} \sum_{j=1}^{m} t_i u_j P(A_iB_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} t_i u_j P(A_i)P(B_j) = (\sum_{i=1}^{n} t_i P(A_i))(\sum_{j=1}^{m} u_j P(B_j)) = E[X]E[Y]$$

Thus the product rule holds for independent simple random variables.

Example 11.2.9: approximating simple functions for an independent pair

Suppose $\{X, Y\}$ is an independent pair, with an approximating simple pair $\{X_s, Y_s\}$. As functions of X and Y, respectively, the pair $\{X_s, Y_s\}$ is independent. According to Example, above, the product rule $E[X_sY_s] = E[X_s]E[Y_s]$ must hold.

Example 11.2.10. product rule for an independent pair

For $X \geq 0$, $Y \geq 0$, there exist nondecreasing sequences $\{X_n : 1 \leq n\}$ and $\{Y_n : 1 \leq n\}$ of simple random variables increasing to X and Y, respectively. The sequence $\{X_nY_n : 1 \leq n\}$ is also a nondecreasing sequence of simple random variables, increasing to XY. By the monotone convergence theorem (MC)

$$E[X_n] \nearrow E[X],\ E[Y_n] \nearrow E[Y], \text{ and } E[X_nY_n] \nearrow E[XY]$$

Since $E[X_nY_n] = E[X_n]E[Y_n]$ for each n, we conclude $E[XY] = E[X]E[Y]$

In the general case,

$$XY = (X^+ - X^-)(Y^+ - Y^-) = X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^-$$

Application of the product rule to each nonnegative pair and the use of linearity gives the product rule for the pair $\{X, Y\}$

Remark. It should be apparent that the product rule can be extended to any finite independent class.

Example 11.2.11: the joint distribution of three random variables

The class $\{X, Y, Z\}$ is independent, with the marginal distributions shown below. Let

$W = g(X, Y, Z) = 3X^2 + 2XY - 3XYZ$. Determine $E[W]$.

```
X = 0:4;
Y = 1:2:7;
Z = 0:3:12;
PX = 0.1*[1 3 2 3 1];
PY = 0.1*[2 2 3 3];
PZ = 0.1*[2 2 1 3 2];
```

```
icalc3                                          % Setup for joint dbn for {X,Y,Z}
Enter row matrix of X-values   X
Enter row matrix of Y-values   Y
Enter row matrix of Z-values   Z
Enter X probabilities  PX
Enter Y probabilities  PY
Enter Z probabilities  PZ
Use array operations on matrices  X, Y, Z,
PX, PY, PZ, t, u, v, and P
EX = X*PX'                                      % E[X]
EX =    2
EX2 = (X.^2)*PX'                                % E[X^2]
EX2 = 5.4000
EY = Y*PY'                                      % E[Y]
EY =  4.4000
EZ = Z*PZ'                                      % E[Z]
EZ =  6.3000
G = 3*t.^2 + 2*t.*u - 3*t.*u.*v;                % W = g(X,Y,Z) = 3X^2 + 2XY - 2XYZ
```

Example 11.2.12. a function with a compound definition: truncated exponential

Suppose X ~ exponential (0, 3). Let

$$Z = \begin{cases} X^2 & \text{for } X \le 4 \\ 16 & \text{for } X > 4 \end{cases} = I_{[0,4]}(X)X^2 + I_{(4,\infty]}(X)16$$

Determine $E(Z)$.

Analytic Solution

$$E[g(X)] = \int g(t) f_X(t)\, dt = \int_0^\infty I_{[0,4]}(t) t^2 0.3e^{-0.3t}\, dt + 16E[I_{(4,\infty]}(X)]$$

$$= \int_0^4 t^2 0.3e^{-0.3t}\, dt + 16P(X > 4) \approx 7.4972 \text{ (by Maple)}$$

APPROXIMATION

To obtain a simple aproximation, we must approximate the exponential by a bounded random variable. Since $P(X > 50) = e^{-15} \approx 3 \cdot 10^{-7}$ we may safety truncate X at 50.

```
tappr
Enter matrix [a b] of x-range endpoints [0 50]
Enter number of x approximation points 1000
Enter density as a function of t 0.3*exp(-0.3*t)
Use row matrices X and PX as in the simple case
M = X <= 4
G = M.*X.^2 + 16*(1 - M); % g(X)
```

```
EG = G*PX'                % E[g(X)]
EG = 7.4972
[Z,PZ] = csort(G,PX);     % Distribution for Z = g(X)
EZ = Z*PZ'                % E[Z] from distribution
EZ = 7.4972
```

Because of the large number of approximation points, the results agree quite closely with the theoretical value.

Example 11.2.13. stocking for random demand (see exercise 4 from "Problems on functions of random variables")

The manager of a department store is planning for the holiday season. A certain item costs c dollars per unit and sells for p dollars per unit. If the demand exceeds the amount m ordered, additional units can be special ordered for s dollars per unit $(s > c)$. If demand is less than amount ordered, the remaining stock can be returned (or otherwise disposed of) at r dollars per unit $(r < c)$. Demand D for the season is asumed to be a random variable with Poisson (μ) distribution. Suppose $\mu = 50$, $c = 30$, $p = 50$, $s = 40$, $r = 20$. What about m should the manager order to maximize the expected profit?

PROBLEM FORMULATION

Suppose D is the demand and X is the profit. Then

For $D \le m$, $X = D(p - c) - (m - D)(c - r) = D(p - r) + m(r - c)$
For $D > m$, $X = m(p - c) + (D - m)(p - s) = D(p - s) + m(s - c)$

It is convenient to write the expression for X in terms of I_M, where $M = (-\infty, m]$. Thus

$$X = I_M(D)[D(p - r) + m(r - c)] + [1 - I_M(D)][D(p - s) + m(s - c)]$$

$$= D(p - s) + m(s - c) + I_M(D)[D(p - r) + m(r - c) - D(p - s) - m(s - c)]$$

$$= D(p - s) + m(s - c) + I_M(D)(s - r)(D - m)$$

Then \(E[X] = (p - c) E[D] + m(s - c) + (s - r) E[I_M(D) D] - (s - r) m E[I_M (D)].

Analytic Solution

For D ~ Poisson (μ), $E[D] = \mu$ and $E[I_M(D)] = P(D \le m)$

$$E[I_M(D)D] = e^{-\mu} \sum_{k=1}^{m} k\frac{\mu^k}{k!} = \mu e^{-\mu} \sum_{k=1}^{m} \frac{\mu^{k-1}}{(k-1)!} = \mu P(D \le m - 1)$$

Hence,

$$E[X] = (p - s)E[D] + m(s - c) + (s - r)E[I_M(D)D] - (s - r)mE[I_M(D)]$$

$$= (p - s)\mu + m(s - c) + (s - r)\mu P(D \le m - 1) - (s - r)mP(D \le m)$$

Because of the discrete nature of the problem, we cannot solve for the optimum m by ordinary calculus. We may solve for various m about $m = \mu$ and determine the optimum. We do so with the aid of MATLAB and the m-function cpoisson.

```
mu = 50;
c  = 30;
p  = 50;
s  = 40;
r  = 20;
m  = 45:55;
```

```
EX = (p - s)*mu + m*(s - c) + (s - r)*mu*(1 - cpoisson(mu, m))...
-(s - r)*m.*(1 - cpoisson(mu,m+1));
disp([m;EX]')
    45.0000    930.8604
    46.0000    935.5231
    47.0000    939.1895
    48.0000    941.7962
    49.0000    943.2988
    50.0000    943.6750            % Optimum m = 50
    51.0000    942.9247
    52.0000    941.0699
    53.0000    938.1532
    54.0000    934.2347
    55.0000    929.3886
```

A direct, solution may be obtained by MATLAB, using finite approximation for the Poisson distribution.

APPROXIMATION

```
ptest = cpoisson(mu,100)            %Check for suitable value of n
ptest = 3.2001e-10
n = 100;
t = 0:n;
pD = ipoisson(mu,t);
for i = 1:length(m)                 % Step by step calculation for various m
    M = t > m(i);
    G(i,:) = t*(p - r) - M.*(t - m(i))*(s - r) - m(i)*(c - r);
end
EG = G*pD';                         % Value agree with theoretical to four decimals
```

An advantage of the second solution, based on simple approximation to *D*, is that the distribution of gain for each m could be studied — e.g., the maximum and minimum gains.

— □

Example 11.2.14. a jointly distributed pair

Suppose the pair $\{X, Y\}$ has joint density $f_{XY}(t, u) = 3u$ on the triangular region bounded by $u = 0$, $u = 1 + t$, $u = 1 - t$ (see Figure 11.2.1). Let $Z = g(X, Y) = X^2 + 2XY$. Determine $E[Z]$.

Figure 1 is a density drawing, with a horizontal axis labeled as t, and a vertical axis labeled as u. A triangle of width 2, with a base sitting on the horizontal axis, from t=-1 to t=1. The third point of the triangle (the one not on the horizontal axis) is directly above, on the vertical axis. The drawing of the triangle is thus divided in equal halves by the vertical axis. The side of the triangle on the horizontal axis has no direct label. The side of the triangle on the left is labeled u = 1 + t, and the side of the triangle on the right is labeled u = 1- t. A caption below the triangle reads f_xy (t, u) - 3u on the triangle.

Figure 11.2.1. The density for Example 11.2.14.

Analytic Solution

$E[Z] = \int\int (t^2 + 2tu) f_{XY}(t, u)\ dudt$

$$= 3\int_{-1}^{0}\int_{0}^{1+t}(t^2u + 2tu^2)\,dudt + 3\int_{0}^{1}\int_{0}^{1-t}(t^2u + 2tu^2)\,dudt = 1/10$$

APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints [-1 1]
Enter matrix [c d] of Y-range endpoints [0 1]
Enter number of X approximation points 400
Enter number of Y approximation points 200
Enter expression for joint density 3*u.*(u<=min(1+t,1-t))
Use array operations on X, Y, PX, PY, t, u, and P
G = t.^2 + 2*t.*u;                % g(X,Y) = X^2 + 2XY
EG = total(G.*P)                  % E[g(X,Y)]
EG = 0.1006                       % Theoretical value = 1/10
[Z, PZ] = csort(G,P);             % Distribution for Z
EZ = Z*PZ'                        % E[Z] from distribution
EZ = 0.1006
```

Example 11.2.15. Afunction with a compound definition

The pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = 1/2$ on the square region bounded to $u = 1 + t, u = 1 - t, u = 3 - t$, and $u = t - 1$ (see Figure 11.2.2).

$$W = \begin{cases} X & \text{for } \max\{X,Y\} \le 1 \\ 2Y & \text{for } \max\{X,Y\} > 1 \end{cases} = I_Q(X,Y)X + I_{Q^c}(X,Y)2Y$$

where $Q = \{(t,u) : \max\{t,u\} \le 1\} = \{(t,u) : t \le 1, u \le 1\}$. Determine $E[W]$.

Figure 2 is a density drawing, with a horizontal axis labeled as t, and a vertical axis labeled u. The drawing is a shaded square rotated 45 degrees to be sitting with one point on the horizontal axis. The point sits on (1, 0) and a second point sits against the vertical axis, at (0, 1). In looking at the drawing it can be deduced that the third vertex is at (1, 2), and that the fourth vertex is at (2, 1). Each side of the square is labeled with an equation. Starting with the side between the vertices that are sitting on the axes, an reading them clockwise, the equations are listed as u= 1 - t, u = 1 + t, u= 3 - t, and u = t - 1. There is also an equation inside the shaded square, reading f_xy (t, u) = 1/2.

Figure 11.2.2. The density for Example 11.2.15

Analytic Solution

The intersection of the region Q and the square is the set for which $0 \le t \le 1$ and $1 - t \le u \le 1$. Reference to the figure shows three regions of integration.

$$E[W] = \frac{1}{2}\int_0^1\int_{1-t}^{1} t\,dudt + \frac{1}{2}\int_0^1\int_1^{1+t} 2u\,dudt + \frac{1}{2}\int_1^2\int_{t-1}^{3-t} 2u\,dudt = 11/6 \approx 1.8333$$

APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints [0 2]
Enter matrix [c d] of Y-range endpoints [0 2]
Enter number of X approximation points 200
Enter number of Y approximation points 200
```

```
Enter expression for joint density ((u<=min(t+1,3-t))& ...
        (u>=max(1-t,t-1))/2
Use array operation on X, Y, PX, PY, t, u, and P
M = max(t,u)<=1;
G = t.*M + 2*u.*(1 - M);    % Z = g(X,Y)
EG = total(G.*P)            % E[g(X,Y)]
EG = 1.8340                 % Theoretical 11/6 = 1.8333
[Z,PZ] = csort(G,P);        % Distribution for Z
EZ = dot(Z,PZ)              % E[Z] from distribution
EZ = 1.8340
```

Special forms for expectation

The various special forms related to property (E20a) are often useful. The general result, which we do not need, is usually derived by an argument which employs a general form of what is known as Fubini's theorem. The special form (E20b)

$$E[X] = \int_{-\infty}^{\infty} [u(t) - F_X(t)]\ dt$$

may be derived from (E20a) by use of integration by parts for Stieltjes integrals. However, we use the relationship between the graph of the distribution function and the graph of the quantile function to show the equivalence of (E20b) and (E20f). The latter property is readily established by elementary arguments.

Example 11.2.16. The property (e20f)

If Q is the quantile function for the distribution function F_X, then

$$E[g(X)] = \int_0^1 g[G(u)]\ du$$

VERIFICATION

If $Y = Q(U)$, where U ~ uniform on (0, 1), then Y has the same distribution as X. Hence,

$$E[g(X)] = E[g(Q(U))] = \int g(Q(u)) f_U(u)\ du = \int_0^1 g(Q(u))\ du$$

Example 11.2.17. Reliability and expectation

In reliability, if X is the life duration (time to failure) for a device, the reliability function is the probability at any time t the device is still operative. Thus

$$R(t) = P(X > t) = 1 - F_X(t)$$

According to property (E20b)

$$E[X] = \int_0^{\infty} R(t)\ dt$$

Example 11.2.18. Use of the quantile function

Suppose $F_X(t) = t^a$, $a > 0$, $0 \le t \le 1$. Then $Q(u) = u^{1/a}$, $0 \le u \le a$.

$$E[X] = \int_0^1 u^{1/a}\ du = \frac{1}{1 + 1/a} = \frac{a}{a+1}$$

The same result could be obtained by using $f_X(t) = F_X'(t)$ and evaluating $\int t f_X(t)\ dt$.

Example 11.2.19. Equivalence of (e20b) and (e20f)

For the special case $g(X) = X$. Figure 3(a) shows \int_{0}^{1} Q(u) \ du\) is the difference in the shaded areas

$\int_0^1 Q(u)\ du = \text{Area } A - \text{Area } B$

The corresponding graph of the distribution function *F* is shown in Figure 11.2.3(b). Because of the construction, the areas of the regions marked A and B are the same in the two figures. As may be seen,

$$\text{Area } A = \int_0^\infty [1 - F(t)]\ dt \text{ and Area } B = \int_{-\infty}^0 F(t)\ dt$$

Use of the unit step function $u(t) = 1$ for $t > 0$ and 0 for $t < 0$ (defined arbitrarily at $t = 0$) enables us to combine the two expressions to get

$$\int_0^1 Q(u)\ du = \text{Area } A - \text{Area } B = \int_{-\infty}^\infty [u(t) - F(t)]\ dt$$

Figure three contains two graphs. The first graph has a horizontal axis labeled t, and a vertical axis labeled u. The large label of the graph reads, u = Q(t). A dashed vertical line along t = 1 bounds an increasing curved plot. The curve starts with a vertical asymptote along the vertical axis below the horizontal axis, and as it approaches the horizontal axis, the slope becomes more shallow. The curve's slope shallows until it is midway in horizontal distance between the vertical axis and the dashed vertical line. At this point, the slope begins to increase again, until it reaches a vertical asymptote along the dashed line at t = 1. The horizontal and vertical axes, along with the curve itself, create a bounded shape. A small right triangle loosely fits this bounded shape, and is labeled as B. The dashed line, horizontal axis, and the segment of the curve above the horizontal axis create a larger bounded shape, and a larger right triangle loosely fits this bounded shape, labeled A. The second graph is roughly similar. The axes are in the same place, but with this figure, s dashed line is now drawn horizontally along u = 1. A curve of the same shape now begins at the origin, and approaches the dashed line as a horizontal asymptote to the t - axis. It increases in slope at an increasing rate for half of the vertical distance, and then decreases in slope back to a horizontal asymptote at u = 1. The same triangles fitting the first figure are used in the second figure, only because of the rotated nature of the new curve, these triangles are rotated in the same fashion.

Figure 11.2.3. Equivalent areas in (E20b) and (E20f).

Property (E20c) is a direct result of linearity and (E20b), with the unit step functions cancelling out.

Example 11.2.20. Property (e20d) useful inequalities

Suppose $X \geq 0$. Then

$$\sum_{n=0}^\infty P(X \geq n+1) \leq E[X] \leq \sum_{n=0}^\infty P(X \geq n) \leq N \sum_{k=0}^\infty P(X \geq kN), \text{ for all } N \geq 1$$

VERIFICATION

For $X \geq 0$, by (E20b)

$$E[X] = \int_0^\infty [1 - F(t)]\ dt = \int_0^\infty P(X > t)\ dt$$

Since F can have only a countable number of jumps on any interval and $P(X > t$ and $P(X \geq t)$ differ only at jump points, we may assert

$$\int_a^b P(X > t)\ dt = \int_a^b P(X \geq t)\ dt$$

For each nonnegative integer n, let $E_n = [n, n+1]$. By the countable additivity of expectation

$$E[X] = \sum_{n=0}^{\infty} E[I_{E_n} X] = \sum_{n=0}^{\infty} \int_{E_n} P(X \geq t)\, dt$$

Since $P(X \geq t)$ is decreasing with t and each E_n has unit length, we have by the mean value theorem

$$P(X \geq n+1) \leq E[I_{E_n} X] \leq P(X \geq n)$$

The third inequality follows from the fact that

$$\int_{kN}^{(k+1)N} P(X \geq t)\, dt \leq N \int_{E_{kN}} P(X \geq t)\, dt \leq NP(X \geq kN)$$

Remark. Property (E20d) is used primarily for theoretical purposes. The special case (E20e) is more frequently used.

Example 11.2.21. Property (e20e)

If X is nonnegative, integer valued, then

$$E[X] = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

VERIFICATION

The result follows as a special case of (E20d). For integer valued random variables,

$$P(X \geq t) = P(X \geq n) \text{ on } E_n \text{ and } P(X \geq t) = P(X > n) = P(X \geq n+1) \text{ on } E_{n+1}$$

An elementary derivation of (E20e) can be constructed as follows.

Example 11.2.22. (e20e) for integer-valued random variables

By definition

$$E[X] = \sum_{k=1}^{\infty} kP(X = k) = \lim_n \sum_{k=1}^{n} kP(X = k)$$

Now for each finite n,

$$\sum_{k=1}^{n} kP(X = k) = \sum_{k=1}^{n} \sum_{j=1}^{k} P(X = k) = \sum_{j=1}^{n} \sum_{k=j}^{n} P(X = k) = \sum_{j=1}^{n} P(X \geq j)$$

Taking limits as $n \to \infty$ yields the desired result.

Example 11.2.23. the geometric distribution

Suppose X ~ geometric (p). Then $P(X \geq k) = q^k$. Use of (E20e) gives

$$E[X] = \sum_{k=1}^{\infty} q^k = q \sum_{k=0}^{\infty} q^k = \frac{q}{1-q} = q/p$$

This page titled 11.2: Mathematical Expectation and General Random Variables is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

11.3: Problems on Mathematical Expectation

Exercise 11.3.1

(See Exercise 1 from "Problems on Distribution and Density Functions", m-file npr07_01.m). The class $\{C_j : 1 \leq j \leq 10\}$ is a partition. Random variable X has values {1, 3, 2, 3, 4, 2, 1, 3, 5, 2} on C_1 through C_{10}, respectively, with probabilities 0.08, 0.13, 0.06, 0.09, 0.14, 0.11, 0.12, 0.07, 0.11, 0.09. Determine $E[X]$

Answer available at the end of this section.

Exercise 11.3.2

(See Exercise 2 from "Problems on Distribution and Density Functions", m-file npr07_02.m). A store has eight items for sale. The prices are \$3.50, \$5.00, \$3.50, \$7.50, \$5.00, \$5.00, \$3.50, and \$7.50, respectively. A customer comes in. She purchases one of the items with probabilities 0.10, 0.15, 0.15, 0.20, 0.10 0.05, 0.10 0.15. The random variable expressing the amount of her purchase may be written

$$X = 3.5I_{C_1} + 5.0I_{C_2} + 3.5I_{C_3} + 7.5I_{C_4} + 5.0I_{C_5} + 5.0I_{C_6} + 3.5I_{C_7} + 7.5I_{C_8}$$

Determine the expection $E[X]$ of the value of her purchase.

Answer available at the end of this section.

Exercise 11.3.3

See Exercise 12 from "Problems on Random Variables and Probabilities", and Exercise 3 from "Problems on Distribution and Density Functions," m-file npr06_12.m). The class $\{A, B, C, D\}$ has minterm probabilities

$$pm = 0.001 * [5\ 7\ 6\ 8\ 9\ 14\ 22\ 33\ 21\ 32\ 50\ 75\ 86\ 129\ 201\ 302]$$

Determine the mathematical expection for the random variable $X = I_A + I_B + I_C + I_D$, which counts the number of the events which occur on a trial.

Answer available at the end of this section.

Exercise 11.3.4

(See Exercise 5 from "Problems on Distribution and Density Functions"). In a thunderstorm in a national park there are 127 lightning strikes. Experience shows that the probability of of a lightning strike starting a fire is about 0.0083. Determine the expected number of fires.

Answer available at the end of this section.

Exercise 11.3.5

(See Exercise 8 from "Problems on Distribution and Density Functions"). Two coins are flipped twenty times. Let X be the number of matches (both heads or both tails). Determine $E[X]$

Answer available at the end of this section.

Exercise 11.3.6

(See Exercise 12 from "Problems on Distribution and Density Functions"). A residential College plans to raise money by selling "chances" on a board. Fifty chances are sold. A player pays \$10 to play; he or she wins \$30 with probability $p = 0.2$. The profit to the College is

$$X = 50 \cdot 10 - 30N, \text{ where } N \text{ is the numbe of winners}$$

Determine the expected profit $E[X]$.

Answer available at the end of this section.

Exercise 11.3.7

(See Exercise 19 from "Problems on Distribution and Density Functions"). The number of noise pulses arriving on a power circuit in an hour is a random quantity having Poisson (7) distribution. What is the expected number of pulses in an hour?

Answer available at the end of this section.

Exercise 11.3.8

(See Exercise 24 and Exercise 25 from "Problems on Distribution and Density Functions"). The total operating time for the units in Exercise 24 is a random variable T ~ gamma (20, 0.0002). What is the expected operating time?

Answer available at the end of this section.

Exercise 11.3.9

(See Exercise 41 from "Problems on Distribution and Density Functions"). Random variable X has density function

$$f_X(t) = \begin{cases} (6/5)t^2 & \text{for } 0 \le t \le 1 \\ (6/5)(2-t) & \text{for } 1 \le t \le 2 \end{cases} = I_{[0,1]}(t)\frac{6}{5}t^2 + I_{(1,2]}(t)\frac{6}{5}(2-t).$$

What is the expected value $E[X]$?

Answer available at the end of this section.

Exercise 11.3.10

Truncated exponential. Suppose X ~ exponential (λ) and $Y = I_{[0,a]}(X)X + I_{a,\infty}(X)a$.

a. Use the fact that

$$\int_0^\infty te^{-\lambda t}\,dt = \frac{1}{\lambda^2} \text{ and } \int_a^\infty te^{-\lambda t}\,dt = \frac{1}{\lambda^2}e^{-\lambda t}(1+\lambda a)$$

to determine an expression for $E[Y]$.

b. Use the approximation method, with $\lambda = 1/50$, $a = 30$. Approximate the exponential at 10,000 points for $0 \le t \le 1000$. Compare the approximate result with the theoretical result of part (a).

Answer available at the end of this section.

Exercise 11.3.11

(See Exercise 1 from "Problems On Random Vectors and Joint Distributions", m-file npr08_01.m). Two cards are selected at random, without replacement, from a standard deck. Let X be the number of aces and Y be the number of spades. Under the usual assumptions, determine the joint distribution. Determine $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$, and $E[XY]$.

Answer available at the end of this section.

Exercise 11.3.12

(See Exercise 2 from "Problems On Random Vectors and Joint Distributions", m-file npr08_02.m). Two positions for campus jobs are open. Two sophomores, three juniors, and three seniors apply. It is decided to select two at random (each possible pair equally likely). Let X be the number of sophomores and Y be the number of juniors who are selected. Determine the joint distribution for $\{X, Y\}$ and $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$, and $E[XY]$.

Answer available at the end of this section.

Exercise 11.3.13

(See Exercise 3 from "Problems On Random Vectors and Joint Distributions", m-file npr08_03.m). A die is rolled. Let *X* be the number of spots that turn up. A coin is flipped X times. Let Y be the number of heads that turn up. Determine the joint distribution for the pair $\{X, Y\}$. Assume $P(X = k) = 1/6$ for $1 \leq k \leq 6$ and for each k, $P(Y = j|X = k)$ has the binomial $(k, 1/2)$ distribution. Arrange the joint matrix as on the plane, with values of Y increasing upward. Determine the expected value $E[Y]$

Answer available at the end of this section.

Exercise 11.3.14

(See Exercise 4 from "Problems On Random Vectors and Joint Distributions", m-file npr08_04.m). As a variation of Exercise, suppose a pair of dice is rolled instead of a single die. Determine the joint distribution for $\{X, Y\}$ and determine $E[Y]$.

Answer available at the end of this section.

Exercise 11.3.15

(See Exercise 5 from "Problems On Random Vectors and Joint Distributions", m-file npr08_05.m). Suppose a pair of dice is rolled. Let X be the total number of spots which turn up. Roll the pair an additional X times. Let Y be the number of sevens that are thrown on the X rolls. Determine the joint distribution for $\{X, Y\}$ and determine $E[Y]$

Answer available at the end of this section.

Exercise 11.3.16

(See Exercise 6 from "Problems On Random Vectors and Joint Distributions", m-file npr08_06.m). The pair $\{X, Y\}$ has the joint distribution:

$$X = [-2.3\ -0.7\ 1.1\ 3.9\ 5.1]\ Y = [1.3\ 2.5\ 4.1\ 5.3]$$

$$P = \begin{bmatrix} 0.0483 & 0.0357 & 0.0420 & 0.0399 & 0.0441 \\ 0.0437 & 0.0323 & 0.0380 & 0.0361 & 0.0399 \\ 0.0713 & 0.0527 & 0.0620 & 0.0609 & 0.0551 \\ 0.0667 & 0.0493 & 0.0580 & 0.0651 & 0.0589 \end{bmatrix}$$

Determine $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ and $E[XY]$.

Answer available at the end of this section.

Exercise 11.3.17

(See Exercise 7 from "Problems On Random Vectors and Joint Distributions", m-file npr08_07.m). The pair $\{X, Y\}$ has the joint distribution:

$$P(X = t, Y = u)$$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189
-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

Determine $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ and $E[XY]$.

Answer available at the end of this section.

Exercise 11.3.18

(See Exercise 8 from "Problems On Random Vectors and Joint Distributions", m-file npr08_08.m). The pair $\{X, Y\}$ has the joint distribution:

$$P(X = t, Y = u)$$

t=	1	3	5	7	9	11	13	15	17	19
u = 12	0.0156	0.0191	0.0081	0.0035	0.0091	0.0070	0.0098	0.0056	0.0091	0.0049
10	0.0064	0.0204	0.0108	0.0040	0.0054	0.0080	0.0112	0.0064	0.0104	0.0056
9	0.0196	0.0256	0.0126	0.0060	0.0156	0.0120	0.0168	0.0096	0.0056	0.0084
5	0.0112	0.0182	0.0108	0.0070	0.0182	0.0140	0.0196	0.0012	0.0182	0.0038
3	0.0060	0.0260	0.0162	0.0050	0.0160	0.0200	0.0280	0.0060	0.0160	0.0040
-1	0.0096	0.0056	0.0072	0.0060	0.0256	0.0120	0.0268	0.0096	0.0256	0.0084
-3	0.0044	0.0134	0.0180	0.0140	0.0234	0.0180	0.0252	0.0244	0.0234	0.0126
-5	0.0072	0.0017	0.0063	0.0045	0.0167	0.0090	0.0026	0.0172	0.0217	0.0223

Determine $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ and $E[XY]$.

Answer available at the end of this section.

Exercise 11.3.19

(See Exercise 9 from "Problems On Random Vectors and Joint Distributions", m-file npr08_09.m). Data were kept on the effect of training time on the time to perform a job on a production line. X is the amount of training, in hours, and Y is the time to perform the task, in minutes. The data are as follows:

$$P(X = t, Y = u)$$

t =	1	1.5	2	2.5	3
u = 5	0.039	0.011	0.005	0.001	0.001
4	0.065	0.070	0.050	0.015	0.010
3	0.031	0.061	0.137	0.051	0.033
2	0.012	0.049	0.163	0.058	0.039
1	0.003	0.009	0.045	0.025	0.017

Determine $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ and $E[XY]$.

Answer available at the end of this section.

For the joint densities in Exercise 20-32 below

a. Determine analytically $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ and $E[XY]$.

b. Use a discrete approximation for $E[X]$, $E[Y]$, $E[X^2]$, $E[Y^2]$ and $E[XY]$.

Exercise 11.3.20

(See Exercise 10 from "Problems On Random Vectors and Joint Distributions"). $f_{XY}(t, u) = 1$ for $0 \le t \le 1$. $0 \le u \le 2(1 - t)$.

$$f_X(t) = 2(1 - t),\ 0 \le t \le 1,\ f_Y(u) = 1 - u/2,\ 0 \le u \le 2$$

Answer available at the end of this section.

Exercise 11.3.21

(See Exercise 11 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t, u) = 1/2$ on the square with vertices at (1, 0), (2, 1) (1, 2), (0, 1).

$$f_X(t) = f_Y(t) = I_{[0,1]}(t)t + I_{(1,2]}(t)(2 - t)$$

Answer available at the end of this section.

Exercise 11.3.22

(See Exercise 12 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t, u) = 4t(1 - u)$ for $0 \le t \le 1$. $0 \le u \le 1$

$$f_X(t) = 2t,\ 0 \le t \le 1,\ f_Y(u) = 2(1 - u),\ 0 \le u \le 1$$

Answer available at the end of this section.

Exercise 11.3.23

(See Exercise 13 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{1}{8}(t+u)$ for $0 \le t \le 2$, $0 \le u \le 2$

$$f_X(t) = f_Y(t) = \frac{1}{4}(t+1),\ 0 \le t \le 2$$

Answer available at the end of this section.

Exercise 11.3.24

(See Exercise 14 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = 4ue^{-2t}$ for $0 \le t, 0 \le u \le 1$

$$f_X(t) = 2e^{-2t},\ 0 \le t,\ f_Y(u) = 2u,\ 0 \le u \le 1$$

Answer available at the end of this section.

Exercise 11.3.25

(See Exercise 15 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{3}{88}(2t+3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1 + t$.

$$f_X(t) = \frac{3}{88}(1+t)(1+4t+t^2) = \frac{3}{88}(1+5t+5t^2+t^3),\ 0 \le t \le 2$$

$$f_Y(t) = I_{[0,1]}(u)\frac{3}{88}(6u^2+4) + I_{(1,3]}(u)\frac{3}{88}(3+2u+8u^2-3u^3)$$

Answer available at the end of this section.

Exercise 11.3.26

(See Exercise 16 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = 12t^2u$ on the parallelogram with vertices

(-1, 0), (0, 0), (1, 1), (0, 1)

$$f_X(t) = I_{[-1,0]}(t)6t^2(t+1)^2 + I_{(0,1]}(t)6t^2(1-t^2),\ f_Y(u)12u^3 - 12u^2 + 4u,\ 0 \le u \le 1$$

Answer available at the end of this section.

Exercise 11.3.27

(See Exercise 17 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2-t\}$.

$$f_X(t) = I_{[0,1]}(t)\frac{12}{11}t + I_{(1,2]}(t)\frac{12}{11}t(2-t)^2,\ f_Y(u) = \frac{12}{11}u(u-2)^2,\ 0 \le u \le 1$$

Answer available at the end of this section.

Exercise 11.3.28

(See Exercise 18 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{3}{23}(t+2u)$ for $0 \leq t \leq 2$, $0 \leq u \leq \max\{2-t, t\}$.

$$f_X(t) = I_{[0,1]}(t)\frac{6}{23}(2-t) + I_{(1,2]}(t)\frac{6}{23}t^2,\ f_Y(u) = I_{[0,1]}(u)\frac{6}{23}(2u+1) + I_{(1,2]}(u)\frac{3}{23}(4+6u-4u^2)$$

Answer available at the end of this section.

Exercise 11.3.29

(See Exercise 19 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{12}{179}(3t^2+u)$, for $0 \leq t \leq 2$, $0 \leq u \leq \min\{2, 3-t\}$.

$$f_X(t) = I_{[0,1]}(t)\frac{24}{179}(3t^2+1) + I_{(1,2]}(t)\frac{6}{179}(9-6t+19t^2-6t^3)$$

$$f_Y(u) = I_{[0,1]}(t)\frac{24}{179}(4+u) + I_{(1,2]}(t)\frac{12}{179}(27-24u+8u^2-u^3)$$

Answer available at the end of this section.

Exercise 11.3.30

(See Exercise 20 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{12}{227}(3t+2tu)$, for $0 \leq t \leq 2$, $0 \leq u \leq \min\{1+t, 2\}$.

$$f_X(t) = I_{[0,1]}(t)\frac{12}{227}(t^3+5t^2+4t) + I_{(1,2]}(t)\frac{120}{227}t$$

$$f_Y(u) = I_{[0,1]}(t)\frac{24}{227}(2u+3) + I_{(1,2]}(u)\frac{6}{227}(2u+3)(3+2u-u^2)$$

$$= I_{[0,1]}(u)\frac{24}{227}(2u+3) + I_{(1,2]}(u)\frac{6}{227}(9+12u+u^2-2u^3)$$

Answer available at the end of this section.

Exercise 11.3.31

(See Exercise 21 from "Problems On Random Vectors and Joint Distribution"). $f_{XY}(t,u) = \frac{2}{13}(t+2u)$, for $0 \leq t \leq 2$, $0 \leq u \leq \min\{2t, 3-t\}$.

$$f_X(t) = I_{[0,1]}(t)\frac{12}{13}t^2 + I_{(1,2]}(t)\frac{6}{13}(3-t)$$

$$f_Y(u) = I_{[0,1]}(u)(\frac{4}{13} + \frac{8}{13}u - \frac{9}{52}u^2) + I_{(1,2]}(u)(\frac{9}{13} + \frac{6}{13}u - \frac{51}{52}u^2)$$

Answer available at the end of this section.

Exercise 11.3.32

(See Exercise 22 from "Problems On Random Vectors and Joint Distribution").

$$f_{XY}(t,u) = I_{[0,1]}(t)\frac{3}{8}(t^2+2u) + I_{(1,2]}(t)\frac{9}{14}t^2u^2, \text{ for } 0 \le u \le 1.$$

$$f_X(t) = I_{[0,1]}(t)\frac{3}{8}(t^2+1) + I_{(1,2]}(t)\frac{3}{14}t^2,\ f_Y(u) = \frac{1}{8} + \frac{3}{4}u + \frac{3}{2}u^2 \text{ (0 \le u \le 1)}$$

Answer available at the end of this section.

Exercise 11.3.33

The class $\{X, Y, Z\}$ of random variables is iid(independent, identically distributed) with common distribution

$$X = \text{[-5 -1 3 4 7]}\ PX = 0.01 * \text{[15 20 30 25 10]}$$

Let $W = 3X - 4Y + 2Z$. Determine $E[W]$. Do this using icalc, then repeat with icalc3 and compare results.

Answer available at the end of this section.

Exercise 11.3.34

(See Exercise 5 from "Problems on Functions of Random Variables") The cultural committee of a student organization has arranged a special deal for tickets to a concert. The agreement is that the organization will purchase ten tickets at \$20 each (regardless of the number of individual buyers). Additional tickets are available according to the following schedule:

11-20, \$18 each; 21-30 \$16 each; 31-50, \$15 each; 51-100, \$13 each

If the number of purchasers is a random variable X, the total cost (in dollars) is a random quantity $Z = g(X)$ described by

$$g(X) = 200 + 18I_{M1}(X)(X-10) + (16-18)I_{M2}(X)(X-20) +$$

$$(15-16)I_{M3}(X)(X-30) + (13-15)I_{M4}(X)(X-50)$$

where $M1 = [10, \infty)$, $M2 = [20, \infty)$, $M3 = [30, \infty)$, $M4 = [50, \infty)$

Suppose X ~ Poisson (75). Approximate the Poisson distribution by truncating at 150. Determine $E[Z]$ and $E[Z^2]$.

Answer available at the end of this section.

Exercise 11.3.35

The pair $\{X, Y\}$ has the joint distribution (in m-file npr08_07.m):

$P(X = t, Y = u)$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189

-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

Let $Z = g(X, Y) = 3X^2 + 2XY - Y^2)$. Determine $E[Z]$ and $E[Z^2]$.

Answer available at the end of this section.

Exercise 11.3.36

For the pair $\{X, Y\}$ in Exercise 11.3.35, let

$$W = g(X, Y) = \begin{cases} X & \text{for } X + Y \leq 4 \\ 2Y & \text{for } X + Y > 4 \end{cases} = I_M(X, Y)X + I_{M^c}(X, Y)2Y$$

Determine $E[W]$ and $E[W^2]$.

Answer available at the end of this section.

For the distribution in Exercises 37-41 below

a. Determine analytically $E[Z]$ and $E[Z^2]$
b. Use a discrete approximation to calculate the same quantities.

Exercise 11.3.37

$f_{XY}(t, u) = \dfrac{3}{88}(2t + 3u^2)$ for $0 \leq t \leq 2$, $0 \leq u \leq 1 + t$ (see Exercise 25).

$$Z = I_{[0,1]}(X)4X + I_{(1,2]}(X)(X + Y)$$

Answer available at the end of this section.

Exercise 11.3.38

$f_{XY}(t, u) = \dfrac{24}{11}tu$ for $0 \leq t \leq 2$, $0 \leq u \leq \min\{1, 2 - t\}$ (see Exercise 27)

$$Z = I_M(X, Y)\frac{1}{2}X + I_{M^c}(X, Y)Y^2, \; M = \{(t, u) : u > t\}$$

Answer available at the end of this section.

Exercise 11.3.39

$f_{XY}(t, u) = \dfrac{3}{23}(t + 2u)$ for $0 \leq t \leq 2$, $0 \leq u \leq \max\{2 - t, t\}$ (see Exercise 28)

$$Z = I_M(X, Y)(X + Y) + I_{M^c}(X, Y)2Y, \; M = \{(t, u) : \max(t, u) \leq 1\}$$

Answer available at the end of this section.

Exercise 11.3.40

$f_{XY}(t,u) = \frac{12}{179}(3t^2 + u)$, for $0 \le t \le 2$, $0 \le u \le \min\{2, 3-t\}$ (see Exercise 19)

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y^2,\ M = \{(t,u) : t \le 1, u \ge 1\}$$

Answer available at the end of this section.

Exercise 11.3.41

$f_{XY}(t,u) = \frac{12}{227}(2t + 2tu)$, for $0 \le t \le 2$, $0 \le u \le \min\{1+t, 2\}$ (see Exercise 30).

$$Z = I_M(X,Y)X + I_{M^c}(X,Y)XY,\ M = \{(t,u) : u \le \min(1, 2-t)\}$$

Answer available at the end of this section.

Exercise 11.3.42

The class $\{X, Y, Z\}$ is independent. (See Exercise 16 from "Problems on Functions of Random Variables", m-file npr10_16.m)

$X = -2I_A + I_B + 3I_C$. Minterm probabilities are (in the usual order)

0.255 0.025 0.375 0.045 0.108 0.012 0.162 0.018

$Y = I_D + 3I_E + I_F - 3$. The class $\{D, E, F\}$ is independent with

$$P(D) = 0.32\ P(E) = 0.56\ P(F) = 0.40$$

Z has distribution

Value	-1.3	1.2	2.7	3.4	5.8
Probability	0.12	0.24	0.43	0.13	0.08

$W = X^2 + 3XY^2 - 3Z$. Determine $E[W]$ and $E[W^2]$.

Answer available at the end of this section.

Answers

Exercise 11.3.1

```
% file npr07_01.m
% Data for Exercise 1 from "Problems on Distribution and Density Functions"
T = [1 3 2 3 4 2 1 3 5 2];
pc = 0.01*[8 13 6 9 14 11 12 7 11 9];
disp('Data are in T and pc')
npr07_01
Data are in T and pc
EX = T*pc'
EX = 2.7000
[X,PX] csort(T,pc): % Alternate using X, PX
ex = X*PX'
```

```
ex = 2.7000
```

Exercise 11.3.2

```
% file npr07_02.m
% Data for Exercise 2 from "Problems on Distribution and Density Functions"
T = [3.5 5.0 3.5 7.5 5.0 5.0 3.5 7.5];
pc = 0.01*[10 15 15 20 10 5 10 15];
disp('Data are in T and pc')
npr07_02
Data are in T and pc
EX = T*pc'
EX = 5.3500
[X,PX] csort(T,pc)
ex = X*PX'
ex = 5.3500
```

Exercise 11.3.3

```
% file npr06_12.m
% Data for Exercise 12 from "Problems on Random Variables and Probabilities"
pm = 0.001*[5 7 6 8 9 14 22 33 21 32 50 75 86 129 201 302];
c = [1 1 1 1 0];
disp('Minterm probabilities in pm, coefficients in c')
npr06_12
Minterm probabilities in pm, coefficients in c
canonic
 Enter row vector of coefficients c
 Enter row vector of minterm probabilities pm
Use row matrices X and PX for calculations
call for XDBN to view the distribution
EX = X*PX'
EX = 2.9890
T = sum(mintable(4));
[x,px] = csort(T,pm);
ex = x*px
ex = 2.9890
```

Exercise 11.3.4

X ~ binomial (127, 0.0083), $E[X] = 127 \cdot 0.0083 = 1.0541$

Exercise 11.3.5

X ~ binomial (20, 1/2). $E[X] = 20 \cdot 0.5 = 10$

Exercise 11.3.6

N ~ binomial (50, 0.2). $E[N] = 50 \cdot 0.2 = 10$. $E[X] = 500 - 30E[N] = 200$.

Exercise 11.3.7

X ~ Poisson (7). $E[X] = 7$.

Exercise 11.3.8

X ~ gamma (20, 0.0002). $E[X] = 20/0.0002 = 100,000$.

Exercise 11.3.9

$$E[X] = \int t f_X(t)\, dt = \frac{6}{5}\int_0^1 t^3\, dt + \frac{6}{5}\int_1^2 (2t - t^2)\, dt = \frac{11}{10}$$

Exercise 11.3.10

$$E[Y] = \int g(t) f_X(t)\, dt = \int_0^a t\lambda e^{-\lambda t}\, dt + aP(X > a) =$$
$$\frac{\lambda}{\lambda^2}[1 - e^{-\lambda a}(1 + \lambda a)] + ae^{-\lambda a} = \frac{1}{\lambda}(1 - e^{-\lambda a})$$

```
tappr
Enter matrix [a b] of x-range endpoints [0 1000]
Enter number of x approximation points 10000
Enter density as a function of t (1/50)*exp(-t/50)
```

```
Use row matrices X and PX as in the simple case
G = X.*(X<=30) + 30*(X>30);
EZ = G8PX'
EZ = 22.5594
ez = 50*(1-exp(-30/50))     %Theoretical value
ez = 22.5594
```

Exercise 11.3.11

```
npr08_01
Data in Pn, P, X, Y
jcalc
Enter JOINT PROBABILITIES (as on the plane) P
Enter row marix of VALUES of X    X
Enter row marix of VALUES of Y    Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
EX = X*PX'
EX = 0.1538

ex = total(t.*P)            % Alternate
ex = 0.1538
EY = Y*PY'
EY = 0.5000
EX2 = (X.^2)*PX'
EX2 = 0.1629
EY2 = (Y.^2)*PY'
EY2 = 0.6176
EXY = total(t.*u.*P)
EXY = 0.0769
```

Exercise 11.3.12

```
npr08_02
Data are in X, Y, Pn, P
jcalc
-----------------------
EX = X*PX'
EX = 0.5000
EY = Y*PY'
EY = 0.7500
EX2 = (X.^2)*PX'
EX2 = 0.5714
EY2 = (Y.^2)*PY'
EY2 = 0.9643
EXY = total(t.*u.*P)
EXY = 0.2143
```

Exercise 11.3.13

```
npr08_03
Data are in X, Y, P, PY
jcalc
-----------------------
EX = X*PX'
EX = 3.5000
EY = Y*PY'
EY = 1.7500
EX2 = (X.^2)*PX'
EX2 = 15.1667
```

```
EY2 = (Y.^2)*PY'
EY2 = 4.6667
EXY = total(t.*u.*P)
EXY = 7.5833
```

Exercise 11.3.14

```
npr08_04
Data are in X, Y, P
jcalc
-----------------------
EX = X*PX'
EX = 7
EY = Y*PY'
EY = 3.5000
EX2 = (X.^2)*PX'
EX2 = 54.8333
EY2 = (Y.^2)*PY'
EY2 = 15.4583
```

Exercise 11.3.15

```
npr08_05
Data are in X, Y, P, PY
jcalc
-----------------------
EX = X*PX'
EX = 7.0000
EY = Y*PY'
EY = 1.1667
```

Exercise 11.3.16

```
npr08_06
Data are in X, Y, P
jcalc
----------------------
EX = X*PX'
EX = 1.3696
EY = Y*PY'
EY = 3.0344
EX2 = (X.^2)*PX'
EX2 = 9.7644
EY2 = (Y.^2)*PY'
EY2 = 11.4839
EXY = total(t.*u.*P)
EXY = 4.1423
```

Exercise 11.3.17

```
npr08_07
Data are in X, Y, P
jcalc
----------------------
EX = X*PX'
EX = 0.8590
EY = Y*PY'
EY = 1.1455
EX2 = (X.^2)*PX'
EX2 = 5.8495
```

```
EY2 = (Y.^2)*PY'
EY2 = 19.6115
EXY = total(t.*u.*P)
EXY = 3.6803
```

Exercise 11.3.18

```
npr08_08
Data are in X, Y, P
jcalc
---------------------
EX = X*PX'
EX = 10.1000
EY = Y*PY'
EY = 3.0016
EX2 = (X.^2)*PX'
EX2 = 133.0800
EY2 = (Y.^2)*PY'
EY2 = 41.5564
EXY = total(t.*u.*P)
EXY = 22.2890
```

Exercise 11.3.19

```
npr08_09
Data are in X, Y, P
jcalc
---------------------
EX = X*PX'
EX = 1.9250
EY = Y*PY'
EY = 2.8050
EX2 = (X.^2)*PX'
EX2 = 4.0375
EY2 = (Y.^2)*PY'           EXY = total(t.*u.*P)
EY2 = 8.9850               EXY = 5.1410
```

Exercise 11.3.20

$$E[X] = \int_0^1 2t(1-t)\,dt = 1/3,\ E[Y] = 2/3,\ E[X^2] = 1/6,\ E[Y^2] = 2/3$$
$$E[XY] = \int_0^1 \int_0^{2(1-t)} tu\,dudt = 1/6$$

```
tuappr: [0 1] [0 2] 200 400  u<=2*(1-t)
EX = 0.3333    EY = 0.6667    EX2 = 0.1667    EY2 = 0.6667
EXY = 0.1667 (use t, u, P)
```

Exercise 11.3.21

$$E[X] = E[Y] = \int_0^1 t^2\,dt + \int_1^t (2t - t^2)\,dt = 1,\ E[X^2] = E[Y^2] = 7/6$$
$$E[XY] = (1/2)\int_0^1 \int_{1-t}^{1+t} dtdt + (1/2)\int_1^2 \int_{t-1}^{3-t} dudt = 1$$

```
tuappr: [0 2] [0 2] 200 200  0.5*(u<=min(t+1,3-t))&(u>=max(1-t,t-1))
EX = 1.0000    EY = 1.0002    EX2 = 1.1684    EY2 = 1.1687    EXY = 1.0002
```

Exercise 11.3.22

$$E[X] = 2/3,\ E[Y] = 1/3,\ E[X^2] = 1/2,\ E[Y^2] = 1/6,\ E[XY] = 2/9$$

```
tuappr: [0 1] [0 1] 200 200  4*t.*(1-u)
EX = 0.6667    EY = 0.3333    EX2 = 0.5000    EY2 = 0.1667    EXY = 0.2222
```

Exercise 11.3.23

Extra close brace or missing open brace , $E[X^2] = E[Y^2] = 5/3$

$$E[XY] = \frac{1}{8}\int_0^2 \int_0^2 (t^2u + tu^2)\,dudt = \frac{4}{3}$$

```
tuappr: [0 1] [0 1] 200 200  4*t.*(1-u)
EX = 1.1667    EY = 1.1667    EX2 = 1.6667    EY2 = 1.6667    EXY = 1.3333
```

Exercise 11.3.24

$$E[X] = \int_0^\infty 2te^{-2t}\,dt = \frac{1}{2},\ E[Y] = \frac{2}{3},\ E[X^2] = \frac{1}{2},\ E[Y^2] = \frac{1}{2},\ E[XY] = \frac{1}{3}$$

```
tuappr: [0 6] [0 1] 600 200  4*u.*exp(-2*t)
EX = 0.5000    EY = 0.6667    EX2 = 0.4998    EY2 = 0.5000    EXY = 0.3333
```

Exercise 11.3.25

$$E[X] = \frac{313}{220},\ E[Y] = \frac{1429}{880},\ E[X^2] = \frac{49}{22},\ E[Y^2] = \frac{172}{55},\ E[XY] = \frac{2153}{880}$$

```
tuappr: [0 2] [0 3] 200 300  (3/88)*(2*t + 3*u.^2).*(u<1+t)
EX = 1.4229    EY = 1.6202    EX2 = 2.2277    EY2 = 3.1141    EXY = 2.4415
```

Exercise 11.3.26

$$E[X] = \frac{2}{5},\ E[Y] = \frac{11}{15},\ E[X^2] = \frac{2}{5},\ E[Y^2] = \frac{3}{5},\ E[XY] = \frac{2}{5}$$

```
tuappr: [-1 1] [0 1] 400 300  12*t.^2.*u.*(u>=max(0,t)).*(u<=min(1+t,1))
EX = 0.4035    EY = 0.7342    EX2 = 0.4016    EY2 = 0.6009    EXY = 0.4021
```

Exercise 11.3.27

$$E[X] = \frac{52}{55},\ E[Y] = \frac{32}{55},\ E[X^2] = \frac{57}{55},\ E[Y^2] = \frac{2}{5},\ E[XY] = \frac{28}{55}$$

```
tuappr: [0 2] [0 1] 400 200  (24/11)*t.*u.*(u<=min(1,2-t))
EX = 0.9458    EY = 0.5822    EX2 = 1.0368    EY2 = 0.4004    EXY = 0.5098
```

Exercise 11.3.28

$$E[X] = \frac{53}{46},\ E[Y] = \frac{22}{23},\ E[X^2] = \frac{397}{230},\ E[Y^2] = \frac{261}{230},\ E[XY] = \frac{251}{230}$$

```
tuappr: [0 2] [0 2] 200 200  (3/23)*(t + 2*u).*(u<=max(2-t,t))
EX = 1.1518    EY = 0.9596    EX2 = 1.7251    EY2 = 1.1417    EXY = 1.0944
```

Exercise 11.3.29

$$E[X] = \frac{2313}{1790},\ E[Y] = \frac{778}{895},\ E[X^2] = \frac{1711}{895},\ E[Y^2] = \frac{916}{895},\ E[XY] = \frac{1811}{1790}$$

```
tuappr: [0 2] [0 2] 400 400  (12/179)*(3*t.^2 + u).*(u<=min(2,3-t))
EX = 1.2923    EY = 0.8695    EX2 = 1.9119    EY2 = 1.0239    EXY = 1.0122
```

Exercise 11.3.30

$$E[X] = \frac{1567}{1135},\ E[Y] = \frac{2491}{2270},\ E[X^2] = \frac{476}{227},\ E[Y^2] = \frac{1716}{1135},\ E[XY] = \frac{5261}{3405}$$

```
tuappr: [0 2] [0 2] 400 400  (12/227)*(3*t + 2*t.*u).*(u<=min(1+t,2))
EX = 1.3805    EY = 1.0974    EX2 = 2.0967    EY2 = 1.5120    EXY = 1.5450
```

Exercise 11.3.31

$$E[X] = \frac{16}{13},\ E[Y] = \frac{11}{12},\ E[X^2] = \frac{219}{130},\ E[Y^2] = \frac{83}{78},\ E[XY] = \frac{431}{390}$$

```
tuappr: [0 2] [0 2] 400 400  (2/13)*(t + 2*u).*(u<=min(2*t,3-t))
EX = 1.2309    EY = 0.9169    EX2 = 1.6849    EY2 = 1.0647    EXY = 1.1056
```

Exercise 11.3.32

$$E[X] = \frac{243}{224},\ E[Y] = \frac{11}{16},\ E[X^2] = \frac{107}{70},\ E[Y^2] = \frac{127}{240},\ E[XY] = \frac{347}{448}$$

```
tuappr: [0 2] [0 1] 400 200  (3/8)*(t.^2 + 2*u).*(t<=1) + (9/14)*(t.^2.*u.^2).*(t > 1)
EX = 1.0848    EY = 0.6875    EX2 = 1.5286    EY2 = 0.5292    EXY = 0.7745
```

Exercise 11.3.33

Use x and px to prevent renaming.

```
x = [-5 -1 3 4 7];
px = 0.01*[15 20 30 25 10];
icalc
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter X probabilities px
Enter Y probabilities px
 Use array operations on matrices X, Y, PX, PY, t, u, and P
 G = 3*t - 4*u
 [R,PR] = csort(G,P);
 icalc
Enter row matrix of X-values  R
Enter row matrix of Y-values  x
Enter X probabilities  PR
Enter Y probabilities  px
 Use array operations on matrices X, Y, PX, PY, t, u, and P
H = t + 2*u;
EH = total(H.*P)
EH = 1.6500
[W,PW] = csort(H,P);  % Alternate
EW = W*PW'
EW = 1.6500
icalc3                % Solution with icalc3
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter row matrix of Z-values  x
Enter X probabilities  px
Enter Y probabilities  px
Enter Z probabilities  px
Use array operations on matrices X, Y, Z,
PX, PY, PZ, t, u, v, and P
K = 3*t - 4*u + 2*v;
EK = total(K.*P)
EK = 1.6500
```

Exercise 11.3.34

```
X = 0:150;
PX = ipoisson(75, X);
G = 200 + 18*(X - 10).*(X>=10) + (16 - 18)*(X - 20).*(X>=20) + ...
      (15 - 16)*(X - 30).*(X>=30) + (13 - 15)*(X>=50);
[Z,PZ] = csort(G,PX);
EZ = Z*PZ'
EZ = 1.1650e+03
EZ2 = (Z.^2)*PZ'
EZ2 = 1/3699e+06
```

Exercise 11.3.35

```
npr08_07
Data are in X, Y, P
jcalc
------------------
G = 3*t.^2 + 2*t.*u - u.^2;
EG = total(G.*P)
EG = 5.2975
ez2 = total(G.^2.*P)
EG2 = 1.0868e+03
```

```
[Z,PZ] = csort(G,P);        % Alternate
EZ = Z*PZ'
EZ = 5.2975
EZ2 = (Z.^2)*PZ'
EZ2 = 1.0868e+03
```

Exercise 11.3.36

```
H = t.*(t+u<=4) + 2*u.*(t+u>4);
EH = total(H.*P)
EH = 4.7379
EH2 = total(H.^2.*P)
EH2 = 61.4351
[W,PW] = csort(H,P);    %Alternate
EW = W*PW'
EW = 4.7379
EW2 = (W.^2)*PW'
EW2 = 61.4351
```

Exercise 11.3.37

$$E[Z] = \frac{3}{88}\int_0^1\int_0^{1+t} 4t(2t+3u^2)\,dudt + \frac{3}{88}\int_1^2\int_0^{1+t}(t+u)(2t+3u^2)\,dudt = \frac{5649}{1760}$$

$$E[Z^2] = \frac{3}{88}\int_0^1\int_0^{1+t} (4t)^2(2t+3u^2)\,dudt + \frac{3}{88}\int_1^2\int_0^{1+t} (t+u)^2(2t+3u^2)\,dudt = \frac{4881}{440}$$

```
tuappr: [0 2] [0 3] 200 300 (3/88)*(2*t+3*u.^2).*(u<=1+t)
G = 4*t.*(t<=1) + (t + u).*(t>1);
EG = total(G.*P)
EG = 3.2086
EG2 = total(G.^2.*P)
EG2 = 11.0872
```

Exercise 11.3.38

$$E[Z] = \frac{12}{11}\int_0^1\int_t^1 t^2u\,dudt + \frac{24}{11}\int_0^1\int_0^t tu^3\,dudt + \frac{24}{11}\int_1^2\int_0^{2-t} tu^3\,dudt = \frac{16}{55}$$

$$E[Z^2] = \frac{6}{11}\int_0^1\int_t^1 t^3u\,dudt + \frac{24}{11}\int_0^1\int_0^t tu^5\,dudt + \frac{24}{11}\int_1^2\int_0^{2-t} tu^5\,dudt = \frac{39}{308}$$

```
tuappr: [0 2] [0 1] 400 200  (24/11)*t.*u.*(u<=min(1,2-t))
G = (1/2)*t.*(u>t) + u.^2.*(u<=t);
EZ = 0.2920 EZ2 = 0.1278
```

Exercise 11.3.39

$$E[Z] = \frac{3}{23}\int_0^1\int_0^1 (t+u)(t+2u)\,dudt + \frac{3}{23}\int_0^1\int_1^{2-t} 2u(1+2u)\,dudt + \frac{3}{23}\int_1^2\int_1^t 2u(t+2u)\,dudt = \frac{175}{92}$$

$$E[Z^2] = \frac{3}{23}\int_0^1\int_0^1 (t+u)^2(t+2u)\,dudt + \frac{3}{23}\int_0^1\int_1^{2-t} 4u^2(1+2u)\,dudt + \frac{3}{23}\int_1^2\int_1^t 4u^2(t+2u)\,dudt =$$

```
tuappr: [0 2] [0 2] 400 400  (3/23)*(t+2*u).*(u<=max(2-t,t))
M = max(t,u)<=1;
G = (t+u).*M + 2*u.*(1-M);
EZ = total(G.*P)
EZ = 1.9048
EZ2 = total(G.^2.*P)
EZ2 = 4.4963
```

Exercise 11.3.40

$$E[Z] = \frac{12}{179}\int_0^1\int_1^2 (t+u)(3t^2+u)\,dudt + \frac{12}{179}\int_0^1\int_0^1 2u^2(3t^2+u)\,dudt + \frac{12}{179}\int_1^2\int_0^{3-t} 2u^2(3t^2+u)\,dudt = \frac{1422}{895}$$

$$E[Z^2] = \frac{12}{179}\int_0^1\int_1^2 (t+u)^2(3t^2+u)\,dudt + \frac{12}{179}\int_0^1\int_0^1 4u^4(3t^2+u)\,dudt + \frac{12}{179}\int_1^2\int_0^{3-t} 4u^4(3t^2+u)\,dudt = \frac{28296}{6265}$$

```
tuappr: [0 2] [0 2] 400 400  (12/179)*(3*t.^2 + u).*(u <= min(2,3-t))
```

```
M = (t<=1)&(u>=1);
G = (t + u).*M + 2*u.^2.*(1 - M);
EZ = total(G.*P)
EZ = 1.5898
EZ2 = total(G.^2.*P)
EZ2 = 4.5224
```

Exercise 11.3.41

$$E[Z] = \frac{12}{227}\int_0^1\int_0^1 t(3t+2tu)\,dudt + \frac{12}{227}\int_1^2\int_0^{2-t} t(3t+2tu)\,dudt +$$

$$\frac{12}{227}\int_0^1\int_1^{1+t} tu(3t+2tu)\,dudt + \frac{12}{227}\int_1^2\int_{2-t}^2 tu(3t+2tu)\,dudt = \frac{5774}{3405}$$

$$E[Z^2] = \frac{56673}{15890}$$

```
tuappr: [0 2] [0 2] 400 400  (12/227)*(3*t + 2*t.*u).*(u <= min(1+t,2))
M = u <= min(1,2-t);
G = t.*M + t.*u.*(1 - M);
EZ = total(G.*P)
EZ = 1.6955
EZ2 = total(G.^2.*P)
EZ2 = 3.5659
```

Exercise 11.3.42

```
npr10_16
Data are in cx, pmx, cy, pmy, Z, PZ
[X,PX] = canonicf(cx,pmx);
[Y,PY] - canonicf(cy,pmy);
icalc3
input: X, Y, Z, PX, PY, PZ
-------------
Use array operations on matrices X, Y, Z.
PX, PY, PZ, t, u, v, and P
G = t.^2 + 3*t.*u.^2 - 3*v;
[W,PW] = csort(G,P);
EW = W*PW'
EW = -1.8673
EW2 = (W.^2)*PW'
EW2 = 426.8529
```

This page titled 11.3: Problems on Mathematical Expectation is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

12: VARIANCE, COVARIANCE, AND LINEAR REGRESSION

12.1: Variance
12.2: Covariance and the Correlation Coefficient
12.3: Linear Regression
12.4: Problems on Variance, Covariance, Linear Regression

This page titled 12: Variance, Covariance, and Linear Regression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Variance

In the treatment of the mathematical expection of a real random variable X, we note that the mean value locates the center of the probability mass distribution induced by X on the real line. In this unit, we examine how expectation may be used for further characterization of the distribution for X. In particular, we deal with the concept of *variance* and its square root the *standard deviation*. In subsequent units, we show how it may be used to characterize the distribution for a pair $\{X, Y\}$ considered jointly with the concepts *covariance*, and *linear regression*

Variance

Location of the center of mass for a distribution is important, but provides limited information. Two markedly different random variables may have the same mean value. It would be helpful to have a measure of the spread of the probability mass about the mean. Among the possibilities, the variance and its square root, the standard deviation, have been found particularly useful.

Definition: Variance & Standard Deviation

The *variance* of a random variable X is the mean square of its variation about the mean value:

$$\text{Var } [X] = \sigma_X^2 = E[(X - \mu_X)^2] \text{ where } \mu_X = E[X]$$

The *standard deviation* for *X* is the positive square root σ_X of the variance.

Remarks

- If $X(\omega)$ is the observed value of X, its variation from the mean is $X(\omega) - \mu_X$. The variance is the probability weighted average of the square of these variances.
- The square of the error treats positive and negative variations alike, and it weights large variations more heavily than smaller ones.
- As in the case of mean value, the variance is a property of the distribution, rather than of the random variable.
- We show below that the standard deviation is a “natural” measure of the variation from the mean.
- In the treatment of mathematical expectation, we show that

$$E[(X - c)^2] \text{ is a minimum off } c = E[X], \text{ in which case } E[(X - E[X])^2] = E[X^2] - E^2[X]$$

This shows that the mean value is the constant which best approximates the random variable, in the mean square sense.

Basic patterns for variance

Since variance is the expectation of a function of the random variable *X*, we utilize properties of expectation in computations. In addition, we find it expedient to identify several patterns for variance which are frequently useful in performing calculations. For one thing, while the variance is defined as $E[(X - \mu_X)^2]$, this is usually not the most convenient form for computation. The result quoted above gives an alternate expression.

(V1): *Calculating formula*. $\text{Var } [X] = E[X^2] - E^2[X]$
(V2): *Shift property*. $\text{Var } [X + b] = \text{Var } [X]$. Adding a constant b to X shifts the distribution (hence its center of mass) by that amount. The variation of the shifted distribution about the shifted center of mass is the same as the variation of the original, unshifted distribution about the original center of mass.
(V3): *Change of scale*. $\text{Var } [aX] = a^2 \text{Var } [X]$. Multiplication of X by constant a changes the scale by a factor $[a]$. The squares of the variations are multiplied by a^2. So also is the mean of the squares of the variations.
(V4): *Linear combinations*.
a. $\text{Var } [aX \pm bY] = a^2 \text{Var } [X] + b^2 \text{Var } [Y] \pm 2ab(E[XY] - E[X]E[Y])$
b. More generally,

$$\text{Var } [\textstyle\sum_{k=1}^n a_k X_k] = \sum_{k=1}^n a_k^2 \text{Var } [X_k] + 2 \sum_{i<j} a_i a_j (E[X_i X_j] - E[X_i]E[X_j])$$

The term $c_{ij} = E[X_iX_j] - E[X_i]E[X_j]$ is the covariance of the pair $\{X_i, X_j\}$, whose role we study in the unit on that topic. If the c_{ij} are all zero, we say the class is *uncorrelated*.

Remarks

- If the pair $\{X, Y\}$ is independent, it is uncorrelated. The converse is not true, as examples in the next section show.
- If the $a_i = \pm 1$ and all pairs are uncorrelated, then

$$\text{Var}\left[\sum_{k=1}^n a_iX_i\right] = \sum_{k=1}^n \text{Var}\left[X_i\right]$$

The variance add even if the coefficients are negative.

We calculate variances for some common distributions. Some details are omitted—usually details of algebraic manipulation or the straightforward evaluation of integrals. In some cases we use well known sums of infinite series or values of definite integrals. A number of pertinent facts are summarized in Appendix B. Some Mathematical Aids. The results below are included in the table in Appendix C.

Variances of some discrete distributions

Indicator function $X = I_E P(E) = p, q = 1 - p$ $E[X] = p$

$$E[X^2] - E^2[X] = E[I_E^2] - p^2 = E[I_E] - p^2 = p - p^2 = p(1-p) - pq$$

Simple random variable $X = \sum_{i=1}^n t_iI_{A_i}$ (primitive form) $P(A_i) = p_i$.

$\text{Var}\left[X\right] = \sum_{i=1}^n t_i^2p_iq_i - 2\sum_{i<j} t_it_jp_ip_j$, since $E[I_{A_i}I_{A_j}] = 0$ $i \neq j$

Binomial(n, p). $X = \sum_{i=1}^n I_{E_i}$ with $\{I_{E_i} : 1 \leq i \leq n\}$ iid $P(E_i) = p$

$$\text{Var}\left[X\right] = \sum_{i=1}^n \text{Var}\left[I_{E_i}\right] = \sum_{i=1}^n pq = npq$$

Geometric(p). $P(X = k) = pq^k$ $\forall k \geq 0$ $E[X] = q/p$

We use a trick: $E[X^2] = E[X(X-1)] + E[X]$

$$E[X^2] = p\sum_{k=0}^\infty k(k-1)q^k + q/p = pq^2\sum_{k=2}^\infty k(k-1)q^{k-2} + q/p = pq^2\frac{2}{(1-q)^3} + q/p = 2\frac{q^2}{p^2} + q/p$$

$$\text{Var}\left[X\right] = 2\frac{q^2}{p^2} + q/p - (q/p)^2 = q/p^2$$

Poisson(\mu) $P(X = k) = e^{-\mu}\dfrac{\mu^k}{k!}$ $\forall k \geq 0$

Using $E[X^2] = E[X(X-1)] + E[X]$, we have

$$E[X^2] = e^{-\mu}\sum_{k=2}^\infty k(k-1)\frac{\mu^k}{k!} + \mu = e^{-\mu}\mu^2\sum_{k=2}^\infty \frac{\mu^{k-2}}{(k-2)!} + \mu = \mu^2 + \mu$$

Thus, $\text{Var}\left[X\right] = \mu^2 + \mu - \mu^2 = \mu$. Note that both the mean and the variance have common value μ

Some absolutely continuous distributions

Uniform on (a, b) $f_X(t) = \dfrac{1}{b-a}$ $a < t < b$ $E[X] = \dfrac{a+b}{2}$

$$E[X^2] = \frac{1}{b-a}\int_a^b t^2\,dt = \frac{b^3-a^3}{3(b-a)} \text{ so } \text{Var}\left[X\right] = \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Symmetric triangular (a, b) Because of the shift property (V2), we may center the distribution at the origin. Then the distribution is symmetric triangular $(-c, c)$, where $c = (b-a)/2$. Because of the symmetry

$$\text{Var}\ [X] = E[X^2] = \int_{-c}^{c} t^2 f_X(t)\ dt = 2\int_0^c t^2 f_X(t)\ dt$$

Now, in this case,

$$f_X(t) = \frac{c-t}{c^2}\ 0 \le t \le c \text{ so that } E[X^2] = \frac{2}{c^2}\int_0^c (ct^2 - t^3)\ dt = \frac{c^3}{6} = \frac{(b-a)^2}{24}$$

Exponential (\lambda) $f_X(t) = \lambda e^{-\lambda t}$, $t \ge 0$ $E[X] = 1/\lambda$

$$E[X^2] = \int_0^\infty \lambda t^2 e^{-\lambda t}\ dt = \frac{2}{\lambda^2} \text{ so that } \text{Var}\ [X] = 1/lambda^2$$

Gamma(α, λ) $f_X(t) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha t^{\alpha=1} e^{-\lambda t}$ $t \ge 0$ $E[X] = \frac{\alpha}{\lambda}$

$$E[X^2] = \frac{1}{\Gamma(\alpha)}\int_0^\infty \lambda^\alpha t^{\alpha+1} e^{-\lambda t}\ dt = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{lambda^2}$$

Hence $\text{Var}\ [X] = \alpha/\lambda^2$.

Normal(μ, σ^2) $E[X] = \mu$

Consider $Y \sim N(0,1)$, $E[Y] = 0$, $\text{Var}\ [Y] = \frac{2}{\sqrt{2\pi}}\int_0^\infty t^2 e^{-t^2/2}\ dt = 1$.

$$X = \sigma Y + \mu \text{ implies } \text{Var}\ [Y] = \sigma^2$$

Extensions of some previous examples

In the unit on expectations, we calculate the mean for a variety of cases. We revisit some of those examples and calculate the variances.

Example 12.1.1 Expected winnings (Example 8 from "Mathematical Expectation: Simple Random Variables")

A bettor places three bets at $2.00 each. The first pays $10.00 with probability 0.15, the second $8.00 with probability 0.20, and the third $20.00 with probability 0.10.

Solution

The net gain may be expressed

$$X = 10I_A + 8I_B + 20I_C - 6, \text{ with } P(A) = 0.15, P(B) = 0.20, P(C) = 0.10$$

We may reasonbly suppose the class $\{A, B, C\}$ is independent (this assumption is not necessary in computing the mean). Then

$$\text{Var}\ [X] = 10^2 P(A)[1 - P(A)] + 8^2 P(B)[1 - P(B)] + 20^2 P(C)[1 - P(C)]$$

Calculation is straightforward. We may use MATLAB to perform the arithmetic.

```
c = [10 8 20];
p = 0.01*[15 20 10];
q = 1 - p;
VX = sum(c.^2.*p.*q)
VX = 58.9900
```

Example 12.1.2 A function of X (Example 9 from "Mathematical Expectation: Simple Random Variables")

Suppose X in a primitive form is

$$X = -3I_{C_1} - I_{C_2} + 2I_{C_3} - 3I_{C_4} + 4I_{C_5} - I_{C_6} + I_{C_7} + 2I_{C_8} + 3I_{C_9} + 2I_{C_{10}}$$

with probabilities $P(C_i) = 0.08, 0.11, 0.06, 0.13, 0.05, 0.08, 0.12, 0.07, 0.14, 0.16$.

Let $g(t) = t^2 + 2t$. Determine $E[g(X)]$ and $\text{Var}[g(X)]$

```
c = [-3 -1 2 -3 4 -1 1 2 3 2];            % Original coefficients
pc = 0.01*[8 11 6 13 5 8 12 7 14 16];     % Probabilities for c_j
G = c.^2 + 2*c                            % g(c_j)
EG = G*pc'                                % Direct calculation E[g(X)]
EG = 6.4200
VG = (G.^2)*pc' - EG^2;                   % Direct calculation Var[g(X)]
VG = 40.8036
[Z,PZ] = csort(G,pc);                     % Distribution for Z = g(X)
EZ = Z*PZ'                                % E[Z]
EZ = 6.4200
VZ = (Z.^2)*PZ' - EZ^2                    % Var[Z]
VZ = 40.8036
```

Example 12.1.3 $Z = g(X, Y)$ (Example 10 from "Mathematical Expectation: Simple Random Variables")

We use the same joint distribution as for Example 10 from "Mathematical Expectation: Simple Random Variables" and let $g(t, u) = t^2 + 2tu - 3u$. To set up for calculations, we use jcalc.

```
jdemo1                          % Call for data
jcalc                           % Set up
Enter JOINT PROBABILITIES (as on the plane)    P
Enter row matrix of VALUES of X    X
Enter row matrix of VALUES of Y    Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = t.^2 + 2*t.*u - 3*u;        % calcculation of matrix of [g(t_i, u_j)]
EG = total(G.*P)                % Direct calculation of E[g(X,Y)]
EG = 3.2529
VG = total(G.^.*P) - EG^2       % Direct calculation of Var[g(X,Y)]
VG = 80.2133
[Z,PZ] = csort(G,P);            % Determination of distribution for Z
EZ = Z*PZ'                      % E[Z] from distribution
EZ = 3.2529
VZ = (Z.^2)*PZ' - EZ^2          % Var[Z] from distribution
VZ = 80.2133
```

Example 12.1.4 A function with compound definition (Example 12 from "Mathematical Expectation: Simple Random Variables")

Suppose \(X) ~ exponential (0.3). Let

$$Z = \begin{cases} X^2 & \text{for } X \le 4 \\ 16 & \text{for } X > 4 \end{cases} = I_{[0,4]}(X)X^2 + I_{(4,\infty]}(X)16$$

Determine $E[Z]$ and $Var[Z]$.

Analytic Solution

$$E[g(X)] = \int g(t) f_X(t)\, dt = \int_0^\infty I_{[0,4]}(t) t^2 0.3 e^{-0.3t}\, dt + 16E[I_{(4,\infty]}(X)]$$

$$= \int_0^4 t^2 0.3 e^{-0.3t}\, dt + 16P(X > 4) \approx 7.4972 \text{ (by Maple)}$$

$$Z^2 - I_{[0,4]}(X)X^4 + I_{(4,\infty]}(X)256$$

$$E[Z^2] = \int_0^\infty I_{[0,4]}(t) t^4 0.3 e^{-0.3t}\, dt + 256E[I_{(4,\infty]}(X)] = \int_0^4 t^4 0.3 e^{-0.3t}\, dt + 256e^{-1.2} \approx 100.0562$$

$$\text{Var}\,[Z] = E[Z^2] - E^2[Z] \approx 43.8486 \text{ (by Maple)}$$

APPROXIMATION

To obtain a simple aproximation, we must approximate by a bounded random variable. Since $P(X > 50) = e^{-15} \approx 3 \cdot 10^{-7}$ we may safely truncate X at 50.

```
tuappr
Enter matrix [a b] of x-range endpoints  [0 50]
Enter number of x approximation points  1000
Enter density as a function of t  0.3*exp(-0.3*t)
Use row matrices X and PX as in the simple case
M = X <= 4;
G = M.*X.^2 + 16*(1 - M);   % g(X)
EG = G*PX'                  % E[g(X)]
EG = 7.4972
VG = (G.^2)*PX' - EG^2      % Var[g(X)]
VG = 43.8472                % Theoretical = 43.8486
[Z,PZ] = csort(G,PX);       % Distribution for Z = g(X)
EZ = Z*PZ'                  % E[Z] from distribution
EZ = 7.4972
VZ = (Z.^2)*PZ' - EZ^2      % Var[Z]
VZ = 43.8472
```

Example 12.1.5 Stocking for random demand (Example 13 from "Mathematical Expectation: Simple Random Variables")

The manager of a department store is planning for the holiday season. A certain item costs c dollars per unit and sells for p dollars per unit. If the demand exceeds the amount m ordered, additional units can be special ordered for s dollars per unit ($s > c$). If demand is less than the amount ordered, the remaining stock can be returned (or otherwise disposed of) at r dollars per unit ($r < c$). Demand D for the season is assumed to be a random variable with Poisson (μ) distribution. Suppose $\mu = 50$, $c = 30$, $p = 50$, $s = 40$, $r = 20$. What amount m should the manager order to maximize the expected profit?

Problem Formulation

Suppose D is the demand and X is the profit. Then

For $D \le m$, $X = D(p - c) - (m - D)(c - r) = D(p - r) + m(r - c)$
For $D > m$, $X = m(p - c) - (D - m)(p - s) = D(p - s) + m(s - c)$

It is convenient to write the expression for X in terms of I_M, where $M = (-\infty, M]$. Thus

$$X = I_M(D)[D(p - r) + m(r - c)] + [1 - I_M(D)][D(p - s) + m(s - c)]$$

$$= D(p-s) + m(s-c) + I_M(D)[D(p-r) + m(r-c) - D(p-s) - m(s-c)]$$

$$D(p-s) + m(s-c) + I_M(D)(s-r)[D-m]$$

Then

$$E[X] = (p-s)E[D] + m(s-c) + (s-r)E[I_M(D)D] - (s-r)mE[I_M(D)]$$

We use the discrete approximation.

APPROXIMATION

```
>> mu = 50;
>> n = 100;
>> t = 0:n;
>> pD = ipoisson(mu,t);            % Approximate distribution for D
>> c = 30;
>> p = 50;
>> s = 40;
>> r = 20;
>> m = 45:55;
>> for i = 1:length(m)             % Step by step calculation for various m
    M = t<=m(i);
    G(i,:) = (p-s)*t + m(i)*(s-c) + (s-r)*M.*(t - m(i));
end
>> EG = G*pD';
>> VG = (G.^2)*pD' - EG.^2;
>> SG = sqrt(VG);
>> disp([EG';VG';SG']')
   1.0e+04 *
    0.0931    1.1561    0.0108
    0.0936    1.3117    0.0115
    0.0939    1.4869    0.0122
    0.0942    1.6799    0.0130
    0.0943    1.8880    0.0137
    0.0944    2.1075    0.0145
    0.0943    2.3343    0.0153
    0.0941    2.5637    0.0160
    0.0938    2.7908    0.0167
    0.0934    3.0112    0.0174
    0.0929    3.2206    0.0179
```

Example 12.1.6 A jointly distributed pair (Example 14 from "Mathematical Expectation: Simple Random Variables")

Suppose the pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = 3u$ on the triangular region bounded by $u = 0$, $u = 1 + t$, $u = 1 - t$. Let $Z = g(X,Y) = X^2 + 2XY$.

Determine $E[Z]$ and $\text{Var}\,[Z]$.

Analytic Solution

$$E[Z] = \int\int (t^2 + 2tu) f_{XY}(t,u)\; dudt = 3\int_{-1}^{0}\int_{0}^{1+t} u(t^2+2tu)\; dudt + 3\int_{0}^{1}\int_{0}^{1-t} u(t^2+2tu)\; dudt = 1/10$$

$$E[Z^2] = 3\int_{-1}^{0}\int_{0}^{1+t} u(t^2+2tu)^2\; dudt + 3\int_{0}^{1}\int_{0}^{1-t} u(t^2+2tu)^2\; dudt = 3/35$$

$$\text{Var}\,[Z] = E[Z^2] - E^2[Z] = 53/700 \approx 0.0757$$

APPROXIMATION

```
tuappr
Enter matrix [a b] of x-range endpoints  [-1 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  400
Enter number of Y approximation points  200
Enter expression for joint density 3*u.*(u<=min(1+t,1-t))
Use array operations on X, Y, PX, PY, t, u, and P
G = t.^2 + 2*t.*u;            % g(X,Y) = X^2 + 2XY
EG = total(G.*P)              % E[g(X,Y)]
EG = 0.1006                   % Theoretical value = 1/10
VG = total(G.^2.*P) - EG^2
VG = 0.0765                   % Theoretical value 53/700 = 0.757
[Z,PZ] = csort(G,P);          % Distribution for Z
EZ = Z*PZ'                    % E[Z] from distribution
EZ = 0.1006
VZ = (Z.^2)*PZ' - EZ^2
VZ = 0.0765
```

Example 12.1.7 A function with compound definition (Example 15 from "Mathematical Expectation: Simple Random Variables")

The pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = 1/2$ on the square region bounded by $u = 1 + t$, $u = 1 - t$, $u = 3 - t$, and $u = t - 1$.

$$W = \begin{cases} X & \text{for } \max\{X,Y\} \le 1 \\ 2Y & \text{for } \max\{X,Y\} > 1 \end{cases} = I_Q(X,Y)X + I_{Q^c}(X,Y)2Y$$

where $Q = \{(t,u) : \max\{t,u\} \le 1\} = \{(t,u) : t \le 1, u \le 1\}$.

Determine $E[W]$ and $\text{Var}\,[W]$.

Solution

The intersection of the region Q and the square is the set for which $0 \le t \le 1$ and $1 - t \le u \le 1$. Reference to Figure 11.3.2 shows three regions of integration.

$$E[W] = \frac{1}{2}\int_0^1\int_{1-t}^1 t\,du\,dt + \frac{1}{2}\int_0^1\int_1^{1+t} 2u\,du\,dt + \frac{1}{2}\int_1^2\int_{t-1}^{3-t} 2u\,du\,dt = 11/6 \approx 1.8333$$

$$E[W^2] = \frac{1}{2}\int_0^1\int_{1-t}^1 t^2\,du\,dt + \frac{1}{2}\int_0^1\int_1^{1+t} 4u^2\,du\,dt + \frac{1}{2}\int_1^2\int_{t-1}^{3-t} 4u^2\,du\,dt = 103/24$$

$$\text{Var}\,[W] = 103/24 - (11/6)^2 = 67/72 \approx 0.9306$$

```
tuappr
Enter matrix [a b] of x-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density ((u<=min(t+1,3-t))& ...
        (u$gt;=max(1-t,t-1))/2
Use array operations on X, Y, PX, PY, t, u, and P
M = max(t,u)<=1;
```

```
G = t.^M + 2*u.*(1 - M);        %Z = g(X,Y)
EG = total(G.*P)                % E[g(X,Y)]
EG = 1.8349=0                   % Theoretical 11/6 = 1.8333
VG = total(G.^2.*P) - EG^2
VG = 0.9368                     % Theoretical 67/72 = 0.9306
[Z,PZ] = csort(G,P);            % Distribution for Z
EZ = Z*PZ'                      % E[Z] from distribution
EZ = 1.8340
VZ = (Z.^2)*PZ' - EZ^2
VZ = 0.9368
```

Example 12.1.8 A function with compound definition

$f_{XY}(t,u) = 3$ on $0 \le u \le t^2 \le 1$

$Z = I_Q(X,Y)X + I_{Q^c}(X,Y)$ for $Q = \{(t,u) : u + t \le 1\}$

The value t_0 where the line $u = 1 - t$ and the curve $u = t^2$ meet satisfies $t_0^2 = 1 - t_0$.

$$E[Z] = 3\int_0^{t_0} t \int_0^{t^2} \, dudt + 3\int_{t_0}^1 t \int_0^{1-t} \, dudt + 3\int_{t_0}^1 \int_{1-t}^{t^2} \, dudt = \frac{3}{4}(5t_0 - 2)$$

For $E[Z^2]$ replace t by t^2 in the integrands to get $E[Z^2] = (25t_0 - 1)/20$.

Using $t_0 = (\sqrt{5} - 1)/2 \approx 0.6180$, we get $\text{Var}\,[Z] = (2125t_0 - 1309)/80 \approx 0.0540$.

APPROXIMATION

```
% Theoretical values
t0 = (sqrt(5) - 1)/2
t0 = 0.6180
EZ = (3/4)*(5*t0 - 2)
EZ = 0.8176
EZ2 = (25*t0 - 1)/20
EZ2 = 0.7225
VZ = (2125*T0 - 1309)/80
VZ = 0.0540
tuappr
Enter matrix [a b] of x-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density 3*(u <= t.^2)
Use array operations on X, Y, t, u, and P
G = (t+u <= 1).*t + (t+u > 1);
EG = total(G.*P)
EG = 0.8169                     % Theoretical = 0.8176
VG = total(G.^2.*P) - EG^2
VG = 0.0540                     % Theoretical = 0.0540
[Z,PZ] = csort(G,P);
EZ = Z*PZ'
EZ = 0.8169
VZ = (Z.^2)*PZ' - EZ^2
VZ = 0.0540
```

Standard deviation and the Chebyshev inequality

In Example 5 from "Functions of a Random Variable," we show that if $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$. Also, $E[X] = \mu$ and $\text{Var}\,[X] = \sigma^2$. Thus

$$P(\frac{|X-\mu|}{\sigma} \le t) = P(|X-\mu| \le t\sigma) = 2\phi(t) - 1$$

For the normal distribution, the standard deviation σ seems to be a natural measure of the variation away from the mean.

For a general distribution with mean μ and variance σ^2, we have the

Chebyshev inequality

$$P(\frac{|X-\mu|}{\sigma} \ge a) \le \frac{1}{a^2} \text{ or } P(|X-\mu| \ge a\sigma) \le \frac{1}{a^2}$$

In this general case, the standard deviation appears as a measure of the variation from the mean value. This inequality is useful in many theoretical applications as well as some practical ones. However, since it must hold for any distribution which has a variance, the bound is not a particularly tight. It may be instructive to compare the bound on the probability given by the Chebyshev inequality with the actual probability for the normal distribution.

```
t = 1:0.5:3;
p = 2*(1 - gaussion(0.1,t));
c = ones(1,length(t))./(t.^2);
r = c./p;
h = ['    t    Chebyshev    Prob    Ratio'];
m = [t;c;p;r]';
disp(h)
        t    Chebyshev    Prob    Ratio
disp(m)
    1.0000    1.0000    0.3173    3.1515
    1.5000    0.4444    0.1336    3.3263
    2.0000    0.2500    0.0455    5.4945
    2.5000    0.1600    0.0124   12.8831
    3.0000    0.1111    0.0027   41.1554
```

— □

DERIVATION OF THE CHEBYSHEV INEQUALITY

Let $A = \{|X-\mu| \ge a\sigma\} = \{(X-\mu)^2 \ge a^2\sigma^2\}$. Then $a^2\sigma^2 I_A \le (X-\mu)^2$.

Upon taking expectations of both sides and using monotonicity, we have

$$a^2\sigma^2 P(A) \le E[(X-\mu)^2] = \sigma^2$$

from which the Chebyshev inequality follows immediately.

— □

We consider three concepts which are useful in many situations.

Definition

A random variable X is *centered* iff $E[X] = 0$.

$X' = X - \mu$ is always centered.

Definition

A random variable X is *standardized* iff $E[X] = 0$ and $\text{Var}[X] = 1$.

$$X^* = \frac{X - \mu}{\sigma} = \frac{X'}{\sigma} \text{ is standardized}$$

Definition

A pair $\{X, Y\}$ of random variables is *uncorrelated* iff

$$E[XY] - E[X]E[Y] = 0$$

It is always possible to derive an uncorrelated pair as a function of a pair $\{X, Y\}$, both of which have finite variances. Consider

$$U = (X^* + Y^*)\ V = (X^* - Y^*), \text{ where } X^* = \frac{X - \mu_X}{\sigma_X},\ Y^* = \frac{Y - \mu_Y}{\sigma_Y}$$

Now $E[U] = E[V] = 0$ and

$$E[UV] = E(X^* + Y^*)(X^* - Y^*)] = E[(X^*)^2] - E[(Y^*)^2] = 1 - 1 = 0$$

so the pair is uncorrelated.

Example 12.1.9 Determining an unvorrelated pair

We use the distribution for Examples Example 10 from "Mathematical Expectation: Simple Random Variables" and Example, for which

$$E[XY] - E[X]E[Y] \neq 0$$

```
jdemo1
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
EX = total(t.*P)
EX = 0.6420
EY = total(u.*P)
EY = 0.0783
EXY = total(t.*u.*P)
EXY = -0.1130
c = EXY - EX*EY
c = -0.1633               % {X, Y} not uncorrelated
```

```
VX = total(t.^2.*P) - EX^2
VX = 3.3016
VY = total(u.^2.*P) - EY^2
VY = 3.6566
SX = sqrt(VX)
SX = 1.8170
SY = sqrt(VY)
SY = 1.9122
```

```
x = (t - EX)/SX;           % Standardized random variables
y = (u - EY)/SY;
uu = x + y;                % Uncorrelated random variables
vv = x - y;
EUV = total(uu.*vv.*P)     % Check for uncorrelated condition
EUV = 9.9755e-06           % Differs from zero because of roundoff
```

This page titled 12.1: Variance is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Covariance and the Correlation Coefficient

The mean value $\mu_X = E[X]$ and the variance $\sigma_X^2 = E[(X - \mu_X)^2]$ give important information about the distribution for real random variable X. Can the expectation of an appropriate function of (X, Y) give useful information about the joint distribution? A clue to one possibility is given in the expression

$$\text{Var}[X \pm Y] = \text{Var}[X] + \text{Var}[Y] \pm 2(E[XY] - E[X]E[Y])$$

The expression $E[XY] - E[X]E[Y]$ vanishes if the pair is independent (and in some other cases). We note also that for $\mu_X = E[X]$ and $\mu_Y = E[Y]$

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X\mu_Y$$

To see this, expand the expression $(X - \mu_X)(Y - \mu_Y)$ and use linearity to get

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y] = E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X\mu_Y$$

which reduces directly to the desired expression. Now for given ω, $X(\omega) - \mu_X$ is the variation of X from its mean and $Y(\omega) - \mu_Y$ is the variation of Y from its mean. For this reason, the following terminology is used.

Definition: Covariance

The quantity $\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$ is called the *covariance* of X and Y.

If we let $X' = X - \mu_X$ and $Y' = Y - \mu_Y$ be the ventered random variables, then

$$\text{Cov}[X, Y] = E[X'Y']$$

Note that the variance of X is the covariance of X with itself.

If we standardize, with $X^* = (X - \mu_X)/\sigma_X$ and $Y^* = (Y - \mu_Y)/\sigma_Y$, we have

Definition: Correlation Coefficient

The *correlation coefficient* $\rho = \rho[X, Y]$ is the quantity

$$\rho[X, Y] = E[X^*Y^*] = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X\sigma_Y}$$

Thus $\rho = \text{Cov}[X, Y]/\sigma_X\sigma_Y$. We examine these concepts for information on the joint distribution. By Schwarz' inequality (E15), we have

$\rho^2 = E^2[X^*Y^*] \le E[(X^*)^2]E[(Y^*)^2] = 1$ with equality iff $Y^* = cX^*$

Now equality holds iff

$1 = c^2E^2[(X^*)^2] = c^2$ which implies $c = \pm 1$ and $\rho = \pm 1$

We conclude $-1 \le \rho \le 1$, with $\rho = \pm 1$ iff $Y^* = \pm X^*$

Relationship between ρ and the joint distribution

- We consider first the distribution for the standardized pair (X^*, Y^*)
- Since $P(X^* \le r, Y^* \le s) = P(\frac{X - \mu_X}{\sigma_X} \le r, \frac{Y - \mu_Y}{\sigma_Y} \le s)$

$= P(X \le t = \sigma_X r + \mu_X, Y \le u = \sigma_Y s + \mu_Y)$

we obtain the results for the distribution for (X, Y) by the mapping

$$t = \sigma_X r + \mu_X$$
$$u = \sigma_Y s + \mu_Y$$

Joint distribution for the standardized variables (X^*, Y^*), $(r, s) = (X^*, Y^*)(\omega)$

$\rho = 1$ iff $X^* = Y^*$ iff all probability mass is on the line $s = r$.
$\rho = -1$ iff $X^* = -Y^*$ iff all probability mass is on the line $s = -r$.

If $-1 < \rho < 1$, then at least some of the mass must fail to be on these lines.

Figure one is comprised of a diagonal line with a right triangle. A portion of the line is the base of the triangle. The line is labeled, s = r. One point of the triangle located on the diagonal line is labeled (r, r). The point of the triangle that is not located on the line is labeled, (r, s). The side of the triangle in between these two labeled points is labeled as the absolute value of s - r. The side of the triangle on the line is not labeled. The third side is labeled as the absolute value of s - r divided by the square root of two.

Figure 12.2.1. Distance from point (r, s) to the line $s = r$.

The $\rho = \pm 1$ lines for the (X, Y) distribution are:

$$\frac{u - \mu_Y}{\sigma_Y} = \pm \frac{t - \mu_X}{\sigma_X} \text{ or } u = \pm \frac{\sigma_Y}{\sigma_X}(t - \mu_X) + \mu_Y$$

Consider $Z = Y^* - X^*$. Then $E[\frac{1}{2}Z^2] = \frac{1}{2}E[(Y^* - X^*)^2]$. Reference to Figure 12.2.1 shows this is the average of the square of the distances of the points $(r, s) = (X^*, Y^*)(\omega)$ from the line $s = r$ (i.e. the variance about the line $s = r$). Similarly for $W = Y^* + X^*$. $E[W^2/2]$ is the variance about $s = -r$. Now

$$\frac{1}{2}E[(Y^* \pm X^*)^2] = \frac{1}{2}\{E[(Y^*)^2] + E[(X^*)^2] \pm 2E[X^*Y^*]\} = 1 \pm \rho$$

Thus

$1 - \rho$ is the variance about $s = r$ (the $\rho = 1$ line)
$1 + \rho$ is the variance about $s = -r$ (the $\rho = -1$ line)

Now since

$$E[(Y^* - X^*)^2] = E[(Y^* + X^*)^2] \text{ iff } \rho = E[X^*Y^*] = 0$$

the condition $\rho = 0$ is the condition for equality of the two variances.

Transformation to the (X, Y) **plane**

$$t = \sigma_X r + \mu_X \; u = \sigma_Y s + \mu_Y \; r = \frac{t - \mu_X}{\sigma_X} \; s = \frac{u - \mu_Y}{\sigma_Y}$$

The $\rho = 1$ line is:

$$\frac{u - \mu_Y}{\sigma_Y} = \frac{t - \mu_X}{\sigma_X} \text{ or } u = \frac{\sigma_Y}{\sigma_X}(t - \mu_X) + \mu_Y$$

The $\rho = -1$ line is:

$$\frac{u - \mu_Y}{\sigma_Y} = \frac{t - \mu_X}{\sigma_X} \text{ or } u = -\frac{\sigma_Y}{\sigma_X}(t - \mu_X) + \mu_Y$$

$1 - \rho$ is proportional to the variance abut the $\rho = 1$ line and $1 + \rho$ is proportional to the variance about the $\rho = -1$ line. $\rho = 0$ iff the variances about both are the same.

Example 12.2.1 Uncorrelated but not independent

Suppose the joint density for $\{X, Y\}$ is constant on the unit circle about the origin. By the rectangle test, the pair cannot be independent. By symmetry, the $\rho = 1$ line is $u = t$ and the $\rho = -1$ line is $u = -t$. By symmetry, also, the variance about each of these lines is the same. Thus $\rho = 0$, which is true iff $\text{Cov}[X, Y] = 0$. This fact can be verified by calculation, if desired.

Example 12.2.2 Uniform marginal distributions

Figure two is comprised of three graphs of multiple shaded squares. All three are standard cartesian graphs, with all four quadrants equal in size, t as the horizontal axis, and u as the vertical axis. The first graph shows one large square centered at the origin with a length of two units on a side. As the square is centered about the origin, the square is divided equally into four smaller squares by the vertical and horizontal axes. A caption below the first graph reads, rho = 0. The second graph contains two smaller squares, one unit to a side, one sitting with two sides along the axes of the graph in the first quadrant, and one sitting with two sides along the axes of the graph in the third quadrant. The caption reads rho = 3/4. The third graph contains two squares of the same size as the second graph, this time with one sitting with two sides along the axes in the second quadrant, and one sitting with two sides along the axes in the fourth quadrant. The caption reads rho = -3/4.
Figure 12.2.2. Uniform marginals but different correlation coefficients.

Consider the three distributions in Figure 12.2.2. In case (a), the distribution is uniform over the square centered at the origin with vertices at (1,1), (-1,1), (-1,-1), (1,-1). In case (b), the distribution is uniform over two squares, in the first and third quadrants with vertices (0,0), (1,0), (1,1), (0,1) and (0,0),

(-1,0), (-1,-1), (0,-1). In case (c) the two squares are in the second and fourth quadrants. The marginals are uniform on (-1,1) in each case, so that in each case

$$E[X] = E[Y] = 0 \text{ and } \text{Var}[X] = \text{Var}[Y] = 1/3$$

This means the $\rho = 1$ line is $u = t$ and the $\rho = -1$ line is $u = -t$.

a. By symmetry, $E[XY] = 0$ (in fact the pair is independent) and $\rho = 0$.
b. For every pair of possible values, the two signs must be the same, so $E[XY] > 0$ which implies $\rho > 0$. The actual value may be calculated to give $\rho = 3/4$. Since $1 - \rho < 1 + \rho$, the variance about the $\rho = 1$ line is less than that about the $\rho = -1$ line. This is evident from the figure.
c. $E[XY] < 0$ and $\rho < 0$. Since $1 + \rho < 1 - \rho$, the variance about the $\rho = -1$ line is less than that about the $\rho = 1$ line. Again, examination of the figure confirms this.

Example 12.2.3 A pair of simple random variables

With the aid of m-functions and MATLAB we can easily caluclate the covariance and the correlation coefficient. We use the joint distribution for Example 9 in "Variance." In that example calculations show

$$E[XY] - E[X]E[Y] = -0.1633 = \text{Cov}[X, Y], \sigma_X = 1.8170 \text{ and } \sigma_Y = 1.9122$$

so that $\rho = -0.04699$.

Example 12.2.4 An absolutely continuous pair

The pair $\{X, Y\}$ has joint density function $f_{XY}(t,u) = \frac{6}{5}(t + 2u)$ on the triangular region bounded by $t = 0$, $u = t$, and $u = 1$. By the usual integration techniques, we have

$$f_X(t) = \frac{6}{5}(1 + t - 2t^2),\ 0 \le t \le 1 \text{ and } f_Y(u) = 3u^2,\ 0 \le u \le 1$$

From this we obtain $E[X] = 2/5$, $\text{Var}[X] = 3/50$, $E[Y] = 3/4$, and $\text{Var}[Y] = 3/80$. To complete the picture we need

$$E[XY] = \frac{6}{5}\int_0^1 \int_t^1 (t^2u + 2tu^2)\, dudt = 8/25$$

Then

$$\text{Cov}[X,Y] = E[XY] - E[X]E[Y] = 2/100 \text{ and } \rho = \frac{\text{Cov}[X,Y]}{\sigma_X\sigma_Y} = \frac{4}{30}\sqrt{10} \approx 0.4216$$

APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  (6/5)*(t + 2*u).*(u>=t)
Use array operations on X, Y, PX, PY, t, u, and P
EX = total(t.*P)
EX =   0.4012                    % Theoretical = 0.4
EY = total(u.*P)
EY =   0.7496                    % Theoretical = 0.75
VX = total(t.^2.*P) - EX^2
VX =   0.0603                    % Theoretical = 0.06
VY = total(u.^2.*P) - EY^2
VY =   0.0376                    % Theoretical = 0.0375
CV = total(t.*u.*P) - EX*EY
CV =   0.0201                    % Theoretical = 0.02
rho = CV/sqrt(VX*VY)
rho =  0.4212                    % Theoretical = 0.4216
```

Coefficient of linear correlation

The parameter ρ is usually called the correlation coefficient. A more descriptive name would be *coefficient of linear correlation*. The following example shows that all probability mass may be on a curve, so that $Y = g(X)$ (i.e., the value of Y is completely determined by the value of X), yet $\rho = 0$.

Example 12.2.5 $Y = g(X)$ but $\rho = 0$

Suppose X ~ uniform (-1, 1), so that $f_X(t) = 1/2$, $-1 < t < 1$ and $E[X] = 0$. Let $Y = g(X) = \cos X$. Then

$$\text{Cov}[X,Y] = E[XY] = \frac{1}{2}\int_{-1}^1 t\cos t\ dt = 0$$

Thus $\rho = 0$. Note that g could be any even function defined on (-1,1). In this case the integrand $tg(t)$ is odd, so that the value of the integral is zero.

Variance and covariance for linear combinations

We generalize the property (V4) on linear combinations. Consider the linear combinations

$$X = \sum_{i=1}^{n} a_i X_i \text{ and } Y = \sum_{j=1}^{m} b_j Y_j$$

We wish to determine $\text{Cov}[X, Y]$ and $\text{Var}[X]$. It is convenient to work with the centered random variables $X' = X - \mu_X$ and $Y' = Y - \mu_Y$. Since by linearity of expectation,

$$\mu_X = \sum_{i=1}^{n} a_i \mu_{X_i} \text{ and } \mu_Y = \sum_{j=1}^{m} b_j \mu_{Y_j}$$

we have

$$X' = \sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_{X_i} = \sum_{i=1}^{n} a_i (X_i - \mu_{X_i}) = \sum_{i=1}^{n} a_i X_i'$$

and similarly for Y'. By definition

$$\text{Cov}(X, Y) = E[X'Y'] = E[\sum_{i,j} a_i b_j X_i' Y_j'] = \sum_{i,j} a_i b_j E[X_i' E_j'] = \sum_{i,j} a_i b_j \text{Cov}(X_i, Y_j)$$

In particular

$$\text{Var}(X) = \text{Cov}(X, X) = \sum_{i,j} a_i a_j \text{Cov}(X_i, X_j) = \sum_{i=1}^{n} a_i^2 \text{Cov}(X_i, X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j)$$

Using the fact that $a_i a_j \text{Cov}(X_i, X_j) = a_j a_i \text{Cov}(X_j, X_i)$, we have

$$\text{Var}[X] = \sum_{i=1}^{n} a_i^2 \text{Var}[X_i] + 2 \sum_{i<j} a_i a_j \text{Cov}(X_i, X_j)$$

Note that a_i^2 does not depend upon the sign of a_i. If the X_i form an independent class, or are otherwise uncorrelated, the expression for variance reduces to

$$\text{Var}[X] = \sum_{i=1}^{n} a_i^2 \text{Var}[X_i]$$

This page titled 12.2: Covariance and the Correlation Coefficient is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Linear Regression

Linear Regression

Suppose that a pair $\{X, Y\}$ of random variables has a joint distribution. A value $X(\omega)$ is observed. It is desired to estimate the corresponding value $Y(\omega)$. Obvious there is no rule for determining $Y(\omega)$ unless Y is a function of X. The best that can be hoped for is some estimate based on an average of the errors, or on the average of some function of the errors.

Suppose $X(\omega)$ is observed, and by some rule an estimate $\widehat{Y}(\omega)$ is returned. The error of the estimate is $Y(\omega) - \widehat{Y}(\omega)$. The most common measure of error is the mean of the square of the error

$$E[(Y - \widehat{Y})^2]$$

The choice of the mean square has two important properties: it treats positive and negative errors alike, and it weights large errors more heavily than smaller ones. In general, we seek a rule (function) r such that the estimate $\widehat{Y}(\omega)$ is $r(X(\omega))$. That is, we seek a function r such that

$$E[(Y - r(X))^2] \text{ is a minimum.}$$

The problem of determining such a function is known as the *regression problem*. In the unit on Regression, we show that this problem is solved by the conditional expectation of Y, given X. At this point, we seek an important partial solution.

The regression line of Y on X

We seek the best straight line function for minimizing the mean squared error. That is, we seek a function r of the form $u = r(t0 = at + b$. The problem is to determine the coefficients a, b such that

$$E[(Y - aX - b)^2] \text{ is a minimum}$$

We write the error in a special form, then square and take the expectation.

$$\text{Error} = Y - aX - b = (Y - \mu_Y) - a(X - \mu_X) + \mu_Y - a\mu_X - b = (Y - \mu_Y) - a(X - \mu_X) - \beta$$

$$\text{Error squared} = (Y - \mu_Y)^2 + a^2(X - \mu_X)^2 + \beta^2 - 2\beta(Y - \mu_Y) + 2\alpha\beta(X - \mu_X) - 2a(Y - \mu_Y)(X - \mu_X)$$

$$E[(Y - aX - b)^2] = \sigma_Y^2 + a^2\sigma_X^2 + \beta^2 - 2a\text{Cov}[X, Y]$$

Standard procedures for determining a minimum (with respect to *a*) show that this occurs for

$$a = \frac{\text{Cov}[X, Y]}{\text{Var}[X]} \quad b = \mu_Y - a\mu_X$$

Thus the optimum line, called the *regression line of Y on X*, is

$$u = \frac{\text{Cov}[X, Y]}{\text{Var}[X]}(t - \mu_X) + \mu_Y = \rho\frac{\sigma_Y}{\sigma_X}(t - \mu_X) + \mu_Y = \alpha(t)$$

The second form is commonly used to define the regression line. For certain theoretical purposes, this is the preferred form. But for *calculation*, the first form is usually the more convenient. Only the covariance (which requres both means) and the variance of X are needed. There is no need to determine $\text{Var}[Y]$ or ρ.

Example 12.3.1 The simple air of Example 3 from "Variance"

```
01 jdemo1
```

```
jcalc
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
EX = total(t.*P)
EX =   0.6420
EY = total(u.*P)
EY =   0.0783
VX = total(t.^2.*P) - EX^2
VX =   3.3016
CV = total(t.*u.*P) - EX*EY
CV =  -0.1633
a = CV/VX
a  =  -0.0495
b = EY - a*EX
b  =   0.1100           % The regression line is u = -0.0495t + 0.11
```

Example 12.3.2 The pair in Example 6 from "Variance"

Suppose the pair $\{X, Y\}$ has joint density $f_{XY}(t, u) = 3u$ on the triangular region bounded by $u = 0$, $u = 1 + t$, $u = 1 - t$. Determine the regression line of Y on X.

Analytic Solution

By symmetry, $E[X] = E[XY] = 0$, so $\text{Cov}[X, Y] = 0$. The regression curve is

$$u = E[Y] = 3\int_0^1 u^2 \int_{u-1}^{1-u} \, dt du = 6\int_0^1 u^2(1-u)\, du = 1/2$$

Note that the pair is uncorrelated, but by the rectangle test is not independent. With zero values of $E[X]$ and $E[XY]$, the approximation procedure is not very satisfactory unless a very large number of approximation points are employed.

Example 12.3.3 Distribution of Example 5 from "Random Vectors and MATLAB" and Example 12 from "Function of Random Vectors"

The pair $\{X, Y\}$ has joint density $f_{XY}(t, u) = \dfrac{6}{37}(t + 2u)$ on the region $0 \le t \le 2$, $0 \le u \le \max\{1, t\}$ (see Figure 12.3.1). Determine the regression line of Y on X. If the value $X(\omega) = 1.7$ is observed, what is the best mean-square linear estimate of $Y(\omega)$?

Figure one contains two lines in the first quadrant of a cartesian graph. The horizontal axis is labeled t, and the vertical axis is labeled u. The title caption reads f_xy (t, u) = (6/37)(t + 2u). The first line crosses the vertical axis one quarter of the way up the graph. It has a positive slope, and is labeled u = 0.3382t + 0.4011. It continues as a linear plot from one side of the graph to the other. The second line begins horizontally as one segment from the left to point (1, 1). The segment is labeled u = 1. After point (1, 1), the line moves upward with a positive, constant slope to point (2, 2). This segment is labeled u = t. At (2, 2) there is a vertical line continuing downward to point (2, 0).

Figure 12.3.1. Regression line for Example 12.3.3

Analytic Solution

$$E[X] = \frac{6}{37}\int_0^1\int_0^1 (t^2 + 2tu)\ dudt + dfrac637\int_1^2\int_0^t (t^2 + 2tu)\ dudt = 50/37$$

The other quantities involve integrals over the same regions with appropriate integrands, as follows:

Quantity	Integrand	Value
$E[X^2]$	$t^3 + 2t^2u$	779/370
$E[Y]$	$tu + 2u^2$	127/148
$E[XY]$	$t^2u + 2tu^2$	232/185

Then

$$\text{Var}[X] = \frac{779}{370} - (\frac{50}{37})^2 = \frac{3823}{13690}\ textCov[X, Y] = \frac{232}{185} - \frac{50}{37}\cdot\frac{127}{148} = \frac{1293}{13690}$$

and

$$a = \text{Cov}[X, Y]/\text{Var}[X] = \frac{1293}{3823} \approx 0.3382,\ b = E[Y] - aE[X] = \frac{6133}{15292} \approx 0.4011$$

The regression line is $u = at + b$. If $X(\omega) = 1.7$, the best linear estimate (in the mean square sense) is $\widehat{Y}(\omega) = 1.7a + b = 0.9760$ (see Figure 12.3.1 for an approximate plot).

APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 2]
Enter matrix [c d] of Y-range endpoints  [0 2]
Enter number of X approximation points  400
Enter number of Y approximation points  400
Enter expression for joint density  (6/37)*(t+2*u).*(u<=max(t,1))
Use array operations on X, Y, PX, PY, t, u, and P
EX = total(t.*P)
EX =  1.3517                    % Theoretical = 1.3514
EY = total(u.*P)
EY =  0.8594                    % Theoretical = 0.8581
VX = total(t.^2.*P) - EX^2
VX =  0.2790                    % Theoretical = 0.2793
CV = total(t.*u.*P) - EX*EY
CV =  0.0947                    % Theoretical = 0.0944
a = CV/VX
a  =  0.3394                    % Theoretical = 0.3382
b = EY - a*EX
b  =  0.4006                    % Theoretical = 0.4011
y = 1.7*a + b
y  =  0.9776                    % Theoretical = 0.9760
```

An interpretation of ρ^2

The analysis above shows the minimum mean squared error is given by

$$E[(Y - \widehat{Y})^2] = E[(Y - \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) - \mu_Y)^2] = \sigma_Y^2 E[(Y^* - \rho X^*)^2]$$

$$= \sigma_Y^2 E[(Y^*)^2 - 2\rho X^*Y^* + \rho^2(X^*)^2] = \sigma_Y^2(1 - 2\rho^2 + \rho^2) = \sigma_Y^2(1 - \rho^2)$$

If $\rho = 0$, then $E[(Y - \widehat{Y})^2] = \sigma_Y^2$, the mean squared error in the case of zero linear correlation. Then, ρ^2 is interpreted as the *fraction*

of uncertainty removed by the linear rule and X. This interpretation should not be pushed too far, but is a common interpretation, often found in the discussion of observations or experimental results.

More general linear regression

Consider a jointly distributed class. $\{Y, X_1, X_2, \cdots, X_n\}$. We wish to deterimine a function U of the form

$$U = \sum_{i=0}^{n} a_i X_i, \text{ with } X_0 = 1, \text{ such that } E[(Y - U)^2] \text{ is a minimum}$$

If U satisfies this minimum condition, then $E[(Y - U)V] = 0$, or, equivalently

$$E[YV] = E[UV] \text{ for all } V \text{ of the form } V = \sum_{i=0}^{n} c_i X_i$$

To see this, set $W = Y - U$ and let $d^2 = E[W^2]$. Now, for any α

$$d^2 \le E[(W + \alpha V)^2] = d^2 + 2\alpha E[WV] + \alpha^2 E[V^2]$$

If we select the special

$$\alpha = -\frac{E[WV]}{E[V^2]} \text{ then } 0 \le -\frac{2E[WV]^2}{E[V^2]} + \frac{E[WV]^2}{E[V^2]^2} E[V^2]$$

This implies $E[WV]^2 \le 0$, which can only be satisfied by $E[WV] = 0$, so that

$$E[YV] = E[UV]$$

On the other hand, if $E[(Y - U)V] = 0$ for all V of the form above, then $E[(Y - U)^2]$ is a minimum. Consider

$$E[(Y - V)^2] = E[(Y - U + U - V)^2] = E[(Y - U)^2] + E[(U - V)^2] + 2E[(Y - U)(U - V)]$$

See $U - V$ is of the same form as V, the last term is zero. The first term is fixed. The second term is nonnegative, with zero value iff $U - V = 0$ a.s. Hence, $E[(Y - V)^2]$ is a minimum when $V = U$.

If we take V to be 1, $X_1, X_2, \cdots, X_n$, successively, we obtain $n + 1$ linear equations in the $n + 1$ unknowns $a_0, a_1, \cdots, a_n$, as follows.

$$E[Y] = a_0 + a_1 E[X_1] + \cdots + a_n E[X_n]$$
$$E[YX_1] = a_0 E[X_i] + a_1 E[X_1 X_i] + \cdots + a_n E[X_n X_i] \text{ for } 1 \le i \le n$$

For each $i = 1, 2, \cdots, n$, we take (2) - $E[X_i] \cdot (1)$ and use the calculating expressions for variance and covariance to get

$$\text{Cov}[Y, X_i] = a_1 \text{Cov}[X_1, X_i] + a_2 \text{Cov}[X_2, X_i] + \cdots + a_n \text{Cov}[X_n, X_i]$$

These n equations plus equation (1) may be solved alagebraically for the a_i.

In the important special case that the X_i are uncorrelated (i.e. $\text{Cov}[X_i, X_j] = 0$ for $i \ne j$), we have

$$a_i = \frac{\text{Cov}[Y, X_i]}{\text{Var}[X_i]} \quad 1 \le i \le n$$

and

$$a_0 = E[Y] - a_1 E[X_1] - a_2 E[X_2] - \cdots - a_n E[X_n]$$

In particular, this condition holds if the class $\{X_i : 1 \le i \le n\}$ is iid as in the case of a simple random sample (see the section on "Simple Random Samples and Statistics").

Examination shows that for $n = 1$, with $X_1 = X$, $a_0 = b$, and $a_1 = a$, the result agrees with that obtained in the treatment of the regression line, above.

Example 12.3.4 Linear regression with two variables.

Suppose $E[Y] = 3$, $E[X_1] = 2$, $E[X_2] = 3$, $\text{Var}[X_1] = 3$, $\text{Var}[X_2] = 8$, $\text{Cov}[Y, X_1] = 5$, $\text{Cov}[Y, X_2] = 7$, and $\text{Cov}[X_1, X_2] = 1$. Then the three equations are

$$a_0 + 2a_2 + 3a_3 = 3$$

$$0 + 3a_1 + 1a_2 = 5$$

$$0 + 1a_1 + 8a_2 = 7$$

Solution of these simultaneous linear equations with MATLAB gives the results

$a_0 = -1.9565$, $a_1 = 1.4348$, and $a_2 = 0.6957$.

This page titled 12.3: Linear Regression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Problems on Variance, Covariance, Linear Regression

Exercise 12.4.1

(See Exercise 1 from "Problems on Distribution and Density Functions ", and Exercise 1 from "Problems on Mathematical Expectation", m-file npr07_01.m). The class $\{C_j : 1 \le j \le 10\}$ is a partition. Random variable X has values {1, 3, 2, 3, 4, 2, 1, 3, 5, 2} on C_1 through C_{10}, respectively, with probabilities 0.08, 0.13, 0.06, 0.09, 0.14, 0.11, 0.12, 0.07, 0.11, 0.09. Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.2

(See Exercise 2 from "Problems on Distribution and Density Functions ", and Exercise 2 from "Problems on Mathematical Expectation", m-file npr07_02.m). A store has eight items for sale. The prices are \$3.50, \$5.00, \$3.50, \$7.50, \$5.00, \$5.00, \$3.50, and \$7.50, respectively. A customer comes in. She purchases one of the items with probabilities 0.10, 0.15, 0.15, 0.20, 0.10 0.05, 0.10 0.15. The random variable expressing the amount of her purchase may be written

$$X = 3.5I_{C_1} + 5.0I_{C_2} + 3.5I_{C_3} + 7.5I_{C_4} + 5.0I_{C_5} + 5.0I_{C_6} + 3.5I_{C_7} + 7.5I_{C_8}$$

Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.3

(See Exercise 12 from "Problems on Random Variables and Probabilities", Exercise 3 from "Problems on Mathematical Expectation", m-file npr06_12.m). The class $\{A, B, C, D\}$ has minterm probabilities

$$pm = 0.001 * [5\ 7\ 6\ 8\ 9\ 14\ 22\ 33\ 21\ 32\ 50\ 75\ 86\ 129\ 201\ 302]$$

Consider $X = I_A + I_B + I_C + I_D$, which counts the number of these events which occur on a trial. Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.4

(See Exercise 4 from "Problems on Mathematical Expectation"). In a thunderstorm in a national park there are 127 lightning strikes. Experience shows that the probability of each lightning strike starting a fire is about 0.0083. Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.5

(See Exercise 5 from "Problems on Mathematical Expectation"). Two coins are flipped twenty times. Let X be the number of matches (both heads or both tails). Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.6

(See Exercise 6 from "Problems on Mathematical Expectation"). A residential College plans to raise money by selling "chances" on a board. Fifty chances are sold. A player pays \$10 to play; he or she wins \$30 with probability $p = 0.2$. The profit to the College is

$$X = 50 \cdot 10 - 30N, \text{ where } N \text{ is the number of winners}$$

Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.7

(See Exercise 7 from "Problems on Mathematical Expectation"). The number of noise pulses arriving on a power circuit in an hour is a random quantity X having Poisson (7) distribution. Determine $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.8

(See Exercise 24 from "Problems on Distribution and Density Functions", and Exercise 8 from "Problems on Mathematical Expectation"). The total operating time for the units in Exercise 24 from "Problems on Distribution and Density Functions" is a random variable T ~ gamma (20, 0.0002). Determine $\mathrm{Var}[T]$.

Answer available at the end of this section.

Exercise 12.4.9

The class $\{A, B, C, D, E, F\}$ is independent, with respective probabilities

0.43, 0.53, 0.46, 0.37, 0.45, 0.39. Let

$$X = 6I_A + 13I_B - 8I_C, Y = -3I_D + 4I_E + I_F - 7$$

a. Use properties of expectation and variance to obtain $E[X]$, $\mathrm{Var}[X]$, $E[Y]$, and $\mathrm{Var}[Y]$. *Note* that it is *not* necessary to obtain the distributions for X or Y.

b. Let $Z = 3Y - 2X$.

Determine $E[Z]$, and $\mathrm{Var}[Z]$.

Answer available at the end of this section.

Exercise 12.4.10

Consider $X = -3.3I_A - 1.7I_B + 2.3I_C + 7.6I_D - 3.4$. The class $\{A, B, C, D\}$ has minterm probabilities (data are in m-file npr12_10.m)

$$\mathrm{pmx} = [0.0475\ 0.0725\ 0.0120\ 0.0180\ 0.1125\ 0.1675\ 0.0280\ 0.0420 \cdots$$

$$0.0480\ 0.0720\ 0.0130\ 0.0170\ 0.1120\ 0.1680\ 0.0270\ 0.0430]$$

a. Calculate $E[X]$ and $\text{Var}[X]$.

b. Let $W = 2X^2 - 3X + 2$.
Calculate $E[W]$ and $\text{Var}[W]$

Answer available at the end of this section.

Exercise 12.4.11

Consider a second random variable $Y = 10I_E + 17I_F + 20I_G - 10$ in addition to that in Exercise 12.4.10. The class $\{E, F, G\}$ has minterm probabilities (in mfile npr12_10.m)

$$\text{pmy} = [0.06\ 0.14\ 0.09\ 0.21\ 0.06\ 0.14\ 0.09\ 0.21]$$

The pair $\{X, Y\}$ is independent.

a. Calculate $E[Y]$ and $\text{Var}[Y]$.

b. Let $Z = X^2 + 2XY - Y$.
Calculate $E[Z]$ and $\text{Var}[Z]$.

Answer available at the end of this section.

Exercise 12.4.12

Suppose the pair $\{X, Y\}$ is independent, with X ~ gamma (3, 0.1) and

Y ~ Poisson (13). Let $Z = 2X - 5Y$. Determine $E[Z]$ and $\text{Var}[Z]$.

Answer available at the end of this section.

Exercise 12.4.13

The pair $\{X, Y\}$ is jointly distributed with the following parameters:

$$E[X] = 3,\ E[Y] = 4,\ E[XY] = 15,\ E[X^2] = 11,\ \text{Var}[Y] = 5$$

Determine $\text{Var}[3X - 2Y]$.

Answer available at the end of this section.

Exercise 12.4.14

The class $\{A, B, C, D, E, F\}$ is independent, with respective probabilities

$$0.47,\ 0.33,\ 0.46,\ 0.27,\ 0.41,\ 0.37$$

Let

$$X = 8I_A + 11I_B - 7I_C,\ Y = -3I_D + 5I_E + I_F - 3, \text{ and } Z = 3Y - 2X$$

a. Use properties of expectation and variance to obtain $E[X]$, $\text{Var}[X]$, $E[Y]$, and $\text{Var}[Y]$.

b. Determine $E[Z]$, and $\text{Var}[Z]$.

c. Use appropriate m-programs to obtain $E[X]$, $\mathrm{Var}[X]$, $E[Y]$, $\mathrm{Var}[Y]$, $E[Z]$, and $\mathrm{Var}[Z]$. Compare with results of parts (a) and (b).

Answer available at the end of this section.

Exercise 12.4.15

For the Beta (r, s) distribution.

a. Determine $E[X^n]$, where n is a positive integer.

b. Use the result of part (a) to determine $E[X]$ and $\mathrm{Var}[X]$.

Answer available at the end of this section.

Exercise 12.4.16

The pair $\{X, Y\}$ has joint distribution. Suppose

$$E[X] = 3,\ E[X^2] = 11,\ E[Y] = 10,\ E[Y^2] = 101,\ E[XY] = 30$$

Determine $\mathrm{Var}[15X - 2Y]$.

Answer available at the end of this section.

Exercise 12.4.17

The pair $\{X, Y\}$ has joint distribution. Suppose

$$E[X] = 2,\ E[X^2] = 5,\ E[Y] = 1,\ E[Y^2] = 2,\ E[XY] = 1$$

Determine $\mathrm{Var}[3X + 2Y]$.

Answer available at the end of this section.

Exercise 12.4.18

The pair $\{X, Y\}$ is independent, with

$$E[X] = 2,\ E[Y] = 1,\ \mathrm{Var}[X] = 6,\ \mathrm{Var}[Y] = 4$$

Let $Z = 2X^2 + XY^2 - 3Y + 4$.

Determine $E[Z]$.

Answer available at the end of this section.

Exercise 12.4.19

(See Exercise 9 from "Problems on Mathematical Expectation"). Random variable *X* has density function

$$f_X(t) = \begin{cases} (6/5)t^2 & \text{for } 0 \le t \le 1 \\ (6/5)(2-t) & \text{for } 1 < t \le 2 \end{cases} = I_{[0,1]}(t)\frac{6}{5}t^2 + I_{(1,2]}(t)\frac{6}{5}(2-t)$$

$E[X] = 11/10$. Determine $\text{Var}[X]$.

Answer available at the end of this section.

For the distributions in Exercises 20-22

Determine $\text{Var}[X]$, $\text{Cov}[X, Y]$, and the regression line of Y on X.

Exercise 12.4.20

(See Exercise 7 from "Problems On Random Vectors and Joint Distributions", and Exercise 17 from "Problems on Mathematical Expectation"). The pair $\{X, Y\}$ has the joint distribution (in file npr08_07.m):

$$P(X = t, Y = u)$$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189
-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

Answer available at the end of this section.

Exercise 12.4.21

(See Exercise 8 from "Problems On Random Vectors and Joint Distributions", and Exercise 18 from "Problems on Mathematical Expectation"). The pair $\{X, Y\}$ has the joint distribution (in file npr08_08.m):

$$P(X = t, Y = u)$$

t =	1	3	5	7	9	11	13	15	17	19
u = 12	0.0156	0.0191	0.0081	0.0035	0.0091	0.0070	0.0098	0.0056	0.0091	0.0049
10	0.0064	0.0204	0.0108	0.0040	0.0054	0.0080	0.0112	0.0064	0.0104	0.0056
9	0.0196	0.0256	0.0126	0.0060	0.0156	0.0120	0.0168	0.0096	0.0056	0.0084
5	0.0112	0.0182	0.0108	0.0070	0.0182	0.0140	0.0196	0.0012	0.0182	0.0038
3	0.0060	0.0260	0.0162	0.0050	0.0160	0.0200	0.0280	0.0060	0.0160	0.0040
-1	0.0096	0.0056	0.0072	0.0060	0.0256	0.0120	0.0268	0.0096	0.0256	0.0084
-3	0.0044	0.0134	0.0180	0.0140	0.0234	0.0180	0.0252	0.0244	0.0234	0.0126
-5	0.0072	0.0017	0.0063	0.0045	0.0167	0.0090	0.0026	0.0172	0.0217	0.0223

Answer available at the end of this section.

Exercise 12.4.22

(See Exercise 9 from "Problems On Random Vectors and Joint Distributions", and Exercise 19 from "Problems on Mathematical Expectation"). Data were kept on the effect of training time on the time to perform a job on a production line. X is the amount of training, in hours, and Y is the time to perform the task, in minutes. The data are as follows (in file npr08_09.m):

$$P(X = t, Y = u)$$

t =	1	1.5	2	2.5	3
u = 5	0.039	0.011	0.005	0.001	0.001
4	0.065	0.070	0.050	0.015	0.010
3	0.031	0.061	0.137	0.051	0.033
2	0.012	0.049	0.163	0.058	0.039
1	0.003	0.009	0.045	0.025	0.017

Answer available at the end of this section.

For the joint densities in Exercises 23-30 below

a. Determine analytically $\mathrm{Var}[X]$ $\mathrm{Cov}[X, Y]$, and the regression line of Y on X.
b. Check these with a discrete approximation.

Exercise 12.4.23

(See Exercise 10 from "Problems On Random Vectors and Joint Distributions", and Exercise 20 from "Problems on Mathematical Expectation"). $f_{XY}(t, u) = 1$ for $0 \le t \le 1$, $0 \le u \le 2(1 - t)$.

$$E[X] = \frac{1}{3},\ E[X^2] = \frac{1}{6},\ E[Y] = \frac{2}{3}$$

Answer available at the end of this section.

Exercise 12.4.24

(See Exercise 13 from "Problems On Random Vectors and Joint Distributions", and Exercise 23 from "Problems on Mathematical Expectation"). $f_{XY}(t, u) = \frac{1}{8}(t + u)$ for $0 \le t \le 2$, $0 \le u \le 2$.

$$E[X] = E[Y] = \frac{7}{6},\ E[X^2] = \frac{5}{3}$$

Answer available at the end of this section.

Exercise 12.4.25

(See Exercise 15 from "Problems On Random Vectors and Joint Distributions", and Exercise 25 from "Problems on Mathematical Expectation"). $f_{XY}(t, u) = \frac{3}{88}(2t + 3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1 + t$.

$$E[X] = \frac{313}{220},\ E[Y] = \frac{1429}{880},\ E[X^2] = \frac{49}{22}$$

Answer available at the end of this section.

Exercise 12.4.26

(See Exercise 16 from "Problems On Random Vectors and Joint Distributions", and Exercise 26 from "Problems on Mathematical Expectation"). $f_{XY}(t,u) = 12t^2u$ on the parallelogram with vertices

(-1, 0), (0, 0), (1, 1), (0, 1)

$$E[X] = \frac{2}{5},\ E[Y] = \frac{11}{15},\ E[X^2] = \frac{2}{5}$$

Answer available at the end of this section.

Exercise 12.4.27

(See Exercise 17 from "Problems On Random Vectors and Joint Distributions", and Exercise 27 from "Problems on Mathematical Expectation"). $f_{XY}(t,u) = \frac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2-t\}$.

$$E[X] = \frac{52}{55},\ E[Y] = \frac{32}{55},\ E[X^2] = \frac{627}{605}$$

Answer available at the end of this section.

Exercise 12.4.28

(See Exercise 18 from "Problems On Random Vectors and Joint Distributions", and Exercise 28 from "Problems on Mathematical Expectation"). $f_{XY}(t,u) = \frac{3}{23}(t+2u)$, for $0 \le t \le 2$, $0 \le u \le \max\{2-t, t\}$.

$$E[X] = \frac{53}{46},\ E[Y] = \frac{22}{23},\ E[X^2] = \frac{9131}{5290}$$

Answer available at the end of this section.

Exercise 12.4.29

(See Exercise 21 from "Problems On Random Vectors and Joint Distributions", and Exercise 31 from "Problems on Mathematical Expectation"). $f_{XY}(t,u) = \frac{2}{13}(t+2u)$, for $0 \le t \le 2$, $0 \le u \le \min\{2t, 3-t\}$.

$$E[X] = \frac{16}{13},\ E[Y] = \frac{11}{12},\ E[X^2] = \frac{2847}{1690}$$

Answer available at the end of this section.

Exercise 12.4.30

(See Exercise 22 from "Problems On Random Vectors and Joint Distributions", and Exercise 32 from "Problems on Mathematical Expectation"). $f_{XY}(t,u) = I_{[0,1]}(t)\frac{3}{8}(t^2+2u) + I_{(1,2]}(t)\frac{9}{14}t^2u^2$, for $0 \le u \le 1$.

$$E[X] = \frac{243}{224},\ E[Y] = \frac{11}{16},\ E[X^2] = \frac{107}{70}$$

Answer available at the end of this section.

Exercise 12.4.31

The class $\{X, Y, Z\}$ of random variables is iid (independent, identically distributed) with common distribution

$$X = \text{[-5 -1 3 4 7]}\ PX = 0.01 * \text{[15 20 30 25 10]}$$

Let $W = 3X - 4Y + 2Z$. Determine $E[W]$ and $\text{Var}[W]$. Do this using icalc, then repeat with icalc3 and compare results.

Answer available at the end of this section.

Exercise 12.4.32

$f_{XY}(t, u) = \dfrac{3}{88}(2t + 3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1 + t$ (see Exercise 25 and Exercise 37 from "Problems on Mathematical Expectation").

$$Z = I_{[0,1]}(X)4X + I_{(1,2]}(X)(X + Y)$$

$$E[X] = \frac{313}{220},\ E[Z] = \frac{5649}{1760},\ E[Z^2] = \frac{4881}{440}$$

Determine $\text{Var}[Z]$ and $\text{Cov}[X, Z]$. Check with discrete approximation.

Answer available at the end of this section.

Exercise 12.4.33

$f_{XY}(t, u) = \dfrac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2 - t\}$ (see Exercise 27 and Exercise 38 from "Problems on Mathematical Expectation").

$$Z = I_M(X, Y)(X + Y) + I_{M^c}(X, Y)2Y,\ M = \{(t, u) : \max(t, u) \le 1\}$$

$$E[X] = \frac{52}{55},\ E[Z] = \frac{16}{55},\ E[Z^2] = \frac{39}{308}$$

Determine $\text{Var}[Z]$ and $\text{Cov}[X, Z]$. Check with discrete approximation.

Answer available at the end of this section.

Exercise 12.4.34

$f_{XY}(t, u) = \dfrac{3}{23}(t + 2u)$ for $0 \le t \le 2$, $0 \le u \le \max\{2 - t, t\}$ (see Exercise 28 and Exercise 39 from "Problems on Mathematical Expectation").

$$Z = I_M(X, Y)(X + Y) + I_{M^c}(X, Y)2Y,\ M = \{(t, u) : \max(t, u) \le 1\}$$

$$E[X] = \frac{53}{46},\ E[Z] = \frac{175}{92},\ E[Z^2] = \frac{2063}{460}$$

Determine $\text{Var}[Z]$ and $\text{Cov}[Z]$. Check with discrete approximation.

Answer available at the end of this section.

Exercise 12.4.35

$f_{XY}(t,u) = \frac{12}{179}(3t^2 + u)$, for $0 \le t \le 2$, $0 \le u \le \min\{2, 3 - t\}$ (see Exercise 29 and Exercise 40 from "Problems on Mathematical Expectation").

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y^2,\ M = \{(t,u) : t \le 1, u \ge 1\}$$

$$E[X] = \frac{2313}{1790},\ E[Z] = \frac{1422}{895},\ E[Z^2] = \frac{28296}{6265}$$

Determine $\mathrm{Var}[Z]$ and $\mathrm{Cov}[X,Z]$. Check with discrete approximation.

Answer available at the end of this section.

Exercise 12.4.36

$f_{XY}(t,u) = \frac{12}{227}(3t + 2tu)$, for $0 \le t \le 2$, $0 \le u \le \min\{1 + t, 2\}$ (see Exercise 30 and Exercise 41 from "Problems on Mathematical Expectation").

$$Z = I_M(X,Y)X + I_{M^c}(X,Y)XY,\ M = \{(t,u) : u \le \min(1, 2 - t)\}$$

$$E[X] = \frac{1567}{1135},\ E[Z] = \frac{5774}{3405},\ E[Z^2] = \frac{56673}{15890}$$

Determine $\mathrm{Var}[Z]$ and $\mathrm{Cov}[X,Z]$. Check with discrete approximation.

Answer available at the end of this section.

Exercise 12.4.37

(See Exercise 12.4.20, and Exercises 9 and 10 from "Problems on Functions of Random Variables"). For the pair $\{X,Y\}$ in Exercise 12.4.20, let

$$Z = g(X,Y) = 3X^2 + 2XY - Y^2$$

$$W = h(X,Y) = \begin{cases} X & \text{for } X + Y \le 4 \\ 2Y & \text{for } X + Y > 4 \end{cases} = I_M(X,Y)X + I_{M^c}(X,Y)2Y$$

Determine the joint distribution for the pair $\{Z,W\}$ and determine the regression line of W on Z.

Answer available at the end of this section.

Answers

Exercise 12.4.1

```
npr07_01
Data are in T and pc
EX = T*pc'
EX =  2.7000
VX = (T.^2)*pc' - EX^2
VX =  1.5500
[X,PX] = csort(T,pc);    % Alternate
Ex = X*PX'
Ex =  2.7000
```

```
Vx = (X.^2)*PX' - EX^2
Vx =  1.5500
```

Exercise 12.4.2

```
npr07_02
Data are in T, pc
EX = T*pc';
VX = (T.^2)*pc' - EX^2
VX =  2.8525
```

Exercise 12.4.3

```
npr06_12
Minterm probabilities in pm, coefficients in c
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  pm
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
VX = (X.^2)*PX' - (X*PX')^2
VX =  0.7309
```

Exercise 12.4.4

X ~ binomial (127, 0.0083). $\mathrm{Var}[X] = 127 \cdot 0.0083 \cdot (1 - 0.0083) = 1.0454$.

Exercise 12.4.5

X ~ binomial (20, 1/2). $\mathrm{Var}[X] = 20 \cdot (1/2)^2 = 5$.

Exercise 12.4.6

N ~ binomial (50, 0.2). $\mathrm{Var}[N] = 50 \cdot 0.2 \cdot 0.8 = 8$. $\mathrm{Var}[X] = 30^2\ \mathrm{Var}[N] = 7200$.

Exercise 12.4.7

X ~ Poisson (7). $\mathrm{Var}[X] = \mu = 7$.

Exercise 12.4.8

T ~ gamma (20, 0.0002). $\mathrm{Var}[T] = 20/0.0002^2 = 500,000,000$.

Exercise 12.4.9

```
cx = [6 13 -8 0];
cy = [-3 4 1 -7];
px = 0.01*[43 53 46 100];
py = 0.01*[37 45 39 100];
EX = dot(cx,px)
EX =   5.7900
EY = dot(cy,py)
EY =  -5.9200
VX = sum(cx.^2.*px.*(1-px))
VX =  66.8191
VY = sum(cy.^2.*py.*(1-py))
VY =   6.2958
EZ = 3*EY - 2*EX
EZ = -29.3400
VZ = 9*VY + 4*VX
VZ = 323.9386
```

Exercise 12.4.10

```
npr12_10
Data are in cx, cy, pmx and pmy
canonic
 Enter row vector of coefficients  cx
 Enter row vector of minterm probabilities  pmx
Use row matrices X and PX for calculations
```

```
Call for XDBN to view the distribution
EX = dot(X,PX)
EX =  -1.2200
VX = dot(X.^2,PX) - EX^2
VX =  18.0253
G = 2*X.^2 - 3*X + 2;
[W,PW] = csort(G,PX);
EW = dot(W,PW)
EW =  44.6874
VW = dot(W.^2,PW) - EW^2
VW =  2.8659e+03
```

Exercise 12.4.11

(Continuation of Exercise 12.4.10)

```
[Y,PY] = canonicf(cy,pmy);
EY = dot(Y,PY)
EY =  19.2000
VY = dot(Y.^2,PY) - EY^2
VY = 178.3600
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
H = t.^2 + 2*t.*u - u;
[Z,PZ] = csort(H,P);
EZ = dot(Z,PZ)
EZ = -46.5343
VZ = dot(Z.^2,PZ) - EZ^2
VZ =  3.7165e+04
```

Exercise 12.4.12

X ~ gamma (3, 0.1) implies $E[X] = 30$ and $\text{Var}[X] = 300$. Y ~ Poisson (13) implies $E[Y] = \text{Var}[Y] = 13$. Then

$$E[Z] = 2 \cdot 30 - 5 \cdot 13 = -5, \text{Var}[Z] = 4 \cdot 300 + 25 \cdot 13 = 1525.$$

Exercise 12.4.13

```
EX = 3;
EY = 4;
EXY = 15;
EX2 = 11;
VY = 5;
VX = EX2 - EX^2
VX =  2
CV = EXY - EX*EY
CV =  3
VZ = 9*VX + 4*VY - 6*2*CV
VZ =  2
```

Exercise 12.4.14

```
px = 0.01*[47 33 46 100];
py = 0.01*[27 41 37 100];
cx = [8 11 -7 0];
cy = [-3 5 1 -3];
ex = dot(cx,px)
ex =   4.1700
ey = dot(cy,py)
```

```
ey =  -1.3900
vx = sum(cx.^2.*px.*(1 - px))
vx =  54.8671
vy = sum(cy.^2.*py.*(1-py))
vy =   8.0545
[X,PX] = canonicf(cx,minprob(px(1:3)));
[Y,PY] = canonicf(cy,minprob(py(1:3)));
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
EX = dot(X,PX)
EX =   4.1700
EY = dot(Y,PY)
EY =  -1.3900
VX = dot(X.^2,PX) - EX^2
VX =  54.8671
VY = dot(Y.^2,PY) - EY^2
VY =   8.0545
EZ = 3*EY - 2*EX
EZ = -12.5100
VZ = 9*VY + 4*VX
VZ = 291.9589
```

Exercise 12.4.15

$$E[X^n] = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_0^1 t^{r+n-1} dt = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \cdot \frac{\Gamma(r+n)\Gamma(s)}{\Gamma(r+s+n)} = \frac{\Gamma(r+n)\Gamma(r+s)}{\Gamma(r+s+n)\Gamma(r)}$$

Using $\Gamma(x+1) = x\Gamma(x)$ we have

$$E[X] = \frac{r}{r+s}, E[X^2] = \frac{r(r+1)}{(r+s)(r+s+1)}$$

Some algebraic manipulations show that

$$\text{Var}[X] = E[X^2] - E^2[X] = \frac{rs}{(r+s)^2(r+s+1)}$$

Exercise 12.4.16

```
EX = 3;
EX2 = 11;
EY = 10;
EY2 = 101;
EXY = 30;
VX = EX2 - EX^2
VX =    2
VY = EY2 - EY^2
VY =    1
CV = EXY - EX*EY
CV =    0
VZ = 15^2*VX + 2^2*VY
VZ =  454
```

Exercise 12.4.17

```
EX = 2;
EX2 = 5;
EY = 1;
EY2 = 2;
EXY = 1;
VX = EX2 - EX^2
```

```
VX =    1
VY = EY2 - EY^2
VY =    1
CV = EXY - EX*EY
CV =   -1
VZ = 9*VX + 4*VY + 2*6*CV
VZ =    1
```

Exercise 12.4.18

```
EX = 2;
EY = 1;
VX = 6;
VY = 4;
EX2 = VX + EX^2
EX2 =  10
EY2 = VY + EY^2
EY2 =   5
EZ = 2*EX2 + EX*EY2 - 3*EY + 4
EZ =   31
```

Exercise 12.4.19

$$E[X^2] = \int t^2 f_X(t)\, dt = \frac{6}{5}\int_0^1 t^4\, dt + \frac{6}{5}\int_1^2 (2t^2 - t^3)\, dt = \frac{67}{50}$$

$$\text{Var}[X] = E[X^2] - E^2[X] = \frac{13}{100}$$

Exercise 12.4.20

```
npr08_07
Data are in X, Y, P
jcalc
- - - - - - - - - - -
EX = dot(X,PX);
EY = dot(Y,PY);
VX = dot(X.^2,PX) - EX^2
VX =   5.1116
CV = total(t.*u.*P) - EX*EY
CV =   2.6963
a = CV/VX
a =    0.5275
b = EY - a*EX
b =    0.6924       % Regression line: u = at + b
```

Exercise 12.4.21

```
npr08_08
Data are in X, Y, P
jcalc
- - - - - - - - - - -
EX = dot(X,PX);
EY = dot(Y,PY);
VX = dot(X.^2,PX) - EX^2
VX =  31.0700
CV = total(t.*u.*P) - EX*EY
CV =  -8.0272
a  = CV/VX
a  =  -0.2584
b = EY - a*EX
b  =   5.6110       % Regression line: u = at + b
```

Exercise 12.4.22

```
npr08_09
Data are in X, Y, P
jcalc
- - - - - - - - - - - -
EX = dot(X,PX);
EY = dot(Y,PY);
VX = dot(X.^2,PX) - EX^2
VX =   0.3319
CV = total(t.*u.*P) - EX*EY
CV =  -0.2586
a  = CV/VX
a  =  -0.77937/6;
b = EY - a*EX
b  =   4.3051       % Regression line: u = at + b
```

Exercise 12.4.23

$$E[XY] = \int_0^1 \int_0^{2(1-t)} tu \; dudt = 1/6$$
$$\text{Cov}[X,Y] = \frac{1}{6} - \frac{1}{3}\cdot\frac{2}{3} = -1/18 \; \text{Var}[X] = 1/6 - (1/3)^2 = 1/18$$
$$a = \text{Cov}[X,Y]/\text{Var}[X] = -1 \; b = E[Y] - aE[X] = 1$$

```
tuappr: [0 1] [0 2] 200 400  u<=2*(1-t)
EX = dot(X,PX);
EY = dot(Y,PY);
VX = dot(X.^2,PX) - EX^2
VX =   0.0556
CV = total(t.*u.*P) - EX*EY
CV =  -0.0556
a = CV/VX
a =   -1.0000
b = EY - a*EX
b =    1.0000
```

Exercise 12.4.24

$$E[XY] = \frac{1}{8}\int_0^2 \int_0^2 tu(t+u) \; dudt = 4/3, \; \text{Cov}[X,Y] = -1/36, \; \text{Var}[X] = 11/36$$
$$a = \text{Cov}[X,Y]/\text{Var}[X] = -1/11, \; b = E[Y] - aE[X] = 14/11$$

```
tuappr:  [0 2] [0 2] 200 200 (1/8)*(t+u)
VX =  0.3055  CV = -0.0278  a = -0.0909  b =  1.2727
```

Exercise 12.4.25

$$E[XY] = \frac{3}{88}\int_0^2 \int_0^{1+t} tu(2t+3u^2) \; dudt = 2153/880, \; \text{Cov}[X,Y] = 26383/1933600, \; \text{Var}[X] = 9831/48400$$
$$a = \text{Cov}[X,Y]/\text{Var}[X] = 26383/39324, \; b = E[Y] - aE[X] = 26321/39324$$

```
tuappr:  [0 2] [0 3] 200 300  (3/88)*(2*t + 3*u.^2).*(u<=1+t)
VX =  0.2036  CV = 0.1364 a = 0.6700  b = 0.6736
```

Exercise 12.4.26

$$E[XY] = 12\int_{-1}^0 \int_0^{t+1} t^3u^2 \; dudt + 12\int_0^1 \int_t^1 t^3u^2 \; dudt = \frac{2}{5}$$
$$\text{Cov}[X,Y] = \frac{8}{75}, \; \text{Var}[X] = \frac{6}{25}$$
$$a = \text{Cov}[X,Y]/\text{Var}[X] = 4/9, \; b = E[Y] - aE[X] = 5/9$$

```
tuappr: [-1 1] [0 1] 400 200  12*t.^2.*u.*(u>= max(0,t)).*(u<= min(1+t,1))
VX = 0.2383  CV = 0.1056  a = 0.4432  b = 0.5553
```

Exercise 12.4.27

$$E[XY] = \frac{24}{11}\int_0^1 \int_0^1 t^2u^2 \; dudt + \frac{24}{11}\int_1^2 \int_0^{2-t} t^2u^2 \; dudt = \frac{28}{55}$$
$$\text{Cov}[X,Y] = -\frac{124}{3025}, \; \text{Var}[X] = \frac{431}{3025}$$

$$a = \text{Cov}[X,Y]/\text{Var}[X] = -\frac{124}{431},\ b = E[Y] - aE[X] = \frac{368}{431}$$

```
tuappr: [0 2] [0 1] 400 200 (24/11)*t.*u.*(u<=min(1,2-t))
VX = 0.1425  CV =-0.0409  a = -0.2867 b = 0.8535
```

Exercise 12.4.28

$$E[XY] = \frac{3}{23}\int_0^1\int_0^{2-t} tu(t+2u)\,dudt + \frac{3}{23}\int_1^2\int_0^t tu(t+2u)\,dudt = \frac{251}{230}$$

$$\text{Cov}[X,Y] = -\frac{57}{5290},\ \text{Var}[X] = \frac{4217}{10580}$$

$$a = \text{Cov}[X,Y]/\text{Var}[X] = -\frac{114}{4217},\ b = E[Y] - aE[X] = \frac{4165}{4217}$$

```
tuappr: [0 2] [0 2] 200 200 (3/23)*(t + 2*u).*(u<=max(2-t,t))
VX = 0.3984 CV = -0.0108  a = -0.0272  b = 0.9909
```

Exercise 12.4.29

$$E[XY] = \frac{2}{13}\int_0^1\int_0^{3-t} tu(t+2u)\,dudt + \frac{2}{13}\int_1^2\int_0^{2t} tu(t+2u)\,dudt = \frac{431}{390}$$

$$\text{Cov}[X,Y] = -\frac{3}{130},\ \text{Var}[X] = \frac{287}{1690}$$

$$a = \text{Cov}[X,Y]/\text{Var}[X] = -\frac{39}{297},\ b = E[Y] - aE[X] = \frac{3733}{3444}$$

```
tuappr: [0 2] [0 2] 400 400 (2/13)*(t + 2*u).*(u<=min(2*t,3-t))
VX = 0.1698  CV = -0.0229  a = -0.1350  b = 1.0839
```

Exercise 12.4.30

$$E[XY] = \frac{3}{8}\int_0^1\int_0^1 tu(t^2+2u)\,dudt + \frac{9}{14}\int_1^2\int_0^1 t^3u^3\,dudt = \frac{347}{448}$$

$$\text{Cov}[X,Y] = -\frac{103}{3584},\ \text{Var}[X] = \frac{88243}{250880}$$

$$a = \text{Cov}[X,Y]/\text{Var}[X] = -\frac{7210}{88243},\ b = E[Y] - aE[X] = \frac{105691}{176486}$$

```
tuappr: [0 2] [0 1] 400 200 (3/8)*(t.^2 + 2*u).*(t<=1) + (9/14)*t.^2.*u.^2.*(t>1)
VX = 0.3517  CV = 0.0287 a = 0.0817  b = 0.5989
```

Exercise 12.4.31

```
x = [-5 -1 3 4 7];
px = 0.01*[15 20 30 25 10];
EX = dot(x,px)                % Use of properties
EX =   1.6500
VX = dot(x.^2,px) - EX^2
VX =  12.8275
EW = (3 - 4+ 2)*EX
EW =   1.6500
VW = (3^2 + 4^2 + 2^2)*VX
VW = 371.9975
icalc                         % Iterated use of icalc
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter X probabilities  px
Enter Y probabilities  px
 Use array operations on matrices X, Y, PX, PY, t, u, and P
G = 3*t - 4*u;
[R,PR] = csort(G,P);
icalc
Enter row matrix of X-values  R
Enter row matrix of Y-values  x
Enter X probabilities  PR
Enter Y probabilities  px
 Use array operations on matrices X, Y, PX, PY, t, u, and P
H = t + 2*u;
```

```
[W,PW] = csort(H,P);
EW = dot(W,PW)
EW =   1.6500
VW = dot(W.^2,PW) - EW^2
VW = 371.9975
icalc3                        % Use of icalc3
Enter row matrix of X-values  x
Enter row matrix of Y-values  x
Enter row matrix of Z-values  x
Enter X probabilities  px
Enter Y probabilities  px
Enter Z probabilities  px
Use array operations on matrices X, Y, Z,
PX, PY, PZ, t, u, v, and P
S = 3*t - 4*u + 2*v;
[w,pw] = csort(S,P);
Ew = dot(w,pw)
Ew =   1.6500
Vw = dot(w.^2,pw) - Ew^2
Vw = 371.9975
```

Exercise 12.4.32

$$E[XZ] = \frac{3}{88}\int_0^1\int_0^{1+t} 4t^2(2t+2u^2)\,dudt + \frac{3}{88}\int_1^2\int_0^{1+t} t(t+u)(2t+3u^2)\,dudt = \frac{16931}{3520}$$
$$\text{Var}[Z] = E[Z^2] - E^2[Z] = \frac{2451039}{3097600}\ \text{Cov}[X,Z] = E[XZ] - E[X]E[Z] = \frac{94273}{387200}$$

```
tuappr: [0 2] [0 3] 200 300 (3/88)*(2*t+3*u.^2).*(u<=1+t)
G = 4*t.*(t<=1) + (t+u).*(t>1);
EZ = total(G.*P)
EZ = 3.2110
EX = dot(X,PX)
EX = 1.4220
CV = total(G.*t.*P) - EX*EZ
CV = 0.2445                       % Theoretical 0.2435
VZ = total(G.^2.*P) - EZ^2
VZ = 0.7934                       % Theoretical 0.7913
```

Exercise 12.4.33

$$E[XZ] = \frac{24}{11}\int_0^1\int_t^1 t(t/2)tu\,dudt + \frac{24}{11}\int_0^1\int_0^t tu^2tu\,dudt\frac{24}{11}\int_1^2\int_0^{2-t} ttu^2tu\,dudt = \frac{211}{770}$$
$$\text{Var}[Z] = E[Z^2] - E^2[Z] = \frac{3557}{84700}\ \text{Cov}[Z,X] = E[XZ] - E[X]E[Z] = -\frac{43}{42350}$$

```
tuappr:  [0 2] [0 1] 400 200 (24/11)*t.*u.*(u<=min(1,2-t))
G = (t/2).*(u>t) + u.^2.*(u<=t);
VZ = total(G.^2.*P) - EZ^2
VZ =   0.0425
CV = total(t.*G.*P) - EZ*dot(X,PX)
CV = -9.2940e-04
```

Exercise 12.4.34

$$E[ZX] = \frac{3}{23}\int_0^1\int_0^1 t(t+u)(t+2u)\,dudt + \frac{3}{23}\int_0^1\int_1^{2-t} 2tu(t+2u)\,dudt +$$
$$\frac{3}{23}\int_1^2\int_1^t 2tu(t+2u)\,dudt = \frac{1009}{460}$$
$$\text{Var}[Z] = E[Z^2] - E^2[Z] = \frac{36671}{42320}\ \text{Cov}[Z,X] = E[ZX] - E[Z]E[X] = \frac{39}{21160}$$

```
tuappr:  [0 2] [0 2] 400 400 (3/23)*(t+2*u).*(u<=max(2-t,t))
M = max(t,u)<=1;
G = (t+u).*M + 2*u.*(1-M);
EZ = total(G.*P);
EX = dot(X,PX);
```

```
CV = total(t.*G.*P) - EX*EZ
CV =  0.0017
```

Exercise 12.4.35

$$E[ZX] = \frac{12}{179}\int_0^1\int_1^2 t(t+u)(3t^2+u)\,dudt + \frac{12}{179}\int_0^1\int_0^1 2tu^2(3t^2+u)\,dudt + \frac{12}{179}\int_1^2\int_0^{3-t} 2tu^2(3t^2+u)\,dudt = \frac{24029}{12530}$$

$$\text{Var}[Z] = E[Z^2] - E^2[Z] = \frac{11170332}{5607175}\ \text{Cov}[Z,X] = E[ZX] - E[Z]E[X] = -\frac{1517647}{11214350}$$

```
tuappr:  [0 2] [0 2] 400 400 (12/179)*(3*t.^2 + u).*(u <= min(2,3-t))
M = (t<=1)&(u>=1);
G = (t + u).*M + 2*u.^2.*(1 - M);
EZ = total(G.*P);
EX = dot(X,PX);
CV = total(t.*G.*P) - EZ*EX
CV = -0.1347
```

Exercise 12.4.36

$$E[ZX] = \frac{12}{227}\int_0^1\int_0^1 t^2(3t+2tu)\,dudt + \frac{12}{227}\int_1^2\int_0^{2-t} t^2(3t+2tu)\,dudt + \frac{12}{227}\int_0^1\int_1^{1+t} t^2u(3t+2tu)\,dudt + \frac{12}{227}\int_1^2\int_{2-t}^2 t^2u(3t+2tu)\,dudt = \frac{20338}{7945}$$

$$\text{Var}[Z] = E[Z^2] - E^2[Z] = \frac{112167631}{162316350}\ \text{Cov}[Z,X] = E[ZX] - E[Z]E[X] = \frac{5915884}{27052725}$$

```
tuappr: [0 2] [0 2] 400 400 (12/227)*(3*t + 2*t.*u).*(u <= min(1+t,2))
EX = dot(X,PX);
M = u <= min(1,2-t);
G = t.*M + t.*u.*(1 - M);
EZ = total(G.*P);
EZX = total(t.*G.*P)
EZX =  2.5597
CV = EZX - EX*EZ
CV =   0.2188
VZ = total(G.^2.*P) - EZ^2
VZ =   0.6907
```

Exercise 12.4.37

```
npr08_07
Data are in X, Y, P
jointzw
Enter joint prob for (X,Y) P
Enter values for X X
Enter values for Y Y
Enter expression for g(t,u) 3*t.^2 + 2*t.*u - u.^2
Enter expression for h(t,u) t.*(t+u<=4) + 2*u.*(t+u>4)
Use array operations on Z, W, PZ, PW, v, w, PZW
EZ = dot(Z,PZ)
EZ =    5.2975
EW = dot(W,PW)
EW =    4.7379
VZ = dot(Z.^2,PZ) - EZ^2
VZ =    1.0588e+03
CZW = total(v.*w.*PZW) - EZ*EW
CZW = -12.1697
a = CZW/VZ
a =   -0.0115
b = EW - a*EZ
b =    4.7988                 % Regression line: w = av + b
```

This page titled 12.4: Problems on Variance, Covariance, Linear Regression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

13: TRANSFORM METHODS

13.1: Transform Methods
13.2: Convergence and the Central Limit Theorem
13.3: Simple Random Samples and Statistics
13.4: Problems on Transform Methods

This page titled 13: Transform Methods is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Transform Methods

As pointed out in the units on Expectation and Variance, the mathematical expectation $E[X] = \mu_X$ of a random variable X locates the center of mass for the induced distribution, and the expectation

$$E[g(X)] = E[(X - E[X])^2] = \text{Var}[X] = \sigma_X^2$$

measures the spread of the distribution about its center of mass. These quantities are also known, respectively, as the mean (moment) of X and the second moment of X about the mean. Other moments give added information. For example, the third moment about the mean $E[(X - \mu_X)^3]$ gives information about the skew, or asymetry, of the distribution about the mean. We investigate further along these lines by examining the expectation of certain functions of X. Each of these functions involves a parameter, in a manner that completely determines the distribution. For reasons noted below, we refer to these as *transforms*. We consider three of the most useful of these.

Three basic transforms

We define each of three transforms, determine some key properties, and use them to study various probability distributions associated with random variables. In the section on integral transforms, we show their relationship to well known integral transforms. These have been studied extensively and used in many other applications, which makes it possible to utilize the considerable literature on these transforms.

Definition

The *moment generating function* M_X for random variable X (i.e., for its distribution) is the function

$$M_X(s) = E[e^{sX}] \text{ (s is a real or complex parameter)}$$

The *characteristic function* ϕ_X for random variable X is

$$\varphi_X(u) = E[e^{iuX}] \text{ ($i^2 = -1$, u is a real parameter)}$$

The *generating function* $g_X(s)$ for a nonnegative, integer-valued random variable X is

$$g_X(s) = E[s^X] = \sum_k s^k P(X = k)$$

The generating function $E[s^X]$ has meaning for more general random variables, but its usefulness is greatest for nonnegative, integer-valued variables, and we limit our consideration to that case.

The defining expressions display similarities which show useful relationships. We note two which are particularly useful.

$$M_X(s) = E[e^{sX}] = E[(e^s)^X] = g_X(e^s) \text{ and } \varphi_X(u) = E[e^{iuX}] = M_X(iu)$$

Because of the latter relationship, we ordinarily use the moment generating function instead of the characteristic function to avoid writing the complex unit *i*. When desirable, we convert easily by the change of variable.

The integral transform character of these entities implies that there is essentially a one-to-one relationship between the transform and the distribution.

Moments

The name and some of the importance of the moment generating function arise from the fact that the derivatives of M_X evaluateed at $s = 0$ are the moments about the origin. Specifically

$$M_X^{(k)}(0) = E[X^k], \text{ provided the } k\text{th moment exists}$$

Since expectation is an integral and because of the regularity of the integrand, we may differentiate inside the integral with respect to the parameter.

$$M_X'(s) = \frac{d}{ds}E[e^{sX}] = E[\frac{d}{ds}e^{sX}] = E[Xe^{sX}]$$

Upon setting $s = 0$, we have $M_X'(0) = E[X]$. Repeated differentiation gives the general result. The corresponding result for the characteristic function is $\varphi^{(k)}(0) = i^k E[X^k]$.

Example 13.1.1 The exponential distribution

The density function is $f_X(t) = \lambda e^{-\lambda t}$ for $t \geq 0$.

$$M_X(s) = E[e^{sX}] = \int_0^\infty \lambda e^{-(\lambda - s)t}\, dt = \frac{\lambda}{\lambda - s}$$

$$M_X'(s) = \frac{\lambda}{(\lambda - s)^2}\ M_X''(s0 = \frac{2\lambda}{(\lambda - s)^3}$$

$$E[X] = M_X'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}\ E[X^2] = M_X''(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

From this we obtain $\text{Var}[X] = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$.

The generating function does not lend itself readily to computing moments, except that

$$g_X'(s) = \sum_{k=1}^\infty k s^{k-1} P(X = k) \text{ so that } g_X'(1) = \sum_{k=1}^\infty k P(X = k) = E[X]$$

For higher order moments, we may convert the generating function to the moment generating function by replacing s with e^s, then work with M_X and its derivatives.

Example 13.1.2 The Poisson (μ) distribution

$P(X = k) = e^{-\mu}\dfrac{\mu^k}{k!}$, $k \geq 0$, so that

$$g_X(s) = e^{-\mu}\sum_{k=0}^\infty s^k \frac{\mu^k}{k!} = e^{-\mu}\sum_{k=0}^\infty \frac{(s\mu)^k}{k!} = e^{-\mu}e^{\mu s} = e^{\mu(s-1)}$$

We convert to M_X by replacing s with e^s to get $M_X(s) = e^{u(e^s - 1)}$. Then

$$M_X'(s) = e^{u(e^s-1)}\mu e^s\ M_X''(s) = e^{u(e^s-1)}[\mu^2 e^{2s} + \mu e^s]$$

so that

$$E[X] = M_X'(0) = \mu,\ E[X^2] = M_X''(0) = \mu^2 + \mu, \text{ and } \text{Var}[X] = \mu^2 + \mu - \mu^2 = \mu$$

These results agree, of course, with those found by direct computation with the distribution.

Operational properties

We refer to the following as *operational properties*.

(T1): If $Z = aX + b$, then

$$M_Z(s) = e^{bs}M_X(as),\ \varphi_Z(u) = e^{iub}\varphi_X(au),\ g_Z(s) = s^b g_X(s^a)$$

For the moment generating function, this pattern follows from

$$E[e^{(aX+b)s}] = s^{bs}E[e^{(as)X}]$$

Similar arguments hold for the other two.

(T2): If the pair $\{X, Y\}$ is independent, then

$$M_{X+Y}(s) = M_X(s)M_Y(s),\ \varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u),\ g_{X+Y}(s) = g_X(s)g_Y(s)$$

For the moment generating function, e^{sX} and e^{sY} form an independent pair for each value of the parameter s. By the product rule for expectation

$$E[e^{s(X+Y)}] = E[e^{sX}e^{sY}] = E[e^{sX}]E[e^{sY}]$$

Similar arguments are used for the other two transforms.

A partial converse for (T2) is as follows:

(T3): If $M_{X+Y}(s) = M_X(s)M_Y(s)$, then the pair $\{X + Y\}$ is uncorrelated. To show this, we obtain two expressions for $E[(X + Y)^2]$, one by direct expansion and use of linearity, and the other by taking the second derivative of the moment generating function.

$$E[(X + Y)^2] = E[X^2] + E[Y^2] + 2E[XY]$$

$$M''_{X+Y}(s) = [M_X(s)M_Y(s)]'' = M''_X(s)M_Y(s) + M_X(s)M''_Y(s) + 2M'_X(s)M'_Y(s)$$

On setting $s = 0$ and using the fact that $M_X(0) = M_Y(0) = 1$, we have

$$E[(X + Y)^2] = E[X^2] + E[Y^2] + 2E[X]E[Y]$$

which implies the equality $E[XY] = E[X]E[Y]$.

Note that we have *not* shown that being uncorrelated implies the product rule.

We utilize these properties in determining the moment generating and generating functions for several of our common distributions.

Some discrete distributions

Indicator function $X = I_E$ $P(E) = p$

$$g_X(s) = s^0q + s^1p = q + ps\ M_X(s) = g_X(e^s) = q + pe^s$$

Simple random variable $X = \sum_{i=1}^n t_iI_{A_i}$ (primitive form) $P(A_i) = p_i$

$$M_X(s) = \sum_{i=1}^n e^{st_i}p_i$$

Binomial (n, p). $X = \sum_{i=1}^n I_{E_i}$ with $\{I_{E_i} : 1 \le i \le n\}$ iid $P(E_i) = p$

We use the product rule for sums of independent random variables and the generating function for the indicator function.

$$g_X(s) = \prod_{i=1}^n (q + ps) = (q + ps)^n\ M_X(s) = (q + pe^s)^n$$

Geometric (p). $P(X = k) = pq^k\ \forall k \ge 0\ E[X] = q/p$ We use the formula for the geometric series to get

$$g_X(s) = \sum_{k=0}^\infty pq^ks^k = p\sum_{k=0}^\infty (qs)^k = \frac{p}{1 - qs}\, M_X(s) = \frac{p}{1 - qe^s}$$

Negative binomial (m, p) If Y_m is the number of the trial in a Bernoulli sequence on which the mth success occurs, and $X_m = Y_m - m$ is the number of failures before the mth success, then

$$P(X_m = k) = P(Y_m - m = k) = C(-m, k)(-q)^kp^m$$

$$\text{where } C(-m,k) = \frac{-m(-m-1)(-m-2)\cdots(-m-k+1)}{k!}$$

The power series expansion about $t = 0$ shows that

$$(1+t)^{-m} = 1 + C(-m,1)t + C(-m,2)t^2 + \cdots \text{ for } -1 < t < 1$$

Hence,

$$M_{X_m}(s) = p^m \sum_{k=0}^{\infty} C(-m,k)(-q)^k e^{sk} = [\frac{p}{1-qe^s}]^m$$

Comparison with the moment generating function for the geometric distribution shows that $X_m = Y_m - m$ has the same distribution as the sum of m iid random variables, each geometric (p). This suggests that the sequence is characterized by independent, successive waiting times to success. This also shows that the expectation and variance of X_m are m times the expectation and variance for the geometric. Thus

$$E[X_m] = mq/p \text{ and } \text{Var}[X_m] = mq/p^2$$

Poisson (μ) $P(X = k) = e^{-\mu}\frac{\mu^k}{k!}$ $\forall k \geq 0$ In Example 13.1.2, above, we establish $g_X(s) = e^{\mu(s-1)}$ and $M_X(s) = e^{\mu(e^s-1)}$. If $\{X, Y\}$ is an independent pair, with X ~ Poisson (λ) and Y ~ Poission (μ), then $Z = X + Y$ ~ Poisson $(\lambda + \mu)$. Follows from (T1) and product of exponentials.

Some absolutely continuous distributions

Uniform on $(a,b) f_X(t) = \frac{1}{b-a}$ $a < t < b$

$$M_X(s) = \int e^{st} f_X(t)\, dt = \frac{1}{b-a}\int_a^b e^{st}\, dt = \frac{e^{sb} - e^{sa}}{s(b-a)}$$

Symmetric triangular $(-c, c)$

$$f_X(t) = I_{[-c,0)}(t)\frac{c+t}{c^2} + I_{[0,c]}(t)\frac{c-t}{c^2}$$

$$M_X(s) = \frac{1}{c^2}\int_{-c}^{0}(c+t)e^{st}\, dt + \frac{1}{c^2}\int_0^c (c-t)e^{st}\, dt = \frac{e^{cs} + e^{-cs} - 2}{c^2 s^2}$$

$$= \frac{e^{cs}-1}{cs}\cdot\frac{1-e^{-cs}}{cs} = M_Y(s)M_Z(-s) = M_Y(s)M_{-Z}(s)$$

where M_Y is the moment generating function for Y ~ uniform $(0, c)$ and similarly for M_Z. Thus, X has the same distribution as the difference of two independent random variables, each uniform on $(0, c)$.

Exponential (λ) $f_X(t) = \lambda e^{-\lambda t}$, $t \geq 0$

In example 1, above, we show that $M_X(s) = \frac{\lambda}{\lambda - s}$.

Gamma(α, λ) $f_X(t) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha t^{\alpha-1}e^{-\lambda t}$ $t \geq 0$

$$M_X(s) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}e^{-(\lambda-s)t}\, dt = [\frac{\lambda}{\lambda-s}]^\alpha$$

For $\alpha = n$, a positive integer,

$$M_X(s) = [\frac{\lambda}{\lambda-s}]^n$$

which shows that in this case X has the distribution of the sum of n independent random variables each exponential (λ).

Normal (μ, σ^2).

- The standardized normal, $Z \sim N(0, 1)$

$$M_Z(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{st} e^{-t^2/2}\, dt$$

Now $st - \frac{t^2}{2} = \frac{s^2}{2} - \frac{1}{2}(t - s)^2$ so that

$$M_Z(s) = e^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t-s)^2/2}\, dt = e^{s^2/2}$$

since the integrand (including the constant $(1/\sqrt{2\pi})$ is the density for $N(s, 1)$.

- $X = \sigma Z + \mu$ implies by property (T1)

$$M_X(s) = e^{s\mu} e^{\sigma^2 s^2/2} = \exp(\frac{\sigma^2 s^2}{2} + s\mu)$$

Example 13.1.3 Affine combination of independent normal random variables

Suppose $\{X, Y\}$ is an independent pair with $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Let $Z = aX + bY + c$. The Z is normal, for by properties of expectation and variance

$$\mu_Z = a\mu_X + b\mu_Y + c \text{ and } \sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$$

and by the operational properties for the moment generating function

$$M_Z(s) = e^{sc} M_X(as) M_Y(bs) = \exp(\frac{(a^2\sigma_X^2 + b^2\sigma_Y^2)s^2}{2} + s(a\mu_X + b\mu_Y + c))$$

$$= \exp(\frac{\sigma_Z^2 s^2}{2} + s\mu_Z)$$

This form of M_Z shows that Z is normally distributed.

Moment generating function and simple random variables

Suppose $X = \sum_{i=1}^{n} t_i I_{A_i}$ in canonical form. That is, A_i is the event $\{X = t_i\}$ for each of the distinct values in the range of X_i with $p_i = P(A_i) = P(X = t_i)$. Then the moment generating function for X is

$$M_X(s) = \sum_{i=1}^{n} p_i e^{st_i}$$

The moment generating function M_X is thus related directly and simply to the distribution for random variable X.

Consider the problem of determining the sum of an independent pair $\{X, Y\}$ of simple random variables. The moment generating function for the sum is the product of the moment generating functions. Now if $Y = \sum_{j=1}^{m} u_j I_{B_j}$, with $P(Y = u_j) = \pi_j$, we have

$$M_X(s) M_Y(s) = (\sum_{i=1}^{n} p_i e^{st_i})(\sum_{j=1}^{m} \pi_j e^{su_j}) = \sum_{i,j} p_i \pi_j e^{s(t_i + u_j)}$$

The various values are sums $t_i + u_j$ of pairs (t_i, u_j) of values. Each of these sums has probability $p_i \pi_j$ for the values corresponding to t_i, u_j. Since more than one pair sum may have the same value, we need to sort the values, consolidate like values and add the probabilties for like values to achieve the distribution for the sum. We have an m-function *mgsum* for achieving this directly. It produces the pair-products for the probabilities and the pair-sums for the values, then performs a csort operation. Although not directly dependent upon the moment generating function analysis, it produces the same result as that produced by multiplying moment generating functions.

Example 13.1.4 Distribution for a sum of independent simple random variables

Suppose the pair $\{X, Y\}$ is independent with distributions

$$X = [1\ 3\ 5\ 7]\ Y = [2\ 3\ 4]\ PX = [0.2\ 0.4\ 0.3\ 0.1]\ PY = [0.3\ 0.5\ 0.2]$$

Determine the distribution for $Z = X + Y$.

```
X = [1 3 5 7];
Y = 2:4;
PX = 0.1*[2 4 3 1];
PY = 0.1*[3 5 2];
[Z,PZ] = mgsum(X,Y,PX,PY);
disp([Z;PZ]')
    3.0000    0.0600
    4.0000    0.1000
    5.0000    0.1600
    6.0000    0.2000
    7.0000    0.1700
    8.0000    0.1500
    9.0000    0.0900
   10.0000    0.0500
   11.0000    0.0200
```

This could, of course, have been achieved by using icalc and csort, which has the advantage that other functions of X and Y may be handled. Also, since the random variables are nonnegative, integer-valued, the MATLAB convolution function may be used (see Example 13.1.7). By repeated use of the function mgsum, we may obtain the distribution for the sum of more than two simple random variables. The m-functions mgsum3 and mgsum4 utilize this strategy.

The techniques for simple random variables may be used with the simple approximations to absolutely continuous random variables.

Example 13.1.5 Difference of uniform distribution

The moment generating functions for the uniform and the symmetric triangular show that the latter appears naturally as the difference of two uniformly distributed random variables. We consider X and Y iid, uniform on [0,1].

```
tappr
Enter matrix [a b] of x-range endpoints  [0 1]
Enter number of x approximation points  200
Enter density as a function of t  t<=1
Use row matrices X and PX as in the simple case
[Z,PZ] = mgsum(X,-X,PX,PX);
plot(Z,PZ/d)              % Divide by d to recover f(t)
%  plotting details   ---  see Figure 13.1.1
```

Figure one is a density graph. It is titled, Density for difference two variables, each uniform (0, 1). The horizontal axis of the graph is labeled, t, and the vertical graph is labeled fZ(t). The plot of the density is triangular, beginning at (-1, 0), and increasing at a constant slope to point (0, 1). The graph continues after this point downward with a constant negative slope to point (1, 0).

Figure 13.1.1. Density for the difference of an independent pair, uniform (0,1).

The generating function

The form of the generating function for a nonnegative, integer-valued random variable exhibits a number of important properties.

$$X = \sum_{k=0}^{\infty} kI_{A_i} \text{ (canonical form) } p_k = P(A_k) = P(X = k) \; g_X(s) = \sum_{k=0}^{\infty} s^k p_k$$

As a power series in s with nonegative coefficients whose partial sums converge to one, the series converges at least for $|s| \leq 1$.

The coefficients of the power series display the distribution: for value k the probability $p_k = P(X = k)$ is the coefficient of s^k.

The power series expansion about the origin of an analytic function is unique. If the generating function is known in closed form, the unique power series expansion about the origin determines the distribution. If the power series converges to a known closed form, that form characterizes the distribution.

For a simple random variable (i.e. $p_k = 0$ for $k > n$), g_X is a polynomial.

Example 13.1.6 The Poisson distribution

In Example 13.1.2, above, we establish the generating function for X ~ Poisson (μ) from the distribution. Suppose, however, we simply encounter the generating function

$$g_X(s) = e^{m(s-1)} = e^{-m}e^{ms}$$

From the known power series for the exponential, we get

$$g_X(s) = e^{-m} \sum_{k=0}^{\infty} \frac{(ms)^k}{k!} = e^{-m} \sum_{k=0}^{\infty} s^k \frac{m^k}{k!}$$

We conclude that

$$P(X = k) = e^{-m} \frac{m^k}{k!}, \; 0 \leq k$$

which is the Poisson distribution with parameter $\mu = m$.

For simple, nonnegative, integer-valued random variables, the generating functions are polynomials. Because of the product rule (T2), the problem of determining the distribution for the sum of independent random variables may be handled by the process of multiplying polynomials. This may be done quickly and easily with the MATLAB *convolution* function.

Example 13.1.7 Sum of independent simple random variables

Suppose the pair $\{X, Y\}$ is independent, with

$$g_X(s) = \frac{1}{10}(2 + 3s + 3s^2 + 2s^5) \; g_Y(s) = \frac{1}{10}(2s + 4s^2 + 4s^3)$$

In the MATLAB function convolution, all powers of s must be accounted for by including zeros for the missing powers.

```
gx = 0.1*[2 3 3 0 0 2];      % Zeros for missing powers 3, 4
gy = 0.1*[0 2 4 4];          % Zero  for missing power 0
gz = conv(gx,gy);
a = ['       Z         PZ'];
b = [0:8;gz]';
disp(a)
       Z         PZ          % Distribution for Z = X + Y
disp(b)
```

```
09          0         0
10     1.0000    0.0400
11     2.0000    0.1400
12     3.0000    0.2600
13     4.0000    0.2400
14     5.0000    0.1200
15     6.0000    0.0400
16     7.0000    0.0800
17     8.0000    0.0800
```

If mgsum were used, it would not be necessary to be concerned about missing powers and the corresponding zero coefficients.

Integral transforms

We consider briefly the relationship of the moment generating function and the characteristic function with well known integral transforms (hence the name of this chapter).

Moment generating function and the Laplace transform

When we examine the integral forms of the moment generating function, we see that they represent forms of the Laplace transform, widely used in engineering and applied mathematics. Suppose F_X is a probability distribution function with $F_X(-\infty) = 0$. The bilateral Laplace transform for F_X is given by

$$\int_{-\infty}^{\infty} e^{-st} F_X(t)\, dt$$

The Laplace-Stieltjes transform for F_X is

$$\int_{-\infty}^{\infty} e^{-st} F_X(dt)$$

Thus, if M_X is the moment generating function for X, then $M_X(-s)$ is the Laplace-Stieltjes transform for X (or, equivalently, for F_X).

The theory of Laplace-Stieltjes transforms shows that under conditions sufficiently general to include all practical distribution functions

$$M_X(-s) = \int_{-\infty}^{\infty} e^{-st} F_X(dt) = s \int_{-\infty}^{\infty} e^{-st} F_X(t)\, dt$$

Hence

$$\frac{1}{s} M_X(-s) = \int_{-\infty}^{\infty} e^{-st} F_X(t)\, dt$$

The right hand expression is the bilateral Laplace transform of F_X. We may use tables of Laplace transforms to recover F_X when M_X is known. This is particularly useful when the random variable X is nonnegative, so that $F_X(t) = 0$ for $t < 0$.

If X is absolutely continuous, then

$$M_X(-s) = \int_{-\infty}^{\infty} e^{-st} f_X(t)\, dt$$

In this case, $M_X(-s)$ is the bilateral Laplace transform of f_X. For nonnegative random variable X, we may use ordinary tables of the Laplace transform to recover f_X.

Example 13.1.8 Use of Laplace transform

Suppose nonnegative X has moment generating function

$$M_X(s) = \frac{1}{(1-s)}$$

We know that this is the moment generating function for the exponential (1) distribution. Now,

$$\frac{1}{s} M_X(-s) = \frac{1}{s(1+s)} = \frac{1}{s} - \frac{1}{1+s}$$

From a table of Laplace transforms, we find $1/s$ is the transform for the constant 1 (for $t \geq 0$) and $1/(1+s)$ is the transform for e^{-t}, $t \geq 0$, so that $F_X(t) = 1 - e^{-t} t \geq 0$, as expected.

Example 13.1.9 Laplace transform and the density

Suppose the moment generating function for a nonnegative random variable is

$$M_X(s) = [\frac{\lambda}{\lambda - s}]^{\alpha}$$

From a table of Laplace transforms, we find that for $\alpha > 0$.

$$\frac{\Gamma(\alpha)}{(s-a)^{\alpha}} \text{ is the Laplace transform of } t^{\alpha-1} e^{at} \; t \geq 0$$

If we put $a = -\lambda$, we find after some algebraic manipulations

$$f_X(t) = \frac{\lambda^{\alpha} t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}, \; t \geq 0$$

Thus, X ~ gamma (α, λ), in keeping with the determination, above, of the moment generating function for that distribution.

The characteristic function

Since this function differs from the moment generating function by the interchange of parameter s and iu, where i is the imaginary unit, $i^2 = -1$, the integral expressions make that change of parameter. The result is that Laplace transforms become Fourier transforms. The theoretical and applied literature is even more extensive for the characteristic function.

Not only do we have the operational properties (T1) and (T2) and the result on moments as derivatives at the origin, but there is an important expansion for the characteristic function.

An expansion theorem

If $E[[X]^n] < \infty$, then

$$\varphi^{(k)}(0) = i^k E[X^k], \text{ for } 0 \leq k \leq n \text{ and } \varphi(u) = \sum_{k=0}^{n} \frac{(iu)^k}{k!} E[X^k] + o(u^n) \text{ as } u \to 0$$

We note one limit theorem which has very important consequences.

A fundamental limit theorem

Suppose $\{F_n : 1 \leq n\}$ is a sequence of probability distribution functions and $\{\varphi_n : 1 \leq n\}$ is the corresponding sequence of characteristic functions.

> If F is a distribution function such that $F_n(t) \to F(t)$ at every point continuity for F, and ϕ is the characteristic function for F, then
>
> $$\varphi_n(u) \to \varphi(u) \; \forall u$$
>
> If $\varphi_n(u) \to \varphi(u)$ for all u and ϕ is continuous at 0, then ϕ is the characteristic function for distribution function F such that

$F_n(t) \to F(t)$ at each point of continuity of F

— □

This page titled 13.1: Transform Methods is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

13.2: Convergence and the Central Limit Theorem

The Central Limit Theorem

The *central limit theorem* (CLT) asserts that if random variable X is the sum of a large class of independent random variables, each with reasonable distributions, then X is approximately normally distributed. This celebrated theorem has been the object of extensive theoretical research directed toward the discovery of the most general conditions under which it is valid. On the other hand, this theorem serves as the basis of an extraordinary amount of applied work. In the statistics of large samples, the sample average is a constant times the sum of the random variables in the sampling process . Thus, for large samples, the sample average is approximately normal—whether or not the population distribution is normal. In much of the theory of errors of measurement, the observed error is the sum of a large number of independent random quantities which contribute additively to the result. Similarly, in the theory of noise, the noise signal is the sum of a large number of random components, independently produced. In such situations, the assumption of a normal population distribution is frequently quite appropriate.

We consider a form of the CLT under hypotheses which are reasonable assumptions in many practical situations. We sketch a proof of this version of the CLT, known as the Lindeberg-Lévy theorem, which utilizes the limit theorem on characteristic functions, above, along with certain elementary facts from analysis. It illustrates the kind of argument used in more sophisticated proofs required for more general cases.

Consider an independent sequence $\{X_n : 1 \leq n\}$ of random variables. Form the sequence of partial sums

$$S_n = \sum_{i=1}^{n} X_i \ \forall n \geq 1 \text{ with } E[S_n] = \sum_{i=1}^{n} E[X_i] \text{ and } \operatorname{Var}[S_n] = \sum_{i=1}^{n} \operatorname{Var}[X_i]$$

Let S_n^* be the standardized sum and let F_n be the distribution function for S_n^*. The CLT asserts that under appropriate conditions, $F_n(t) \to \phi(t)$ as $n \to \infty$ for all t. We sketch a proof of the theorem under the condition the X_i form an iid class.

Central Limit Theorem **(Lindeberg-Lévy form)**

If $\{X_n : 1 \leq n\}$ is iid, with

$$E[X_i] = \mu, \operatorname{Var}[X_i] = \sigma^2, \text{ and } S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

then

$$F_n(t) \to \phi(t) \text{ as } n \to \infty, \text{ for all } t$$

IDEAS OF A PROOF

There is no loss of generality in assuming $\mu = 0$. Let ϕ be the common characteristic function for the X_i, and for each n let ϕ_n be the characteristic function for S_n^*. We have

$$\varphi(t) = E[e^{itX}] \text{ and } \varphi_n(t) = E[e^{itS_n^*}] = \varphi^n(t/\sigma\sqrt{n})$$

Using the power series expansion of φ about the origin noted above, we have

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + \beta(t) \text{ where } \beta(t) = o(t^2) \text{ as } t \to 0$$

This implies

$$[\varphi(t/\sigma\sqrt{n}) - (1 - t^2/2n)] = [\beta(t/\sigma\sqrt{n})] = o(t^2/\sigma^2 n)$$

so that

$$n[\varphi(t/\sigma\sqrt{n}) - (1 - t^2/2n)] \to 0 \text{ as } n \to \infty$$

A standard lemma of analysis ensures

$(1 - \frac{t^2}{2n})^n \to e^{-t^2/2}$ as $n \to \infty$

so that

$\varphi(t/\sigma\sqrt{n}) \to e^{-t^2/2}$ as $n \to \infty$ for all t

By the convergence theorem on characteristic functions, above, $F_n(t) \to \phi(t)$.

— □

The theorem says that the distribution functions for sums of increasing numbers of the X_i converge to the normal distribution function, but it does not tell how fast. It is instructive to consider some examples, which are easily worked out with the aid of our m-functions.

Demonstration of the central limit theorem

Discrete examples

We first examine the gaussian approximation in two cases. We take the sum of five iid simple random variables in each case. The first variable has six distinct values; the second has only three. The discrete character of the sum is more evident in the second case. Here we use not only the gaussian approximation, but the gaussian approximation shifted one half unit (the so called continuity correction for integer-values random variables). The fit is remarkably good in either case with only five terms.

A principal tool is the m-function *diidsum* (sum of discrete iid random variables). It uses a designated number of iterations of mgsum.

Example 13.2.1 First random variable

```
X = [-3.2 -1.05 2.1 4.6 5.3 7.2];
PX = 0.1*[2 2 1 3 1 1];
EX = X*PX'
EX =  1.9900
VX = dot(X.^2,PX) - EX^2
VX = 13.0904
[x,px] = diidsum(X,PX,5);            % Distribution for the sum of 5 iid rv
F = cumsum(px);                      % Distribution function for the sum
stairs(x,F)                          % Stair step plot
hold on
plot(x,gaussian(5*EX,5*VX,x),'-.')   % Plot of gaussian distribution function
% Plotting details                   (see Figure 13.2.1)
```

Figure one is a distribution graph. It is titled, distribution for the sum of five iid random variables. The horizontal axis is labeled, X values, and the vertical axis is labeled PX. The values on the horizontal axis range from -20 in increments of 10 to 40. The values on the vertical axis begin at 0 and increase in increments of 0.2 to 1.4. There are two captions inside the graph. The first reads, X = [-3.2 -1.05 2.1 4.6 5.3 7.2]. The second reads, PX = 0.1*[2 2 1 3 1 1]. There are two graphs, one, a solid blue line, listed as a sum and the other a dashed and dotted line, listed as gaussian, but they both follow the same path and are nearly indistinguishable as they lay on top of one another. The path begins at the bottom-right corner of the graph. It begins completely flat, but increases in slope at an increasing rate until it is halfway across the graph, at approximately the point (10, 0). At this point, it begins decreasing its positive slope until by the far right side of the graph, approximately the point (35, 1), it has again reduced in slope enough to be a horizontal line.

Figure 13.2.1. Distribution for the sum of five iid random variables.

Example 13.2.2 Second random variable

```
X = 1:3;
PX = [0.3 0.5 0.2];
EX = X*PX'
EX = 1.9000
EX2 = X.^2*PX'
EX2 =  4.1000
VX = EX2 - EX^2
VX =  0.4900
[x,px] = diidsum(X,PX,5);            % Distribution for the sum of 5 iid rv
F = cumsum(px);                      % Distribution function for the sum
stairs(x,F)                          % Stair step plot
hold on
plot(x,gaussian(5*EX,5*VX,x),'-.')   % Plot of gaussian distribution function
plot(x,gaussian(5*EX,5*VX,x+0.5),'o')  % Plot with continuity correction
% Plotting details                   (see Figure 13.2.2)
```

Figure two is a distribution graph. It is titled, distribution for the sum of five iid random variables. The horizontal axis is labeled, X values, and the vertical axis is labeled PX. The values on the horizontal axis range from 5 in increments of 1 to 15. The values on the vertical axis begin at 0 and increase in increments of 0.2 to 1.2. There are two captions inside the graph. The first reads, X = [1 2 3]. The second reads, PX = [0.3 0.5 0.2]. There are two graphs, one, a solid blue line, labeled step dbn fn, and the other a dashed and dotted line, labeled gaussian, but they both follow the same path. The step dbn fn is a series of horizontal line segments followed by vertical line segments in varying sizes that fit the shape of the smoother curve, the gaussian curve. A third labeled item is a series of small blue circles that sit at the upper corners of the steps of the solid lined curve, labeled, shifted gaussian. The path begins at the bottom-right corner of the graph. It begins completely flat, but increases in slope at an increasing rate until it is halfway across the graph, at approximately the point (5, 0). At this point, it begins decreasing its positive slope until by the far right side of the graph, approximately the point (14, 1), it has again reduced in slope enough to be a horizontal line.

Figure 13.2.2. Distribution for the sum of five iid random variables.

As another example, we take the sum of twenty one iid simple random variables with integer values. We examine only part of the distribution function where most of the probability is concentrated. This effectively enlarges the x-scale, so that the nature of the approximation is more readily apparent.

Example 13.2.3 Sum of twenty-one iid random variables

```
X = [0 1 3 5 6];
PX = 0.1*[1 2 3 2 2];
EX = dot(X,PX)
EX =  3.3000
VX = dot(X.^2,PX) - EX^2
VX =  4.2100
[x,px] = diidsum(X,PX,21);
F = cumsum(px);
FG = gaussian(21*EX,21*VX,x);
stairs(40:90,F(40:90))
hold on
plot(40:90,FG(40:90))
% Plotting details                (see Figure 13.2.3)
```

Figure three is a distribution graph. It is titled, partial distribution for sum of 21 iid random variables. The horizontal axis is labeled, x-values, and the vertical axis is labeled PX. The values on the horizontal axis range in value from 40 to 90 at increments of 5, and the vertical axis ranges from 0 to 1 in increments of .1. There are two labeled equations. The first reads, X = [0 1 3 5 6]. The second reads, PX = 0.1*[1 2 3 2 2]. There are two graphs, one a smooth curve, labeled gaussian dbn, and the other a series of steps closely following the smooth curve, labeled Dbn for sum. Both graphs begin at the point (40, 0) at the bottom-left of the graph. The slope of the smooth curve is flat, and increases until approximately (70, 0.5). At this point, the graph continues increasing, but its slope begins decreasing until at approximately (90, 0.99), the path is again nearly flat. The steps follow the smooth curve along the same path.

Figure 13.2.3. Distribution for the sum of twenty one iid random variables.

Absolutely continuous examples

By use of the discrete approximation, we may get approximations to the sums of absolutely continuous random variables. The results on discrete variables indicate that the more values the more quickly the conversion seems to occur. In our next example, we start with a random variable uniform on (0, 1).

Example 13.2.4 Sum of three iid, uniform random variables.

Suppose X ~ uniform (0, 1). Then $E[X] = 0.5$ and $\text{Var}[X] = 1/12$.

```
tappr
Enter matrix [a b] of x-range endpoints  [0 1]
Enter number of x approximation points  100
Enter density as a function of t  t<=1
Use row matrices X and PX as in the simple case
EX = 0.5;
VX = 1/12;
[z,pz] = diidsum(X,PX,3);
F = cumsum(pz);
```

```
FG = gaussian(3*EX,3*VX,z);
length(z)
ans = 298
a = 1:5:296;                    % Plot every fifth point
plot(z(a),F(a),z(a),FG(a),'o')
% Plotting details              (see Figure 13.2.4)
```

Figure four is a distribution graph. It is titled, distribution for the sum of three iid uniform random variables. The horizontal axis is labeled, x-values, and the vertical axis is labeled PX. The values on the horizontal axis range from 0 to 3, in increments of 0.5. The values on the vertical axis range from 0 to 1, in increments of 0.1. There is one labeled statement inside the graph, that reads, X uniform on (0,1). There is one smooth curve in the graph, labeled sum, and one set of many small circles, labeled Gaussian. They follow the same path, which begins at the bottom-left at the point (0, 0). The graph begins increasing at an increasing rate until approximately the point (1.5, 0.5), where it begins increasing at a decreasing rate until it has become a flat line at the top-right of the graph, at approximately point (3, 1).

Figure 13.2.4. Distribution for the sum of three iid uniform random variables.

For the sum of only three random variables, the fit is remarkably good. This is not entirely surprising, since the sum of two gives a symmetric triangular distribution on (0, 2). Other distributions may take many more terms to get a good fit. Consider the following example.

Example 13.2.5 Sum of eight iid random variables

Suppose the density is one on the intervals (-1, -0.5) and (0.5, 1). Although the density is symmetric, it has two separate regions of probability. From symmetry. $E[X] = 0$. Calculations show $\mathrm{Var}[X] = E[X^2] = 7/12$. The MATLAB computations are:

```
tappr
Enter matrix [a b] of x-range endpoints  [-1 1]
Enter number of x approximation points  200
Enter density as a function of t  (t<=-0.5)|(t>=0.5)
Use row matrices X and PX as in the simple case
[z,pz] = diidsum(X,PX,8);
VX = 7/12;
F = cumsum(pz);
FG = gaussian(0,8*VX,z);
plot(z,F,z,FG)
% Plottting details               (see Figure 13.2.5)
```

Figure five is a distribution graph. It is titled, distribution for sum of eight iid random variables. The horizontal axis is labeled, x-values, and the vertical axis is unlabeled. The values on the horizontal axis range from -8 to 8 in increments of 2, and the values on the vertical axis range from 0 to 1 in increments of 0.1. The figure contains a second title inside the graph, which reads, Density = 1 on (-1, -0.5) and (0.5, 1). There are two plots in this figure. The first is a solid line, labeled sum. the second is a dashed, smooth line, labeled gaussian. Both follow the same general shape, except that the solid line is not as smooth, with multiple places along its plot where it is wiggly, as if it is attempting to follow the same path as the gaussian plot but does so only with some imperfection. The gaussian pot is smooth and consistent. The shape of both plots can be described as the following. The plots begin at the bottom-left corner of the graph, at point (-8, 0) and continue to the right horizontally with negligible slope, until point (-6, 0), where the plot begins increasing at an increasing rate. It does so until the midpoint in the graph, approximately (0, 0.5), where it begins to increase at a decreasing rate as it approaches the top-right corner of the graph. By approximately (6, 1) the plot continues horizontally to the top-right corner, (8, 1).

Figure 13.2.5. Distribution for the sum of eight iid uniform random variables.

Although the sum of eight random variables is used, the fit to the gaussian is not as good as that for the sum of three in Example 13.2.4. In either case, the convergence is remarkable fast—only a few terms are needed for good approximation.

Convergence phenomena in probability theory

The central limit theorem exhibits one of several kinds of convergence important in probability theory, namely convergence *in distribution* (sometimes called *weak* convergence). The increasing concentration of values of the sample average random variable $A_n$with increasing n illustrates convergence *in probability*. The convergence of the sample average is a form of the so-called *weak law of large numbers*. For large enough *n* the probability that A_n lies within a given distance of the population mean can be made as near one as desired. The fact that the variance of A_n becomes small for large *n* illustrates convergence *in the mean* (of order 2).

$$E[|A_n - \mu|^2] \to 0 \text{ as } n \to \infty$$

In the calculus, we deal with sequences of numbers. If $\{a_n : 1 \leq n\}$ s a sequence of real numbers, we say the sequence converges iff for N sufficiently large a_n approximates arbitrarily closely some number L for all $n \geq N$. This unique number L is called the limit of the sequence. Convergent sequences are characterized by the fact that for large enough N, the distance $|a_n - a_m|$ between any two terms is arbitrarily small for all n, $m \geq N$. Such a sequence is said to be *fundamental* (or *Cauchy*). To be precise, if we let $\epsilon > 0$ be the error of approximation, then the sequence is

- Convergent iff there exists a number L such that for any $\epsilon > 0$ there is an N such that

$$|L - a_n| \leq \epsilon \text{ for all } n \geq N$$

- Fundamental iff for any $\epsilon > 0$ there is an N such that

$$|a_n - a_m| \leq \epsilon \text{ for all } n, m \geq N$$

As a result of the completeness of the real numbers, it is true that any fundamental sequence converges (i.e., has a limit). And such convergence has certain desirable properties. For example the limit of a linear combination of sequences is that linear combination of the separate limits; and limits of products are the products of the limits.

The notion of convergent and fundamental sequences applies to sequences of real-valued functions with a common domain. For each x in the domain, we have a sequence

$\{f_n(x) : 1 \leq n\}$ of real numbers. The sequence may converge for some x and fail to converge for others.

A somewhat more restrictive condition (and often a more desirable one) for sequences of functions is *uniform* convergence. Here the uniformity is over values of the argument x. In this case, for any $\epsilon > 0$ there exists an N which works for all x (or for some suitable prescribed set of x).

*These concepts may be applied to a sequence of random variable*s, which are real-valued functions with domain Ω and argument ω. Suppose $\{X_n : 1 \le n\}$ is is a sequence of real random variables. For each argument ω we have a sequence $\{X_n(\omega) : 1 \le n\}$ of real numbers. It is quite possible that such a sequence converges for some ω and diverges (fails to converge) for others. As a matter of fact, in many important cases the sequence converges for all ω except possibly a set (event) of probability zero. In this case, we say the seqeunce converges *almost surely* (abbreviated a.s.). The notion of uniform convergence also applies. In probability theory we have the notion of *almost uniform* convergence. This is the case that the sequence converges uniformly for all ω except for a set of arbitrarily small probability.

The notion of convergence *in probability* noted above is a quite different kind of convergence. Rather than deal with the sequence on a pointwise basis, it deals with the random variables as such. In the case of sample average, the "closeness" to a limit is expressed in terms of the probability that the observed value $X_n(\omega)$ should lie close the the value $X(\omega)$ of the limiting random variable. We may state this precisely as follows:

A sequence $\{X_n : 1 \le n\}$ converges to *Xin probability*, designated $X_n \xrightarrow{P} X$ iff for any $\epsilon > 0$.

$$\lim_n P(|X - X_n| > \epsilon) = 0$$

There is a corresponding notion of a sequence fundamental in probability.

The following schematic representation may help to visualize the difference between almost-sure convergence and convergence in probability. In setting up the basic probability model, we think in terms of "balls" drawn from a jar or box. Instead of balls, consider for each possible outcome ω a "tape" on which there is the sequence of values $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, $\cdots$.

- If the sequence of random variable converges a.s. to a random variable X, then there is an set of "exceptional tapes" which has zero probability. For all other tapes, $X_n(\omega) \to X(\omega)$. This means that by going far enough out on *any* such tape, the values $X_n(\omega)$ beyond that point all lie within a prescribed distance of the value $X(\omega)$ of the limit random variable.
- If the sequence converges in probability, the situation may be quite different. A tape is selected. For n sufficiently large, the probability is arbitrarily near one that the observed value $X_n(\omega)$ lies within a prescribed distance of $X(\omega)$. This says nothing about the values $X_m(\omega)$ on the selected tape for any larger m. In fact, the sequence on the selected tape may very well diverge.

It is not difficult to construct examples for which there is convergence in probability but pointwise convergence for *no* ω. It is easy to confuse these two types of convergence. The kind of convergence noted for the sample average is convergence in probability (a "weak" law of large numbers). What is really desired in most cases is a.s. convergence (a "strong" law of large numbers). It turns out that for a sampling process of the kind used in simple statistics, the convergence of the sample average is almost sure (i.e., the strong law holds). To establish this requires much more detailed and sophisticated analysis than we are prepared to make in this treatment.

The notion of mean convergence illustrated by the reduction of $\mathrm{Var}[A_n]$ with increasing n may be expressed more generally and more precisely as follows. A sequence $\{X_n : 1 \le n\}$ converges in the mean of order p to X iff

$E[|X - X_n|^p] \to 0$ as $n \to \infty$ designated $X_n \xrightarrow{L^p} X$; as $n \to \infty$

If the order p is one, we simply say the sequence converges in the mean. For $p = 2$, we speak of *mean-square convergence*.

The introduction of a new type of convergence raises a number of questions.

1. There is the question of fundamental (or Cauchy) sequences and convergent sequences.
2. Do the various types of limits have the usual properties of limits? Is the limit of a linear combination of sequences the linear combination of the limits? Is the limit of products the product of the limits?
3. What conditions imply the various kinds of convergence?
4. What is the relation between the various kinds of convergence?

Before sketching briefly some of the relationships between convergence types, we consider one important condition known as *uniform integrability*. According to the property (E9b) for integrals

X is integrable iff $E[I_{\{|X_t|>a\}}|X_t|] \to 0$ as $a \to \infty$

Roughly speaking, to be integrable a random variable cannot be too large on too large a set. We use this characterization of the integrability of a single random variable to define the notion of the uniform integrability of a class.

Definition

An arbitray class $\{X_t : t \in T\}$ is *uniformly integrable* (abbreviated u.i.) with respect to probability measure P iff

$$\sup_{t\in T} E[I_{\{|X_t|>a\}}|X_t|] \to 0 \text{ as } a \to \infty$$

This condition plays a key role in many aspects of theoretical probability.

The relationships between types of convergence are important. Sometimes only one kind can be established. Also, it may be easier to establish one type which implies another of more immediate interest. We simply state informally some of the important relationships. A somewhat more detailed summary is given in PA, Chapter 17. But for a complete treatment it is necessary to consult more advanced treatments of probability and measure.

Relationships between types of convergence for probability measures

Consider a sequence $\{X_n : 1 \leq n\}$ of random variables.

It converges almost surely iff it converges almost uniformly.
If it converges almost surely, then it converges in probability.
It converges in mean, order p, iff it is uniformly integrable and converges in probability.
If it converges in probability, then it converges in distribution (i.e. weakly).

Various chains of implication can be traced. For example

- Almost sure convergence *implies* convergence in probability *implies* convergence in distribution.
- Almost sure convergence and uniform integrability *implies* convergence in mean p.

We do not develop the underlying theory. While much of it could be treated with elementary ideas, a complete treatment requires considerable development of the underlying measure theory. However, it is important to be aware of these various types of convergence, since they are frequently utilized in advanced treatments of applied probability and of statistics.

This page titled 13.2: Convergence and the Central Limit Theorem is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

13.3: Simple Random Samples and Statistics

Simple Random Samples and Statistics

We formulate the notion of a (simple) *random sample*, which is basic to much of classical statistics. Once formulated, we may apply probability theory to exhibit several basic ideas of statistical analysis.

We begin with the notion of a *population distribution*. A population may be most any collection of individuals or entities. Associated with each member is a quantity or a feature that can be assigned a number. The quantity varies throughout the population. The population distribution is the distribution of that quantity among the members of the population.

If each member could be observed, the population distribution could be determined completely. However, that is not always feasible. In order to obtain information about the population distribution, we select "at random" a subset of the population and observe how the quantity varies over the sample. Hopefully, the sample distribution will give a useful approximation to the population distribution.

The sampling process

We take a sample of size n, which means we select *n* members of the population and observe the quantity associated with each. The selection is done in such a manner that on any trial each member is equally likely to be selected. Also, the sampling is done in such a way that the result of any one selection does not affect, and is not affected by, the others. It appears that we are describing a composite trial. We model the *sampling process* as follows:

> Let X_i, $1 \le i \le n$ be the random variable for the *i*th component trial. Then the class $\{X_i : 1 \le i \le n\}$ is iid, with each member having the population distribution.

This provides a model for sampling either from a very large population (often referred to as an infinite population) or sampling with replacement from a small population.

The goal is to determine as much as possible about the character of the population. Two important parameters are the mean and the variance. We want the population mean and the population variance. If the sample is representative of the population, then the sample mean and the sample variance should approximate the population quantities.

- The *sampling process* is the iid class $\{X_i : 1 \le i \le n\}$.
- A *random sample* is an observation, or realization, $(t_1, t_2, \cdots, t_n)$ of the sampling process.

The sample average and the population mean

Consider the numerical average of the values in the sample $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} t_i$. This is an observation of the *sample average*

$$A_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n} S_n$$

The sample sum S_n and the sample average A_n are random variables. If another observation were made (another sample taken), the observed value of these quantities would probably be different. Now S_n and A_n are functions of the random variables $\{X_i : 1 \le i \le n\}$ in the sampling process. As such, they have distributions related to the *population distribution* (the common distribution of the X_i). According to the central limit theorem, for any reasonable sized sample they should be approximately normally distributed. As the examples demonstrating the central limit theorem show, the sample size need not be large in many cases. Now if the *population* mean $E[X]$ is μ and the *population variance* $\mathrm{Var}[X]$ is σ^2, then

$$E[S_n] = \sum_{i=1}^{n} E[X_i] = nE[X] = n\mu \text{ and } \mathrm{Var}[S_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = n\mathrm{Var}[X] = n\sigma^2$$

so that

$$E[A_n] = \frac{1}{n}E[S_n] = \mu \text{ and } \mathrm{Var}[A_n] = \frac{1}{n^2}\mathrm{Var}[S_n] = \sigma^2/n$$

Herein lies the key to the usefulness of a large sample. The mean of the sample average A_n is the same as the population mean, but the variance of the sample average is $1/n$ times the population variance. Thus, *for large enough sample, the probability is high that the observed value of the sample average will be close to the population mean*. The population standard deviation, as a measure of the variation is reduced by a factor $1/\sqrt{n}$.

Example 13.3.1 Sample size

Suppose a population has mean μ and variance σ^2. A sample of size n is to be taken. There are complementary questions:

1. If n is given, what is the probability the sample average lies within distance *a* from the population mean?
2. What value of n is required to ensure a probability of at least *p* that the sample average lies within distance *a* from the population mean?

Solution

Suppose the sample variance is known or can be approximated reasonably. If the sample size n is reasonably large, depending on the population distribution (as seen in the previous demonstrations), then A_n is approximately $N(\mu, \sigma^2/n)$.

1. Sample size given, probability to be determined.

$$p = P(|A_n - \mu| \le a) = P(|\frac{A_n - \mu}{\sigma/\sqrt{n}}| \le \frac{a\sqrt{n}}{\sigma} = 2\phi(a\sqrt{n}/\sigma) - 1$$

2. Sample size to be determined, probability specified.

$2\phi(a\sqrt{n}/\sigma) - 1 \ge p$ iff $\phi(a\sqrt{n}/\sigma) \ge \frac{p+1}{2}$

Find from a table or by use of the inverse normal function the value of $x = a\sqrt{n}/\sigma$ required to make $\phi(x)$ at least $(p+1)/2$. Then

$$n \ge \sigma^2(x/a)^2 = (\frac{\sigma}{a})^2 x^2$$

We may use the MATLAB function *norminv* to calculate values of x for various p.

```
p = [0.8 0.9 0.95 0.98 0.99];
x = norminv(0,1,(1+p)/2);
disp([p;x;x.^2]')
    0.8000    1.2816    1.6424
    0.9000    1.6449    2.7055
    0.9500    1.9600    3.8415
    0.9800    2.3263    5.4119
    0.9900    2.5758    6.6349
```

For $p = 0.95$, $\sigma = 2$, $a = 0.2$, $n \ge (2/0.2)^2 3.8415 = 384.15$. Use at least 385 or perhaps 400 because of uncertainty about the actual σ^2

The idea of a statistic

As a function of the random variables in the sampling process, the sample average is an example of a statistic.

Definition: statistic

A *statistic* is a function of the class $\{X_i : 1 \le i \le n\}$ which uses explicitly no unknown parameters of the population.

Example 13.3.2 Statistics as functions of the sampling progress

The random variable

$$W = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2, \text{ where } \mu = E[X]$$

is *not* a statistic, since it uses the unknown parameter μ. However, the following *is* a statistic.

$$V_n^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - A_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - A_n^2$$

It would appear that V_n^* might be a reasonable estimate of the population variance. However, the following result shows that a slight modification is desirable.

Example 13.3.3 An estimator for the population variance

The statistic

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - A_n)^2$$

is an estimator for the population variance.

VERIFICATION

Consider the statistic

$$V_n^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - A_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - A_n^2$$

Noting that $E[X^2] = \sigma^2 + \mu^2$, we use the last expression to show

$$E[V_n^*] = \frac{1}{n}n(\sigma^2 + \mu^2) - (\frac{\sigma^2}{n} + \mu^2) = \frac{n-1}{n}\sigma^2$$

The quantity has a *bias* in the average. If we consider

$$V_n = \frac{n}{n-1}V_n^* = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - A_n)^2, \text{ then } E[V_n] = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

The quantity V_n with $1/(n-1)$ rather than $1/n$ is often called the *sample variance* to distinguish it from the population variance. If the set of numbers

$$(t_1, t_2, \cdots, t_N)$$

represent the complete set of values in a population of N members, the variance for the population would be given by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} t_i^2 - (\frac{1}{N}\sum_{i=1}^{N} t_i)^2$$

Here we use $1/N$ rather than $1/(N-1)$.

Since the statistic V_n has mean value σ^2, it seems a reasonable candidate for an estimator of the population variance. If we ask how good is it, we need to consider its variance. As a random variable, it has a variance. An evaluation similar to that for the mean, but more complicated in detail, shows that

$$\text{Var}[V_n] = \frac{1}{n}(\mu_4 - \frac{n-3}{n-1}\sigma^4) \text{ where } \mu_4 = E[(X-\mu)^4]$$

For large n, $\mathrm{Var}[V_n]$ is small, so that V_n is a good large-sample estimator for σ^2.

Example 13.3.4 A sampling demonstration of the CLT

Consider a population random variable X ~ uniform [-1, 1]. Then $E[X] = 0$ and $\mathrm{Var}[X] = 1/3$. We take 100 samples of size 100, and determine the sample sums. This gives a sample of size 100 of the sample sum random variable S_{100}, which has mean zero and variance 100/3. For each observed value of the sample sum random variable, we plot the fraction of observed sums less than or equal to that value. This yields an experimental distribution function for S_{100}, which is compared with the distribution function for a random variable $Y \sim N(0, 100/3)$.

```
rand('seed',0)    % Seeds random number generator for later comparison
tappr                                        % Approximation setup
Enter matrix [a b] of x-range endpoints  [-1 1]
Enter number of x approximation points  100
Enter density as a function of t  0.5*(t<=1)
Use row matrices X and PX as in the simple case
```

```
qsample                                  % Creates sample
Enter row matrix of VALUES  X
Enter row matrix of PROBABILITIES  PX
Sample size n =  10000                   % Master sample size 10,000
Sample average ex = 0.003746
Approximate population mean E(X) = 1.561e-17
Sample variance vx = 0.3344
Approximate population variance V(X) = 0.3333
m = 100;
a = reshape(T,m,m);                      % Forms 100 samples of size 100
A = sum(a);                              % Matrix A of sample sums
[t,f] = csort(A,ones(1,m));              % Sorts A and determines cumulative
p = cumsum(f)/m;                         % fraction of elements <= each value
pg = gaussian(0,100/3,t);                % Gaussian dbn for sample sum values
plot(t,p,'k-',t,pg,'k-.')                % Comparative plot
% Plotting details                       (see Figure 13.3.1)
```

Figure one is a graph of two plots, titled, Central limit theorem for sample sums. The horizontal axis is labeled, sample sum values, and the vertical axis is labeled, cumulative fraction. The values on the horizontal axis range from -15 to 20 in increments of 5. The values on the vertical axis range from 0 to 1 in increments of 0.1. There are two captions inside the graph. The first reads, X uniform on [-1 1], and the second reads, E[X] = 0 Var[X] = 1/3. The first plot is a smooth, dashed line, labeled gaussian. The second plot is a wavering, jagged solid line labeled experimental. Both plots follow generally the same shape. They begin in the bottom right at approximately (-12, 0) with a positive slope, and they move to the right, increasing at an increasing rate. At nearly the midpoint in the graph, approximately (0, 0.5), the graphs adjust and begin increasing at a decreasing rate, approaching the top-right corner of the graph while tapering off to a horizontal line. The gaussian, dashed line follows this path's description more accurately, while the solid experimental line seems to be closely fitted to the gaussian line's path with some imperfections causing it to waver jaggedly at a couple spots along the path.

Figure 13.3.1. The central limit theorem for sample sums.

This page titled 13.3: Simple Random Samples and Statistics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

13.4: Problems on Transform Methods

Exercise 13.4.1

Calculate directly the generating function $g_X(s)$ for the geometric (p) distribution.

Answer available at the end of this section.

Exercise 13.4.2

Calculate directly the generating function $g_X(s)$ for the Poisson (μ) distribution.

Answer available at the end of this section.

Exercise 13.4.3

A projection bulb has life (in hours) represented by X ~ exponential (1/50). The unit will be replaced immediately upon failure or at 60 hours, whichever comes first. Determine the moment generating function for the time Y to replacement.

Answer available at the end of this section.

Exercise 13.4.4

Simple random variable X has distribution

$X =$ [-3 -2 0 1 4] $PX =$ [0.15 0.20 0.30 0.25 0.10]

a. Determine the moment generating function for X
b. Show by direct calculation the $M_X'(0) = E[X]$ and $M_X''(0) = E[X^2]$.

Answer available at the end of this section.

Exercise 13.4.5

Use the moment generating function to obtain the variances for the following distributions

EXponential (λ) Gamma (α, λ) Normal (μ, σ^2)

Answer available at the end of this section.

Exercise 13.4.6

The pair $\{X, Y\}$ is iid with common moment generating function $\frac{\lambda^3}{(\lambda - s)^3}$. Determine the moment generating function for $Z = 2X - 4Y + 3$.

Answer available at the end of this section.

Exercise 13.4.7

The pair $\{X, Y\}$ is iid with common moment generating function $M_X(s) = (0.6 + 0.4e^s)$. Determine the moment generating function for $Z = 5X + 2Y$.

Answer available at the end of this section.

Exercise 13.4.8

Use the moment generating function for the symmetric triangular distribution on $(-c, c)$ as derived in the section "Three Basic Transforms".

a. Obtain an expression for the symmetric triangular distribution on (a, b) for any $a < b$.
b. Use the result of part (a) to show that the sum of two independent random variables uniform on (a, b) has symmetric triangular distribution on $(2a, 2b)$.

Answer available at the end of this section.

Exercise 13.4.9

Random variable X has moment generating function $\dfrac{p^2}{(1 - qe^s)^2}$.

a. Use derivatives to determine $E[X]$ and $\text{Var}[X]$.

b. Recognize the distribution from the form and compare $E[X]$ and $\text{Var}[X]$ with the result of part (a).

Answer available at the end of this section.

Exercise 13.4.10

The pair $\{X, Y\}$ is independent. X ~ Poisson (4) and Y ~ geometric (0, 3). Determine the generating function g_Z for $Z = 3X + 2Y$.

Answer available at the end of this section.

Exercise 13.4.11

Random variable X has moment generating function

$$M_X(s) = \frac{1}{1 - 3s} \cdot \exp(16s^2/2 + 3s)$$

By recognizing forms and using rules of combinations, determine $E[X]$ and $\text{Var}[X]$.

Answer available at the end of this section.

Exercise 13.4.12

Random variable X has moment generating function

$$M_X(s) = \frac{\exp(3(e^s - 1))}{1 - 5s} \cdot \exp(16s^2/2 + 3s)$$

By recognizing forms and using rules of combinations, determine $E[X]$ and $\text{Var}[X]$.

Answer available at the end of this section.

Exercise 13.4.13

Suppose the class $\{A, B, C\}$ of events is independent, with respective probabilities 0.3, 0.5, 0.2. Consider

$$X = -3I_A + 2I_B + 4I_C$$

a. Determine the moment generating functions for and use properties of moment generating functions to determine the moment generating function for X.
b. Use the moment generating function to determine the distribution for X.
c. Use canonic to determine the distribution. Compare with result (b).
d. Use distributions for the separate terms; determine the distribution for the sum with mgsum3. Compare with result (b).

Answer available at the end of this section.

Exercise 13.4.14

Suppose the pair $\{X, Y\}$ is independent, with both X and Y binomial. Use generating functions to show under what condition, if any, $X + Y$ is binomial.

Answer available at the end of this section.

Exercise 13.4.15

Suppose the pair $\{X, Y\}$ is independent, with both X and Y Poisson.

a. Use generating functions to show under what condition $X + Y$ is Poisson.
b. What about $X - Y$? Justify your answer.

Answer available at the end of this section.

Exercise 13.4.16

Suppose the pair $\{X, Y\}$ is independent, Y is nonnegative integer-valued, X is Poisson and $X + Y$ is Poisson. Use the generating functions to show that Y is Poisson.

Answer available at the end of this section.

Exercise 13.4.17

Suppose the pair $\{X, Y\}$ is iid, binomial (6, 0.51). By the result of Exercise 13.4.14

$X + Y$ is binomial. Use mgsum to obtain the distribution for $Z = 2X + 4Y$. Does Z have the binomial distribution? Is the result surprising? Examine the first few possible values for Z. Write the generating function for Z; does it have the form for the binomial distribution?

Answer available at the end of this section.

Exercise 13.4.18

Suppose the pair $\{X, Y\}$ is independent, with X ~ binomial (5, 0.33) and Y ~ binomial (7, 0.47).

Let $G = g(X) = 3X^2 - 2X$ and $H = h(Y) = 2Y^2 + Y + 3$.

a. Use the mgsum to obtain the distribution for $G + H$.
b. Use icalc and csort to obtain the distribution for $G + H$ and compare with the result of part (a).

Answer available at the end of this section.

Exercise 13.4.19

Suppose the pair $\{X, Y\}$ is independent, with X ~ binomial (8, 0.39) and Y ~ uniform on {-1.3, -0.5, 1.3, 2.2, 3.5}. Let

$U = 3X^2 - 2X + 1$ and $V = Y^3 + 2Y - 3$

a. Use mgsum to obtain the distribution for $U + V$.
b. Use icalc and csort to obtain the distribution for $U + V$ and compare with the result of part (a).

Answer available at the end of this section.

Exercise 13.4.20

If X is a nonnegative integer-valued random variable, express the generating function as a power series.

a. Show that the kth derivative at $s = 1$ is

$$g_X^{(k)}(1) = E[X(X-1)(X-2)\cdots(X-k+1)]$$

b. Use this to show the $\text{Var}[X] = g_X''(1) + g_X'(1) - [g_X'(1)]^2$.

Answer available at the end of this section.

Exercise 13.4.21

Let $M_X(\cdot)$ be the moment generating function for X.

a. Show that $\text{Var}[X]$ is the second derivative of $e^{-s\mu}M_X(s)$ evaluated at $s = 0$.
b. Use this fact to show that X ~ $N(\mu, \sigma^2)$, then $\text{Var}[X] = \sigma^2$.

Answer available at the end of this section.

Exercise 13.4.22

Use derivatives of $M_{M_m}(s)$ to obtain the mean and variance of the negative binomial (m, p) distribution.

Answer available at the end of this section.

Exercise 13.4.23

Use moment generating functions to show that variances add for the sum or difference of independent random variables.

Answer available at the end of this section.

Exercise 13.4.24

The pair $\{X, Y\}$ is iid $N(3,5)$. Use the moment generating function to show that $Z = 2X - 2Y + 3$ is normal (see Example 3 from "Transform Methods" for general result).

Answer available at the end of this section.

Exercise 13.4.25

Use the central limit theorem to show that for large enough sample size (usually 20 or more), the sample average

$$A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is approximately $N(\mu, \sigma^2/n)$ for any reasonable population distribution having mean value μ and variance σ^2.

Answer available at the end of this section.

Exercise 13.4.26

A population has standard deviation approximately three. It is desired to determine the sample size *n* needed to ensure that with probability 0.95 the sample average will be within 0.5 of the mean value.

a. Use the Chebyshev inequality to estimate the needed sample size.
b. Use the normal approximation to estimate n (see Example 1 from "Simple Random Samples and Statistics").

Answer available at the end of this section.

Answers

Exercise 13.4.1

$$g_X(s) = E[s^2] = \sum_{k=0}^{\infty} p_k s^k = p\sum_{k=0}^{\infty} q^k s^k = \frac{p}{1-qs} \text{ (geometric series)}$$

Exercise 13.4.2

$$g_X(s) = E[s^X] = \sum_{k=0}^{\infty} p_k s^k = e^{-\mu}\sum_{k=0}^{\infty} \frac{\mu^k s^k}{k!} = e^{-\mu}e^{\mu s} = e^{\mu(s-1)}$$

Exercise 13.4.3

$$Y = I_{[0,a]}(X)X + I_{(a,\infty)}(X)a \text{ \(e^{sY} = I_{[0, a)} (X) e^{sX} + I_{(a, \infty) (X) e^{as}\)}$$
$$M_Y(s) = \int_0^a e^{st}\lambda e^{-\lambda t}\, dt + s^{sa}\int_a^{\infty} \lambda e^{-\lambda t}\, dt$$
$$= \frac{\lambda}{\lambda - s}[1 - e^{(\lambda - s)a}] + e^{-(\lambda - s)a}$$

Exercise 13.4.4

$$M_X(s) = 0.15e^{-3s} + 0.20e^{-2s} + 0.30 + 0.25e^{s} + 0.10e^{4s}$$
$$M_X'(s) = -3 \cdot 0.15e^{-3s} - 2 \cdot 0.20e^{-2s} + 0 + 0.25e^{s} + 4 \cdot 0.10e^{4s}$$
$$M_X''(s) = (-3)^2 \cdot 0.15e^{-3s} + (-2)^2 \cdot 0.20e^{-2s} + 0 + 0.25e^{s} + 4^2 \cdot 0.10e^{4s}$$

Setting $s = 0$ and using $e^0 = 1$ give the desired results.

Exercise 13.4.5

a. Exponential:

$$M_X(s) = \frac{\lambda}{\lambda - s}\ M_X'(s) = \frac{\lambda}{(\lambda - s)^2}\ M_X''(s) = \frac{2\lambda}{(\lambda - s)^3}$$

$$E[X] = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}\ E[X^2] = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}\ \mathrm{Var}[X] = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$$

b. Gamma (α, λ):

$$M_X(s) = (\frac{\lambda}{\lambda - s})^\alpha\ M_X'(s) = \alpha(\frac{\lambda}{\lambda - s})^{\alpha-1}\ \frac{\lambda}{(\lambda - s)^2} = \alpha(\frac{\lambda}{\lambda - s})^\alpha \frac{1}{\lambda - s}$$
$$M_X''(s) = \alpha^2(\frac{\lambda}{\lambda - s})^\alpha \frac{1}{\lambda - s}\ \frac{1}{\lambda - s} + \alpha(\frac{\lambda}{\lambda - s})^\alpha \frac{1}{(\lambda - s)^2}$$
$$E[X] = \frac{\alpha}{\lambda}\ E[X^2] = \frac{\alpha^2 + \alpha}{\lambda^2}\ \mathrm{Var}[X] = \frac{\alpha}{\lambda^2}$$

c. Normal(μ, σ):

$$M_X(s) = \exp(\frac{\sigma^2 s^2}{2} + \mu s)\ M_X'(s) = M_X(s) \cdot (\sigma^2 s + \mu)$$
$$M_X''(s) = M_X(s) \cdot (\sigma^2 s + \mu)^2 + M_X(s)\sigma^2$$
$$E[X] = \mu\ E[X^2] = \mu^2 + \sigma^2\ \mathrm{Var}[X] = \sigma^2$$

Exercise 13.4.6

$$M_Z(s) = e^{3s}(\frac{\lambda}{\lambda - 2s})^3(\frac{\lambda}{\lambda + 4s})^3$$

Exercise 13.4.7

\(M_Z (s) = (0.6 + 0.4e^{5s})(0.6 + 0.4e^{2s})

Exercise 13.4.8

Let $m = (a + b)/2$ and $c = (b - a)/2$. If Y ~ symetric triangular on $(-c, c)$, then $X = Y + m$ is symmetric triangular on $(m - c, m + c) = (a, b)$ and

$$M_X(s) = e^{ms}M_Y(s) = \frac{e^{cs} + e^{-cs} - 2}{c^2 s^2}e^{ms} = \frac{e^{(m+c)s} + e^{(m-c)s} - 2e^{ms}}{c^2 s^2} = \frac{e^{hs} + e^{as} - 2e^{\frac{a+b}{2}s}}{(\frac{b-a}{2})^2 s^2}$$
$$M_{X+Y}(s) = [\frac{e^{sb} - e^{sa}}{s(b - a)}]^2 = \frac{e^{s2b} + e^{s2a} - 2e^{s(b+a)}}{s^2(b - a)^2}$$

Exercise 13.4.9

$$[p^2(1 - qe^s)^{-2}]' = \frac{2p^2 qe^s}{(1 - qe^s)^3} \text{ so that } E[X] = 2q/p$$
$$[p^2(1 - qe^s)^{-2}]'' = \frac{6p^2 q^2 e^s}{(1 - qe^s)^4} + \frac{2p^2 qe^s}{(1 - qe^s)^3} \text{ so that } E[X^2] = \frac{6q^2}{p^2} + \frac{2q}{p}$$
$$\mathrm{Var}[X] = \frac{2q^2}{p^2} + \frac{2q}{p} = \frac{2(q^2 + pq)}{p^2} = \frac{2q}{p^2}$$

X ~ negative binomial $(2, p)$, which has $E[X] = 2q/p$ and $\mathrm{Var}[X] = 2q/p^2$.

Exercise 13.4.10

$$g_Z(s) = g_X(s^3)g_Y(s^2) = e^{4(s^3-1)} \cdot \frac{0.3}{1 - qs^2}$$

Exercise 13.4.11

$X = X_1 + X_2$ with X_1 ~ exponential (1/3) X_2 ~ N(3, 16)
$$E[X] = 3 + 3 = 6\ \mathrm{Var}[X] = 9 + 16 = 25$$

Exercise 13.4.12

$X = X_1 + X_2 + X_3$, with X_1 ~ Poisson (3), X_2 ~ exponential (1/5), X_3\0 \(N (3, 16)
$$E[X] = 3 + 5 + 3 = 11\ \mathrm{Var}[X] = 3 + 25 + 16 = 44$$

Exercise 13.4.13

$$M_X(s) = (0.7 + 0.3e^{-3s})(0.5 + 0.5e^{2s})(0.8 + 0.2e^{4s}) =$$
$$0.12e^{-3s} + 0.12e^{-s} + 0.28 + 0.03e^{s} + 0.28e^{2s} + 0.03e^{3s} + 0.07e^{4s} + 0.07e^{6s}$$

The distribution is

X = [-3 -1 0 1 2 3 4 6] PX = [0.12 0.12 0.28 0.03 0.28 0.03 0.07 0.07]

```
c = [-3 2 4 0];
P = 0.1*[3 5 2];
canonic
 Enter row vector of coefficients  c
 Enter row vector of minterm probabilities  minprob(P)
Use row matrices X and PX for calculations
Call for XDBN to view the distribution
P1 = [0.7 0.3];
P2 = [0.5 0.5];
P3 = [0.8 0.2];
X1 = [0 -3];
X2 = [0 2];
X3 = [0 4];
[x,px] = mgsum3(X1,X2,X3,P1,P2,P3);
```

```
disp([X;PX;x;px]')
   -3.0000    0.1200   -3.0000    0.1200
   -1.0000    0.1200   -1.0000    0.1200
         0    0.2800         0    0.2800
    1.0000    0.0300    1.0000    0.0300
    2.0000    0.2800    2.0000    0.2800
    3.0000    0.0300    3.0000    0.0300
    4.0000    0.0700    4.0000    0.0700
    6.0000    0.0700    6.0000    0.0700
```

Exercise 13.4.14

Binomial iff both have same p, as shown below.

$$g_{X+Y}(s) = (q_1 + p_1 s)^n (q_2 + p_2 s)^m = (q + ps)^{n+m} \text{ iff } p_1 = p_2$$

Exercise 13.4.15

Always Poisson, as the argument below shows.

$$g_{X+Y}(s) = e^{\mu(s-1)} e^{v(s-1)} = e^{(\mu+v)(s-1)}$$

However, $Y \sim X$ could have negative values.

Exercise 13.4.16

$E[X + Y] = \mu + v$, where $v = E[Y] > 0$, $g_X(s) = e^{\mu(s-1)}$ and \(g_{X + Y} (s) = g_X (s) g_Y (s) = e^{(\mu + s) (s - 1)\). Division by $g_X(s)$ gives $g_Y(s) = e^{v(s-1)}$.

Exercise 13.4.17

```
x  = 0:6;
px = ibinom(6,0.51,x);
[Z,PZ] = mgsum(2*x,4*x,px,px);
disp([Z(1:5);PZ(1:5)]')
         0    0.0002       % Cannot be binomial, since odd values missing
    2.0000    0.0012
    4.0000    0.0043
    6.0000    0.0118
    8.0000    0.0259
    - - - - - - - -
```

$$g_X(s) = g_Y(s) = (0.49 + 0.51s)^6 \; g_Z(s) = (0.49 + 0.51s^2)^6 (0.49 + 0.51s^4)^6$$

Exercise 13.4.18

```
X = 0:5;
Y = 0:7;
PX = ibinom(5,0.33,X);
PY = ibinom(7,0.47,Y);
G = 3*X.^2 - 2*X;
H = 2*Y.^2 + Y + 3;
[Z,PZ] = mgsum(G,H,PX,PY);


icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
M = 3*t.^2 - 2*t + 2*u.^2 + u + 3;
[z,pz] = csort(M,P);
e = max(abs(pz - PZ))  % Comparison of p values
e =  0
```

Exercise 13.4.19

```
X = 0:8;
Y = [-1.3 -0.5 1.3 2.2 3.5];
```

```
PX = ibinom(8,0.39,X);
PY = (1/5)*ones(1,5);
U  = 3*X.^2 - 2*X + 1;
V  = Y.^3 + 2*Y - 3;
[Z,PZ] = mgsum(U,V,PX,PY);
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
M = 3*t.^2 - 2*t + 1 + u.^3 + 2*u - 3;
[z,pz] = csort(M,P);
e = max(abs(pz - PZ))
e = 0
```

Exercise 13.4.20

Since power series may be differentiated term by term

$$g_X^{(n)}(s) = \sum_{k=0}^{\infty} k(k-1)\cdot(k-n+1)p_k s^{k-n} \text{ so that}$$
$$g_X^{(n)}(1) = \sum_{k=0}^{\infty} k(k-1)\cdot(k-n+1)p_k = E[X(X-1)\cdots(X-n+1)]$$
$$\text{Var}[X] = E[X^2] - E^2[X] = E[X(X-1)] + E[X] - E^2[X] = g_X''(1) + g_X'(1) - [g_X'(1)]^2$$

Exercise 13.4.21

$$f(s) = e^{-s\mu}M_X(s)\ f''(s) = e^{-s\mu}[-\mu M_X'(s) + \mu^2 M_X(s) + M_X''(s) - \mu M_X'(s)]$$

Setting $s = 0$ and using the result on moments gives

$$f''(0) = -\mu^2 + \mu^2 + E[X^2] - \mu^2 = \text{Var}[X]$$

Exercise 13.4.22

To simplify writing use $f(s)$ for $M_X(S)$.

$$f(s) = \frac{p^m}{(1-qe^s)^m}\ f'(s) = \frac{mp^m qe^s}{(1-qe^s)^{m+1}}\ f''(s) = \frac{mp^m qe^s}{1-qe^s)^{m+1}} + \frac{m(m+1)p^m q^2 e^{2s}}{1-qe^s)^{m+2}}$$
$$E[X] = \frac{mp^m q}{(1-q)^{m+1}} = \frac{mq}{p}\ E[X^2] = \frac{mq}{p} + \frac{m(m+1)p^m q^2}{(1-q)^{m+2}}$$
$$\text{Var}[X] = \frac{mq}{p} + \frac{m(m+1)q^2}{p^2} - \frac{m^2q^2}{p^2} = \frac{mq}{p^2}$$

Exercise 13.4.23

To simplify writing, set $f(s) = M_X(s)$, $g(s) = M_Y(s)$, and $h(s) = M_X(s)M_Y(s)$

$$h'(s) = f'(s)g(s) + f(s)g'(s)\ h''(s) = f''(s)g(s) + f'(s)g'(s) + f'(s)g'(s) + f(s)g''(s)$$

Setting $s = 0$ yields

$$E[X+Y] = E[X] + E[Y]\ E[(X+Y)^2] = E[X^2] + 2E[X]E[Y] + E[Y^2]\ E^2[X+Y] = E^2[X] + 2E[X]E[Y] + E^2[Y]$$

Taking the difference gives $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$. A similar treatment with $g(s)$ replaced by $g(-s)$ shows $\text{Var}[X-Y] = \text{Var}[X] + \text{Var}[Y]$.

Exercise 13.4.24

$$M_{3X}(s) = M_X(3s) = \exp(\frac{9\cdot 5s^2}{2} + 3\cdot 3s)\ M_{-2Y}(s) = M_Y(-2s) = \exp(\frac{4\cdot 5s^2}{2} - 2\cdot 3s)$$
$$M_Z(s) = e^{3s}\exp(\frac{(45+20)s^2}{2} + (9-6)s) = \exp(\frac{65s^2}{2} + 6s)$$

Exercise 13.4.25

$$E[A_n] = \frac{1}{n}\sum_{i=1}^n \mu = \mu\ \text{Var}[A_n] = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$$

By the central limit theorem, A_n is approximately normal, with the mean and variance above.

Exercise 13.4.26

Chevyshev inequality:

$$P(\frac{|A_n - \mu|}{\sigma/\sqrt{n}} \geq \frac{0.5\sqrt{n}}{3}) \leq \frac{3^2}{0.5^2 n} \leq 0.05 \text{ implies } n \geq 720$$

Normal approximation: Use of the table in Example 1 from "Simple Random Samples and Statistics" shows

$$n \geq (3/0.5)^2 3.84 = 128$$

This page titled 13.4: Problems on Transform Methods is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

14: CONDITIONAL EXPECTATION, REGRESSION

14.1: Conditional Expectation, Regression

14.2: Problems on Conditional Expectation, Regression

This page titled 14: Conditional Expectation, Regression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

14.1: Conditional Expectation, Regression

Conditional expectation, given a random vector, plays a fundamental role in much of modern probability theory. Various types of "conditioning" characterize some of the more important random sequences and processes. The notion of conditional independence is expressed in terms of conditional expectation. Conditional independence plays an essential role in the theory of Markov processes and in much of decision theory.

We first consider an elementary form of conditional expectation with respect to an event. Then we consider two highly intuitive special cases of conditional expectation, given a random variable. In examining these, we identify a fundamental property which provides the basis for a very general extension. We discover that conditional expectation is a random quantity. The basic property for conditional expectation and properties of ordinary expectation are used to obtain four fundamental properties which imply the "expectationlike" character of conditional expectation. An extension of the fundamental property leads directly to the solution of the regression problem which, in turn, gives an alternate interpretation of conditional expectation.

Conditioning by an event

If a conditioning event C occurs, we modify the original probabilities by introducing the conditional probability measure $P(\cdot|C)$. In making the change form

$$P(A) \text{ to } P(A|C) = \frac{P(AC)}{P(C)}$$

we effectively do two things:

- We limit the possible outcomes to event C
- We "normalize" the probability mass by taking $P(C)$ as the new unit

It seems reasonable to make a corresponding modification of mathematical expectation when the occurrence of event C is known. The expectation $E[X]$ is the probability weighted average of the values taken on by X. Two possibilities for making the modification are suggested.

- We could replace the prior probability measure $P(\cdot)$ with the conditional probability measure $P(\cdot|C)$ and take the weighted average with respect to these new weights.
- We could continue to use the prior probability measure $P(\cdot)$ and modify the averaging process as follows:
  - Consider the values $P(\omega)$ for only those $\omega \in C$. This may be done by using the random variable $I_C X$ which has value $X(\omega)$ for $\omega \in C$ and zero elsewhere. The expectation $E[I_C X]$ is the probability weighted *sum* of those values taken on in C.
  - The weighted average is obtained by dividing by $P(C)$.

These two approaches are equivalent. For a simple random variable $X = sum_{k=1}^{n} t_k I_{A_k}$ in canonical form

$$E[I_C X]/P(C) = \sum_{k=1}^{n} E[t_k I_C I_{A_k}]/P(C) = \sum_{k=1}^{n} t_k P(CA_k)/P(C) = \sum_{k=1}^{n} t_k P(A_k|C)$$

The final sum is expectation with respect to the conditional probability measure. Arguments using basic theorems on expectation and the approximation of general random variables by simple random variables allow an extension to a general random variable X. The notion of a conditional distribution, given C, and taking weighted averages with respect to the conditional probability is intuitive and natural in this case. However, this point of view is limited. In order to display a natural relationship with more the general concept of conditioning with repspect to a random vector, we adopt the following

Definition

The conditional expectation of X, given event C with positive probability, is the quantity

$$E[X|C] = \frac{E[I_C X]}{P(C)} = \frac{E[I_C X]}{E[I_C]}$$

Remark. The product form $E[X|C]P(C) = E[I_C X]$ is often useful.

Example 14.1.1 A numerical example

Suppose X ~ exponential (λ) and $C = \{1/\lambda \le X \le 2/\lambda\}$. Now $I_C = I_M(X)$ where $M = [1/\lambda, 2/\lambda]$.

$$P(C) = P(X \ge 1/\lambda) - P(X > 2/\lambda) = e^{-1}e^{-2} \text{ and}$$

$$E[I_C X] = \int I_M(t)t\lambda e^{-\lambda t}\, dt = \int_{1/\lambda}^{2/\lambda} t\lambda e^{-\lambda t}\, dt = \frac{1}{\lambda}(2e^{-1} - 3e^{-2})$$

Thus

$$E[X|C] = \frac{2e^{-1} - 3e^{-2}}{\lambda(e^{-1} - e^{-2})} \approx \frac{1.418}{\lambda}$$

Conditioning by a random vector—discrete case

Suppose $X = \sum_{i=1}^{n} t_i I_{A_i}$ and $Y = \sum_{j=1}^{m} u_j I_{B_j}$ in canonical form. We suppose $P(A_i) = P(X = t_i) > 0$ and $P(B_j) = P(Y = u_j) > 0$, for each permissible i, j. Now

$$P(Y = u_j|X = t_i) = \frac{P(X = t_i, Y = u_j)}{P(X = t_i)}$$

We take the expectation relative to the conditional probability $P(\cdot|X = t_i)$ to get

$$E[g(Y)|X = t_i] = \sum_{j=1}^{m} g(u_j)P(Y = u_j|X = t_i) = e(t_i)$$

Since we have a value for each t_i in the range of X, the function $e(\,\cdot\,)$ is defined on the range of X. Now consider any reasonable set M on the real line and determine the expectation

$$E[I_M(X)g(Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} I_M(t_i)g(u_j)P(X = t_i, Y = u_j)$$

$$= \sum_{i=1}^{n} I_M(t_i)[\sum_{j=1}^{m} g(u_j)P(Y = u_j|X = t_i)]P(X = t_i)$$

$$= \sum_{i=1}^{n} I_M(t_i)e(t_i)P(X = t_i) = E[I_M(X)e(X)]$$

We have the pattern

$$(A)\ E[I_M(X)g(Y)] = E[I_M(X)e(X)] \text{ where } e(t_i) = E[g(Y)|X = t_i]$$

for all t_i in the range of X.

We return to examine this property later. But first, consider an example to display the nature of the concept.

Example 14.1.2 Basic calculations and interpretation

Suppose the pair $\{X, Y\}$ has the joint distribution

$P(X = t_i, Y = u_j)$

$X =$	0	1	4	9
$Y = 2$	0.05	0.04	0.21	0.15
0	0.05	0.01	0.09	0.10

-1	0.10	0.05	0.10	0.05
PX	0.20	0.10	0.40	0.30

Calculate $E[Y|X = t_i]$ for each possible value t_i taken on by X

$$E[Y|X = 0] = -1\frac{0.10}{0.20} + 0\frac{0.05}{0.20} + 2\frac{0.05}{0.20}$$
$$= (-1 \cdot 0.10 + 0 \cdot 0.05 + 2 \cdot 0.05)/0.20 = 0$$
$$E[Y|X = 1] = (-1 \cdot 0.05 + 0 \cdot 0.01 + 2 \cdot 0.04)/0.10 = 0.30$$
$$E[Y|X = 4] = (-1 \cdot 0.10 + 0 \cdot 0.09 + 2 \cdot 0.21)/0.40 = 0.80$$
$$E[Y|X = 9] = (-1 \cdot 0.05 + 0 \cdot 0.10 + 2 \cdot 0.15)/0.10 = 0.83$$

The pattern of operation in each case can be described as follows:

- For the i th column, multiply each value u_j by $P(X = t_i, Y = u_j)$, sum, then divide by $P(X = t_i)$.

The following interpretation helps visualize the conditional expectation and points to an important result in the general case.

- For each t_i we use the mass distributed "above" it. This mass is distributed along a vertical line at values u_j taken on by Y. The result of the computation is to determine the *center of mass* for the *conditional distribution* above $t = t_i$. As in the case of ordinary expectations, this should be the best estimate, in the mean-square sense, of Y when $X = ti$. We examine that possibility in the treatment of the regression problem in Section: The regression problem.

Although the calculations are not difficult for a problem of this size, the basic pattern can be implemented simply with MATLAB, making the handling of much larger problems quite easy. This is particularly useful in dealing with the simple approximation to an absolutely continuous pair.

```
X = [0 1 4 9];             % Data for the joint distribution
Y = [-1 0 2];
P = 0.01*[ 5  4 21 15; 5  1  9 10; 10  5 10  5];
jcalc                      % Setup for calculations
Enter JOINT PROBABILITIES (as on the plane)  P
Enter row matrix of VALUES of X  X
Enter row matrix of VALUES of Y  Y
 Use array operations on matrices X, Y, PX, PY, t, u, and P
EYX = sum(u.*P)./sum(P);   % sum(P) = PX  (operation sum yields column sums)
disp([X;EYX]')             % u.*P = u_j P(X = t_i, Y = u_j) for all i, j
         0         0
    1.0000    0.3000
    4.0000    0.8000
    9.0000    0.8333
```

The calculations extend to $E[g(X, Y)|X = t_i]$. Instead of values of u_j we use values of $g(t_i, u_j)$ in the calculations. Suppose $Z = g(X, Y) = Y^2 - 2XY$.

```
G = u.^2 - 2*t.*u;         % Z = g(X,Y) = Y^2 - 2XY
EZX = sum(G.*P)./sum(P);   % E[Z|X=x]
disp([X;EZX]')
         0    1.5000
    1.0000    1.5000
    4.0000   -4.0500
    9.0000  -12.8333
```

Conditioning by a random vector — absolutely continuous case

Suppose the pair $\{X, Y\}$ has joint density function f_{XY}. We seek to use the concept of a conditional distribution, given $X = t$. The fact that $P(X = t) = 0$ for each t requires a modification of the approach adopted in the discrete case. Intuitively, we consider the *conditional density*

$$f_{Y|X}(u|t) \geq 0, \int f_{Y|X}(u|t)\ du = \frac{1}{f_X(t)} \int f_{XY}(t,u)\ du = f_X(t)/f_X(t) = 1$$

We define, in this case,

$$E[g(Y)|X = t] = \int g(u) f_{Y|X}(u|t)\ du = e(t)$$

The function $e(\cdot)$ is defined for $f_X(t) > 0$, hence effectively on the range of X. For any reasonable set M on the real line,

$$E[I_M(X)g(Y)] = \int \int I_M(t)g(u) f_{XY}(t,u)\ dudt = \int I_M(t)[\int g(u) f_{Y|X}(u|t)\ du] f_X(u)\ dt$$

$$= \int I_M(t) e(t) f_X(t)\ dt, \text{ where } e(t) = E[g(Y)|X = t]$$

Thus we have, as in the discrete case, for each t in the range of X.

$$(A)\ E[I_M(X)g(Y)] = E[I_M(X)e(X)] \text{ where } e(t) = E[g(Y)|X = t]$$

Again, we postpone examination of this pattern until we consider a more general case.

Example 14.1.3 Basic calculation and interpretation

Suppose the pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = \frac{6}{5}(t + 2u)$ on the triangular region bounded by $t = 0$, $u = 1$, and $u = t$ (see Figure 14.1.1). Then

$$f_X(t) = \frac{6}{5} \int_t^1 (t + 2u)\ du = \frac{6}{5}(1 + t - 2t^2),\ 0 \leq t \leq 1$$

By definition, then,

$$f_{Y|X}(u|t) = \frac{t + 2u}{1 + t - 2t^2} \text{ on the triangle (zero elsewhere)}$$

We thus have

$$E[Y|X = t] = \int u f_{Y|X}(u|t)\ du = \frac{1}{1 + t - 2t^2} \int_t^1 (tu + 2u^2)\ du = \frac{4 + 3t - 7t^3}{6(1 + t - 2t^2)}\ (0 \leq t < 1)$$

Theoretically, we must rule out $t = 1$ since the denominator is zero for that value of t. This causes no problem in practice.

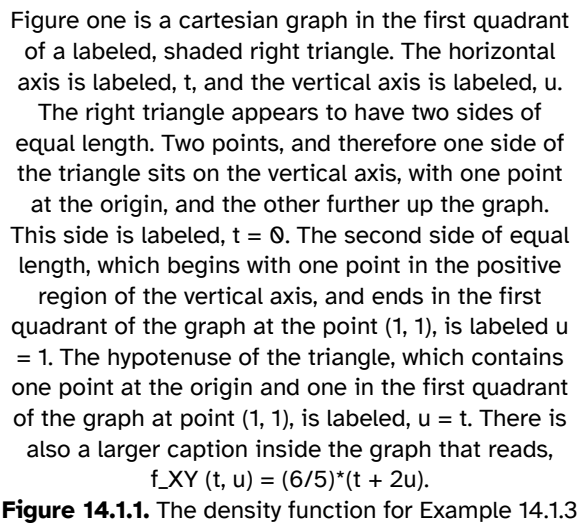


Figure 14.1.1. The density function for Example 14.1.3

We are able to make an interpretation quite analogous to that for the discrete case. This also points the way to practical MATLAB calculations.

- For any t in the range of X (between 0 and 1 in this case), consider a narrow vertical strip of width Δt with the vertical line through t at its center. If the strip is narrow enough, then $f_{XY}(t, u)$ does not vary appreciably with t for any u.
- The mass in the strip is approximately

$$\text{Mass} \approx \Delta t \int f_{XY}(t, u)\ du = \Delta t f_X(t)$$

- The moment of the mass in the strip about the line $u = 0$ is approximately

$$\text{Momemt} \approx \Delta t \int u f_{XY}(t, u)\ du$$

- The center of mass in the strip is

$$\text{Center of mass} = \frac{\text{Moment}}{\text{Mass}} \approx \frac{\Delta \int u f_{XY}(t, u)\ du}{\Delta t f_X(t)} = \int u f_{Y|X}(u|t)\ du = e(t)$$

This interpretation points the way to the use of MATLAB in approximating the conditional expectation. The success of the discrete approach in approximating the theoretical value in turns supports the validity of the interpretation. Also, this points to the general result on regression in the section, "The Regression Problem".

In the MATLAB handling of joint absolutely continuous random variables, we divide the region into narrow vertical strips. Then we deal with each of these by dividing the vertical strips to form the grid structure. The center of mass of the discrete distribution over one of the *t* chosen for the approximation must lie close to the actual center of mass of the probability in the strip. Consider the MATLAB treatment of the example under consideration.

```
f = '(6/5)*(t + 2*u).*(u>=t)';                  % Density as string variable
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  200
Enter number of Y approximation points  200
Enter expression for joint density  eval(f)     % Evaluation of string variable
Use array operations on X, Y, PX, PY, t, u, and P
EYx = sum(u.*P)./sum(P);                        % Approximate values
eYx = (4 + 3*X - 7*X.^3)./(6*(1 + X - 2*X.^2)); % Theoretical expression
plot(X,EYx,X,eYx)
% Plotting details             (see Figure 14.1.2)
```

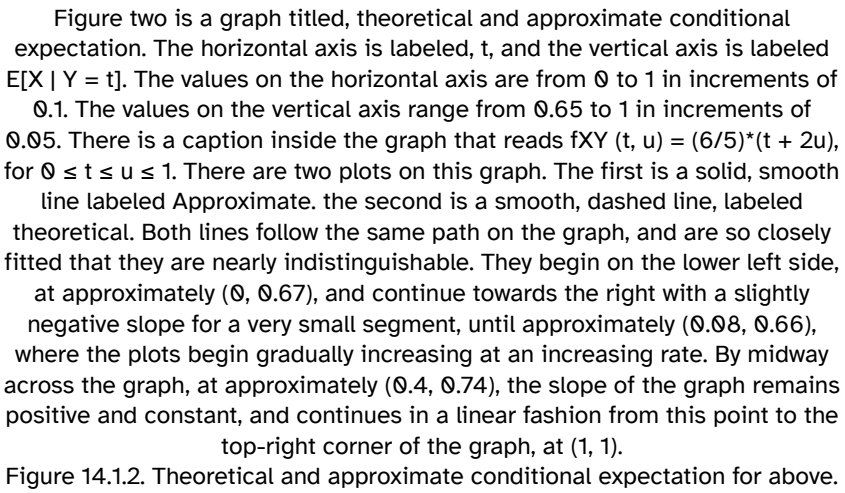


Figure 14.1.2. Theoretical and approximate conditional expectation for above.

The agreement of the theoretical and approximate values is quite good enough for practical purposes. It also indicates that the interpretation is reasonable, since the approximation determines the center of mass of the discretized mass which approximates the center of the actual mass in each vertical strip.

Extension to the general case

Most examples for which we make numerical calculations will be one of the types above. Analysis of these cases is built upon the intuitive notion of conditional distributions. However, these cases and this interpretation are rather limited and do not provide the basis for the range of applications—theoretical and practical—which characterize modern probability theory. We seek a basis for extension (which includes the special cases). In each case examined above, we have the property

$$(A)\ E[I_M(X)g(Y)] = E[I_M(X)e(X)] \text{ where } e(t) = E[g(Y)|X = t]$$

for all t in the range of X.

We have a tie to the simple case of conditioning with respect to an event. If $C = \{X \in M\}$ has positive probability, then using $I_C = I_M(X)$ we have

$$(B)\ E[I_M(X)g(Y)] = E[g(Y)|X \in M]P(X \in M)$$

wo properties of expectation are crucial here:

- By the uniqueness property (E5), since (A) holds for all reasonable (Borel) sets, then $e(X)$ is unique a.s. (i.e., except for a set of ω of probability zero).
- By the special case of the Radon Nikodym theorem (E19), the function $e(\cdot)$ *always exists* and is such that random variable $e(X)$ is unique a.s.

We make a definition based on these facts.

Definition

The *conditional expectation* $E[g(Y)|Y = t] = e(t)$ is the a.s. unique function defined on the range of X such that

$$(A)\ E[I_M(X)g(Y)] = E[I_M(X)e(X)] \text{ for all Borel sets } M$$

Note that $e(X)$ is a random variable and $e(\cdot)$ is a function. Expectation $E[g(Y)]$ is always a constant. The concept is abstract. At this point it has little apparent significance, except that it must include the two special cases studied in the previous sections. Also, it is not clear why the term conditional *expectation* should be used. The justification rests in certain formal properties which

are based on the defining condition (A) and other properties of expectation.

In Appendix F we tabulate a number of key properties of conditional expectation. The condition (A) is called property (CE1). We examine several of these properties. For a detailed treatment and proofs, any of a number of books on measure-theoretic probability may be consulted.

(CE1) **Defining condition.** $e(X) = E[g(Y)|X]$ a.s. iff

$$E[I_M(X)g(Y)] = E[I_M(X)e(X)] \text{ for each Borel set } M \text{ on the codomain of } X$$

Note that X and Y do not need to be real valued, although $g(Y)$ is real valued. This extension to possible vector valued X and Y is extremely important. The next condition is just the property (B) noted above.

(CE1a) If $P(X \in M) > 0$, then $E[I_M(X)e(X)] = E[g(Y)|X \in M]P(X \in M)$

The special case which is obtained by setting M to include the entire range of X so that $I_M(X(\omega)) = 1$ for all ω is useful in many theoretical and applied problems.

(CE1b) Law of total probability. $E[g(Y)] = E\{E[g(Y)|X]\}$

It may seem strange that we should complicate the problem of determining $E[g(Y)]$ by first getting the conditional expectation $e(X) = E[g(Y)|X]$ then taking expectation of that function. Frequently, the data supplied in a problem makes this the expedient procedure.

Exercise 14.1.4 Use of the law of total probability

Suppose the time to failure of a device is a random quantity X ~ exponential (μ), where the parameter u is the value of a parameter random variable H. Thus

$$f_{X|H}(t|u) = ue^{-ut} \text{ for } t \geq 0$$

If the parameter random variable H ~ uniform (a, b), determine the expected life $E[X]$ of the device.

Solution available at the end of this section.

The next three properties, linearity, positivity/monotonicity, and monotone convergence, along with the defining condition provide the "expectation like" character. These properties for expectation yield most of the other essential properties for expectation. A similar development holds for conditional expectation, with some reservation for the fact that $e(X)$ is a random variable, unique a.s. This restriction causes little problem for applications at the level of this treatment.

In order to get some sense of how these properties root in basic properties of expectation, we examine one of them.

(CE2) **Linearity**. For any constants a, b

$$E[ag(Y) + bh(Z)|X] = aE[g(Y)|X] + bE[h(Z)|X] \text{ a.s.}$$

VERIFICATION

Let $e_1(X) = E[g(Y)|X]$, $e_2[X] = E[h(Z)|X]$, and $e(X) = E[ag(Y) + bh(Z)|X]$ a.s.

$$\begin{aligned} E[I_M(X)e(X)] &= E\{I_M(X)[ag(Y) + bh(Z)]\} \text{ a.s.} && \text{by(CE1)} \\ &= aE[I_M(X)g(Y)] + bE[I_M(X)h(Z)] \text{ a.s.} && \text{by linearity of expectation} \\ &= aE[I_M(X)e_1(X)] + bE[I_M(X)e_2(X)] \text{ a.s.} && \text{by (CE1)} \\ &= E\{I_M(X)[ae_1(X) + be_2(X)]\} \text{ a.s.} && \text{by linearity of expectation} \end{aligned}$$

Since the equalities hold for any Borel M, the uniqueness property (E5) for expectation implies

$$e(X) = ae_1(X) = be_2(X) \text{ a.s.}$$

This is property (CE2). An *extension to any finite linear combination* is easily established by mathematical induction.

— □

Property (CE5) provides another condition for independence.

(CE5) **Independence**. $\{X, Y\}$ is an independent pair

iff $E[g(Y)|X] = E[g(Y)]$ a.s. for all Borel functions g
iff $E[I_N(Y)|X] = E[I_N(Y)]$ a.s. for all Borel sets N on the codomain of Y

Since knowledge of X does not affect the likelihood that Y will take on any set of values, then conditional expectation should not be affected by the value of X. The resulting constant value of the conditional expectation must be $E[g(Y)]$ in order for the law of total probability to hold. A formal proof utilizes uniqueness (E5) and the product rule (E18) for expectation.

Property (CE6) forms the basis for the solution of the regresson problem in the next section.

(CE6) $e(X) = E[g(Y)|X]$ a.s. iff $E[h(X)g(Y)] = E[h(X)e(X)]$ a.s. for any Borel function h

Examination shows this to be the result of replacing $I_M(X)$ in (CE1) with arbitrary $h(X)$. Again, Again, to get some insight into how the various properties arise, we sketch the ideas of a proof of (CE6).

IDEAS OF A PROOF OF (CE6)

For $h(X) = I_M(X)$, this is (CE1).

For $h(X) = \sum_{i=1}^{n} a_i I_{M_i}(X)$, the result follows by linearity.

For $h \geq 0$, $g \geq 0$, there is a seqence of nonnegative, simple $h_n nearrow h$. Now by positivity, $e(X) \geq 0$. By monotone convergence (CE4),

$$E[h_n(X)g(Y)] \nearrow E[h(X)g(Y)] \text{ and } E[h_n(X)e(X)] \nearrow E[h(X)e(X)]$$

Since corresponding terms in the sequences are equal, the limits are equal.
For $h = h^+ - h^-$, $g \geq 0$, the result follows by linearity (CE2).
For $g = g^+ - g^-$, the result again follows by linearity.

— □

Properties (CE8) and (CE9) are peculiar to conditional expectation. They play an essential role in many theoretical developments. They are essential in the study of Markov sequences and of a class of random sequences known as submartingales. We list them here (as well as in Appendix F) for reference.

(CE8) $E[h(X)g(Y)|X] = h(X)E[g(Y)|X]$ a.s. for any Borel function h

This property says that any function of the conditioning random vector may be treated as a constant factor. This combined with (CE10) below provide useful aids to computation.

(CE9) **Repeated conditioning**

If $X = h(W)$, then $E\{E[g(Y)|X|W\} = E\{E[g(Y)|W|X\} = E[g(Y)|X]$ a.s.

This somewhat formal property is highly useful in many theoretical developments. We provide an interpretation after the development of regression theory in the next section.

The next property is highly intuitive and very useful. It is easy to establish in the two elementary cases developed in previous sections. Its proof in the general case is quite sophisticated.

(CE10) Under conditions on g that are nearly always met in practice

$$E[g(X, Y)|X = t] = E[g(t, Y)|X = t] \text{ a.s. } [P_X]$$

If $\{X, Y\}$ is independent, then $E[g(X,Y)|X = t] = E[g(t,Y)]$ a.s. $[P_X]$

It certainly seem reasonable to suppose that if $X = t$, then we should be able to replace X by t in $E[g(X,Y)|X = t]$ to get $E[g(t,Y)|X = t]$. Property (CE10) assures this. If $\{X, Y\}$ is an independent pair, then the value of X should not affect the value of Y, so that $E[g(t,Y)|X = t] = E[g(t,Y)]$ a.s.

Example 14.1.5 Use of property (CE10)

Consider again the distribution for Example 14.1.3. The pair $\{X, Y\}$ has density

$f_{XY}(t,u) = \frac{6}{5}(t + 2u)$ on the triangular region bounded by $t = 0$, $u = 1$, and $u = t$

We show in Example 14.1.3 that

$$E[Y|X = t] = \frac{4 + 3t - 7t^3}{6(1 + t - 2t^2)} \quad 0 \le t < 1$$

Let $Z = 3X^2 + 2XY$. Determine $E[Z|X = t]$.

Solution

By linearity, (CE8), and (CE10)

$$E[Z|X = t] = 3t^2 + 2tE[Y|X = t] = 3t^2 + \frac{4t + 3t^2 - 7t^4}{3(1 + t - 2t^2)}$$

Conditional probability

In the treatment of mathematical expectation, we note that probability may be expressed as an expectation

$$P(E) = E[I_E]$$

For conditional probability, given an event, we have

$$E[I_E|C] = \frac{E[I_E I_C]}{P(C)} = \frac{P(EC)}{P(C)} = P(E|C)$$

In this manner, we extend the concept conditional expectation.

Definition

The *conditional probability* of event E, given X, is

$$P(E|X) = E[I_E|X]$$

Thus, there is no need for a separate theory of conditional probability. We may define the conditional distribution function

$$F_{Y|X}(u|X) = P(Y \le u|X) = E[I_{(-\infty,u]}(Y)|X]$$

Then, by the law of total probability (CE1b),

$$F_Y(u) = E[F_{Y|X}(u|X)] = \int F_{Y|X}(u|t) F_X(dt)$$

If there is a conditional density $f_{Y|X}$ such that

$$P(Y \in M|X = t) = \int_M f_{Y|X}(r|t)\, dr$$

then

$$F_{Y|X}(u|t) = \int_{-\infty}^{u} f_{Y|X}(r|t)\, dr \text{ so that } f_{Y|X}(u|t) = \frac{\partial}{\partial u} F_{Y|X}(u|t)$$

A careful, measure-theoretic treatment shows that it may *not* be true that $F_{Y|X}(\cdot|t)$ is a distribution function for all t in the range of X. However, in applications, this is seldom a problem. Modeling assumptions often start with such a family of distribution functions or density functions.

Example 14.1.6 The conditional distribution function

As in Example 14.1.4, suppose X ~ exponential (u), where the parameter u is the value of a parameter random variable H. If the parameter random variable H ~ uniform (a, b), determine the distribution fuction F_X.

Solution

As in Example 14.1.4, take the assumption on the conditional distribution to mean

$$f_{X|H}(t|u) = ue^{-ut}\ t \geq 0$$

Then

$$F_{X|H}(t|u) = \int_0^1 ue^{-us}\, ds = 1 - e^{-ut}\ 0 \leq t$$

By the law of total probability

$$F_X(t) = \int F_{X|H}(t|u) f_H(u)\, du = \frac{1}{b-a} \int_a^b (1 - e^{-ut})\, du = 1 - \frac{1}{b-a} \int_a^b e^{-ut}\, du$$

$$= 1 - \frac{1}{t(b-a)} [e^{-bt} - e^{-at}]$$

Differentiation with respect to t yields the expression for $f_X(t)$

$$f_X(t) = \frac{1}{b-a} [(\frac{1}{t^2} + \frac{b}{t}) e^{-bt} - (\frac{1}{t^2} + \frac{a}{t}) e^{-at}]\ t > 0$$

The following example uses a discrete conditional distribution and marginal distribution to obtain the joint distribution for the pair.

Example 14.1.7 A random number N of Bernoulli trials

A number N is chosen by a random selection from the integers from 1 through 20 (say by drawing a card from a box). A pair of dice is thrown N times. Let S be the number of "matches" (i.e., both ones, both twos, etc.). Determine the joint distribution for $[N, S]$.

Solution

N ~ uniform on the integers 1 through 20. $P(N = i) = 1/20$ for $1 \leq i \leq 20$. Since there are 36 pairs of numbers for the two dice and six possible matches, the probability of a match on any throw is 1/6. Since the i throws of the dice constitute a Bernoulli sequence with probability 1/6 of a success (a match), we have S conditionally binomial (i, 1/6), given $N = i$. For any pair (i, j) , $0 \leq j \leq i$,

$$P(N = i, S = j) = P(S = j|N = i)P(N = i)$$

Now $E[S|N = i) = i/6$, so that

$$E[S] = \frac{1}{6} \cdot \frac{1}{20} \sum_{i=1}^{20} i = \frac{20 \cdot 21}{6 \cdot 20 \cdot 2} = \frac{7}{4} = 1.75$$

The following MATLAB procedure calculates the joint probabilities and arranges them "as on the plane."

```
% file randbern.m
p  = input('Enter the probability of success  ');
N  = input('Enter VALUES of N  ');
PN = input('Enter PROBABILITIES for N  ');
n  = length(N);
m  = max(N);
S  = 0:m;
P  = zeros(n,m+1);
for i = 1:n
  P(i,1:N(i)+1) = PN(i)*ibinom(N(i),p,0:N(i));
end
PS = sum(P);
P  = rot90(P);
disp('Joint distribution N, S, P, and marginal PS')
randbern                          % Call for the procedure
Enter the probability of success  1/6
Enter VALUES of N  1:20
Enter PROBABILITIES for N  0.05*ones(1,20)
Joint distribution N, S, P, and marginal PS
ES = S*PS'
ES =  1.7500                          % Agrees with the theoretical value
```

The regression problem

We introduce the regression problem in the treatment of linear regression. Here we are concerned with more general regression. A pair $\{X, Y\}$ of real random variables has a joint distribution. A value $X(\omega)$ is observed. We desire a rule for obtaining the "best" estimate of the corresponding value $Y(\omega)$. If $Y(\omega)$ is the actual value and $r(X(\omega))$ is the estimate, then $Y(\omega) - r(X(\omega))$ is the error of estimate. The best estimation rule (function) $r(\cdot)$ is taken to be that for which the average square of the error is a minimum. That is, we seek a function r such that

$$E[(Y - r(X))^2] \text{ is a minimum}$$

In the treatment of linear regression, we determine the best affine function, $u = at + b$. The optimum function of this form defines the regression *line* of Y on X. We now turn to the problem of finding the best function r, which may in some cases be an affine function, but more often is not.

We have some hints of possibilities. In the treatment of expectation, we find that the best constant to approximate a random variable in the mean square sense is the mean value, which is the center of mass for the distribution. In the interpretive Example 14.2.1 for the discrete case, we find the conditional expectation $E[Y|X = t_i]$ is the center of mass for the conditional distribution at $X = t_i$. A similar result, considering thin vertical strips, is found in Example 14.1.3 for the absolutely continuous case. This suggests the possibility that $e(t) = E[Y|X = t]$ might be the best estimate for Y when the value $X(\omega) = t$ is observed. We investigate this possibility. The property (CE6) proves to be key to obtaining the result.

Let $e(X) = E[Y|X]$. We may write (CE6) in the form $E[h(X)(Y - e(X))] = 0$ for any reasonable function h. Consider

$$E[(Y - r(X))^2] = E[(Y - e(X) + e(X) - r(X))^2]$$

$$= E[(Y - e(X))^2] + E[(e(X) - r(X))^2] + 2E[(Y - e(X))(r(X) - e(X))]$$

Now $e(X)$ is fixed (a.s.) and for any choice of r we may take $h(X) = r(X) - e(X)$ to assert that

$$E[Y - e(X))(r(X) - e(X))] = E[(Y - e(X))h(X)] = 0$$

Thus

$$E[(Y - r(X))^2] = E[(Y - e(X))^2] + E[(e(X) - r(X))^2]$$

The first term on the right hand side is fixed; the second term is nonnegative, with a minimum at zero iff $r(X) = e(X)$ a.s. Thus, $r = e$ is the best rule. For a given value $X(\omega) = t$ the best mean square esitmate of Y is

$$u = e(t) = E[Y|X = t]$$

The graph of $u = e(t)$ vs t is known as the *regression curve of Y on X*. This is defined for argument t in the range of X, and is unique except possibly on a set N such that $P(X \in N) = 0$. Determination of the regression curve is thus determination of the conditional expectation.

Example 14.1.8 Regression curve for an independent pair

If the pair $\{X, Y\}$ is independent, then $u = E[Y|X = t] = E[Y]$, so that the regression curve of Y on X is the horizontal line through $u = E[Y]$. This, of course, agrees with the regression line, since $\text{Cov}[X, Y] = 0$ and the regression line is $u = 0 = E[Y]$.

The result extends to functions of X and Y. Suppose $Z = g(X, Y)$. Then the pair $\{X, Z\}$ has a joint distribution, and the best mean square estimate of Z given $X = t$ is $E[Z|X = t]$.

Example 14.1.9 Estimate of a function of $\{X, Y\}$

Suppose the pair $\{X, Y\}$ has joint density $f_{XY}(t, u) = 60t^2u$ for $0 \le t \le 1$, $0 \le u \le 1 - t$. This is the triangular region bounded by $t = 0$, $u = 0$, and $u = 1 - t$ (see Figure 14.1.3). Integration shows that

$$f_X(t) = 30t^2(1 - t)^2,\ 0 \le t \le 1 \text{ and } f_{Y|X}(u|t) = \frac{2u}{(1 - t)^2} \text{ on the triangle}$$

Consider

$$Z = \begin{cases} X^2 & \text{for } X \le 1/2 \\ 2Y & \text{for } X > 1/2 \end{cases} = I_M(X)X^2 + I_N(X)2Y$$

where M = [0, 1/2] and N = (1/2, 1]. Determine $E[Z|X = t]$.

Figure three is a cartesian graph in the first quadrant containing a large, shaded right triangle. The horizontal axis is labeled, t, and the vertical axis is labeled, u. It is labeled appropriately that both shorter sides of the triangle sit on the vertical and horizontal axes and are both of length one, with the vertex of the triangle containing the right angle sitting at the origin. The hypotenuse of the triangle, which is along a line from the point (0, 1) to the point (1, 0), is the only labeled side of the triangle, and its label reads, u = 1 - t. Inside the triangle is an equation that reads, f_xy (t, u) = 60t^2 u.

Figure 14.1.3. The density function for Example 14.1.9.

Solution By linearity and (CE8).

$$E[Z|X = t] = E[I_M(X)X^2||X = t] + E[I_N(X)2Y||X = t] = I_M(t)t^2 + I_N(t)2E[Y|X = t]$$

Now

$$E[Y|X=t] = \int u f_{Y|X}(u|t)\, du = \frac{1}{(1-t)^2}\int_0^{1-t} 2u^2\, du = \frac{2}{3}\cdot\frac{(1-t)^3}{(1-t)^2} = \frac{2}{3}(1-t),\ 0 \le t < 1$$

so that

$$E[Z|X=t] = I_M(t)t^2 + I_N(t)\frac{4}{3}(1-t)$$

Note that the indicator functions separate the two expressions. The first holds on the interval $M = [0, 1/2]$ and the second holds on the interval $N = (1/2, 1]$. The two expressions $t^2 \backslash 0 and (4/3) \backslash ((1-t)$ *must not be added*, for this would give an expression *incorrect for all t* in the range of X.

APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints  [0 1]
Enter matrix [c d] of Y-range endpoints  [0 1]
Enter number of X approximation points  100
Enter number of Y approximation points  100
Enter expression for joint density  60*t.^2.*u.*(u<=1-t)
Use array operations on X, Y, PX, PY, t, u, and P
G = (t<=0.5).*t.^2 + 2*(t>0.5).*u;
EZx = sum(G.*P)./sum(P);                          % Approximation
eZx = (X<=0.5).*X.^2 + (4/3)*(X>0.5).*(1-X);   % Theoretical
plot(X,EZx,'k-',X,eZx,'k-.')
% Plotting details                                % See Figure 14.1.4
```

The fit is quite sufficient for practical purposes, in spite of the moderate number of approximation points. The difference in expressions for the two intervals of X values is quite clear.

Figure four is a graph labeled, theoretical and approximate regression curves. The horizontal axis is labeled t, and the vertical axis is labeled E[Z | X = t]. The values on the horizontal axis range from 0 to 1 in increments of 0.1, and the vertical axis ranges in value from 0 to 0.7, in increments of 1. There are two plots on this graph. The first is a dashed line labeled Theoretical, and the second is a solid line labeled approximate. Both lines follow the same path and shape on the graph, except that the solid line is sometimes a little less smooth, wavering but still closely following the more consistent dashed line. The shape of the plot appears in three major connected sections. The first section begins at the bottom-left corner of the graph, and starts to the right with a shallow but increasing slope. The plot increases at an increasing rate until midway across the graph, at approximately (0.5, 0.25). The second section begins at this point, as the path continues vertically from (0.5, 0.25) to (0.5, 0.65). At this point, the third section begins, and is roughly linear, with a constant negative slope moving towards the bottom-right corner of the graph, where it terminates at point (1, 0).

Figure 14.1.4. Theoretical and approximate regression curves for Example 14.1.9

Example 14.1.10 Estimate of a function of $\{X, Y\}$

Suppose the pair $\{X, Y\}$ has joint density $f_{XY}(t,u) = \frac{6}{5}(t^2 + u)$, on the unit square $0 \le t \le 1$, $0 \le u \le 1$ (see Figure 14.1.5). The usual integration shows

$f_X(t) = \frac{3}{5}(2t^2 + 1)$, $0 \le t \le 1$, and $f_{Y|X}(u|t) = 2\frac{t^2 + u}{2t^2 + 1}$ on the square

Consider

$$Z = \begin{cases} 2X^2 & \text{for } X \le Y \\ 3XY & \text{for } X > Y \end{cases} I_Q(X,Y)2X^2 + I_{Q^c}(X,Y)3XY, \text{ where } Q = \{(t,u) : u \ge t\}$$

Determine $E[Z|X = t]$.

Solution

$$E[Z|X = t] = 2t^2 \int I_Q(t,u) f_{Y|X}(u|t) + 3t \int I_{Q^c}(t,u) u f_{Y|X}(u|t)\ du$$

$$= \frac{4t^2}{2t^2 + 1} \int_t^1 (t^2 + u)\ du + \frac{6t}{2t^2 + 1} \int_0^t (t^2 u + u^2)\ du = \frac{-t^5 + 4t^4 + 2t^2}{2t^2 + 1},\ 0 \le t \le 1$$

Figure five is a cartesian graph containing two equal right triangles that put together at their hypotenuse create a large square. The horizontal axis is labeled, t, and the vertical axis is labeled, u. Each axis marked only with the value 1. The points (0, 0), (0, 1), (1, 1), and (1, 0) are vertices of the square. A diagonal dashed line from point (0, 0) through point (1, 1) is labeled u = t and divides the square into two triangles. The two sides of the triangle not sitting on an axis are labeled, with the horizontal side from (0, 1) to (1, 1) labeled, u = 1, and the vertical side from (1, 0) to (1, 1) labeled, t = 1. The triangle above the diagonal line is labeled, Q, and the triangle below is labeled Q^C. A large equation is printed below the graph that reads, f_XY (t, u) = (6/5)*(t^2 + u).

Figure 14.1.5. The density and regions for Example 14.1.10

Note the different role of the indicator functions than in Example 14.1.9. There they provide a separation of two parts of the result. Here they serve to set the effective limits of integration, but sum of the two parts is needed for each t.

Figure six is a graph labeled, theoretical and approximate regression curves. The horizontal axis is labeled, t, and the vertical axis is labeled, E[Z | X = t]. The values on the horizontal axis range from 0 to 1 in increments of 0.1. The values on the vertical axis range from 0 to 1.8 in increments of 0.2. There are two plots on the graph, but both follow the same shape so closely that they are indistinguishable. One is a solid line, labeled Approximate, and the other is a dashed line, labeled Theoretical. The shape begins at the bottom-right corner of the graph at (0, 0). It initially moves to the right at a shallow positive slope. As it continues to move to the right, it begins increasing at an increasing rate until approximately (0.6, 0.7) where it maintains a constant positive slope. The plot continues this slope up to the upper-right corner of the graph, where it terminates at approximately (1, 1.65).

Figure 14.1.6. Theoretical and approximate regression curves for Example 14.1.10

APPROXIMATION

```
01 tuappr
02 Enter matrix [a b] of X-range endpoints  [0 1]
03 Enter matrix [c d] of Y-range endpoints  [0 1]
04 Enter number of X approximation points  200
```

```
Enter number of Y approximation points  200
Enter expression for joint density  (6/5)*(t.^2 + u)
Use array operations on X, Y, PX, PY, t, u, and P
G = 2*t.^2.*(u>=t) + 3*t.*u.*(u<t);
EZx = sum(G.*P)./sum(P);                          % Approximate
eZx = (-X.^5 + 4*X.^4 + 2*X.^2)./(2*X.^2 + 1);  % Theoretical
plot(X,EZx,'k-',X,eZx,'k-.')
% Plotting details                                % See Figure 14.1.4
```

The theoretical and approximate are barely distinguishable on the plot. Although the same number of approximation points are use as in Figure 14.1.4 (Example 14.1.9), the fact that the entire region is included in the grid means a larger number of effective points in this example.

Given our approach to conditional expectation, the fact that it solves the regression problem is a matter that requires proof using properties of of conditional expectation. An alternate approach is simply to *define* the conditional expectation to be the solution to the regression problem, then determine its properties. This yields, in particular, our defining condition (CE1). Once that is established, properties of expectation (including the uniqueness property (E5)) show the essential equivalence of the two concepts. There are some technical differences which do not affect most applications. The alternate approach assumes the second moment $E[X^2]$ is finite. Not all random variables have this property. However, those ordinarily used in applications at the level of this treatment will have a variance, hence a finite second moment.

We use the interpretation of $e(X) = E[g(Y)|X]$ as the best mean square estimator of $g(Y)$, given X, to interpret the formal property (CE9). We examine the special form

(CE9a) $E\{E[g(Y)|X]|X, Z\} = E\{E|g(Y)|X, Z]|X\} = E[g(Y)|X]$

Put $e_1(X, Z) = E[g(Y)|X, Z]$, the best mean square estimator of $g(Y)$, given (X, Z). Then (CE9b) can be expressed

$$E[e(X)|X, Z] = e(X) \text{ a.s. and } E[e_1(X, Z)|X] = e(X) \text{ a.s.}$$

In words, if we take the best estimate of $g(Y)$, given X, then take the best mean sqare estimate of that, given (X, Z), we do not change the estimate of $g(Y)$. On the other hand if we first get the best mean sqare estimate of $g(Y)$, given (X, Z), and then take the best mean square estimate of that, given X, we get the best mean square estimate of $g(Y)$, given X.

Answers

Exercise 14.1.4 **Use of the law of total probability**

We use the law of total probability:

$$E[X] = E\{E[X|H]\} = \int E[X|H = u] f_H(u)\, du$$

Now by assumption

$$E[X|H = u] = 1/u \text{ and } f_H(u) = \frac{1}{b-a}, a < u < b$$

Thus

$$E[X] = \frac{1}{b-a} \int_a^b \frac{1}{u} du = \frac{\ln(b/a)}{b-a}$$

For $a = 1/100$, $b = 2/100$, $E[X] = 100\ln(2) \approx 69.31$.

This page titled 14.1: Conditional Expectation, Regression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

14.2: Problems on Conditional Expectation, Regression

For the distributions in Exercises 1-3

a. Determine the regression *curve* of Y on X and compare with the regression *line* of Y on X.
b. For the function $Z = g(X, Y)$ indicated in each case, determine the regression curve of Z on X.

Exercise 14.2.1

(See Exercise 17 from "Problems on Mathematical Expectation"). The pair $\{X, Y\}$ has the joint distribution (in file npr08_07.m):

$$P(X = t, Y = u)$$

t =	-3.1	-0.5	1.2	2.4	3.7	4.9
u = 7.5	0.0090	0.0396	0.0594	0.0216	0.0440	0.0203
4.1	0.0495	0	0.1089	0.0528	0.0363	0.0231
-2.0	0.0405	0.1320	0.0891	0.0324	0.0297	0.0189
-3.8	0.0510	0.0484	0.0726	0.0132	0	0.0077

The regression line of Y on X is $u = 0.5275t + 0.6924$.

$$Z = X^2Y + |X + Y|$$

Answer available at the end of this section.

Exercise 14.2.2

(See Exercise 18 from "Problems on Mathematical Expectation"). The pair $\{X, Y\}$ has the joint distribution (in file npr08_08.m):

$$P(X = t, Y = u)$$

t =	1	3	5	7	9	11	13	15	17	19
u = 12	0.0156	0.0191	0.0081	0.0035	0.0091	0.0070	0.0098	0.0056	0.0091	0.0049
10	0.0064	0.0204	0.0108	0.0040	0.0054	0.0080	0.0112	0.0064	0.0104	0.0056
9	0.0196	0.0256	0.0126	0.0060	0.0156	0.0120	0.0168	0.0096	0.0056	0.0084
5	0.0112	0.0182	0.0108	0.0070	0.0182	0.0140	0.0196	0.0012	0.0182	0.0038
3	0.0060	0.0260	0.0162	0.0050	0.0160	0.0200	0.0280	0.0060	0.0160	0.0040
-1	0.0096	0.0056	0.0072	0.0060	0.0256	0.0120	0.0268	0.0096	0.0256	0.0084
-3	0.0044	0.0134	0.0180	0.0140	0.0234	0.0180	0.0252	0.0244	0.0234	0.0126
-5	0.0072	0.0017	0.0063	0.0045	0.0167	0.0090	0.0026	0.0172	0.0217	0.0223

The regression line of Y on X is $u = -0.2584t + 5.6110$.

$$Z = I_Q(X, Y)\sqrt{X}(Y - 4) + I_{Q^c}(X, Y)XY^2 \; Q = \{(t, u) : u \leq t\}$$

Answer available at the end of this section.

Exercise 14.2.3

(See Exercise 19 from "Problems on Mathematical Expectation"). Data were kept on the effect of training time on the time to perform a job on a production line. X is the amount of training, in hours, and Y is the time to perform the task, in minutes. The data are as follows (in file npr08_09.m):

$$P(X = t, Y = u)$$

t =	1	1.5	2	2.5	3
u = 5	0.039	0.011	0.005	0.001	0.001
4	0.065	0.070	0.050	0.015	0.010
3	0.031	0.061	0.137	0.051	0.033
2	0.012	0.049	0.163	0.058	0.039
1	0.003	0.009	0.045	0.025	0.017

The regression line of Y on X is $u = -0.7793t + 4.3051$.

$$Z = (Y - 2.8)/X$$

Answer available at the end of this section.

For the joint densities in Exercises 4-11 below

a. Determine analytically the regression *curve* of Y on X and compare with the regression *line* of Y on X.
b. Check these with a discrete approximation.

Exercise 14.2.4

(See Exercise 10 from "Problems On Random Vectors and Joint Distributions", Exercise 20 from "Problems on Mathematical Expectation", and Exercise 23 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t, u) = 1$ for $0 \leq t \leq 1$. $0 \leq u \leq 2(1 - t)$.

The regression line of Y on X is $u = 1 - t$.

$$f_X(t) = 2(1 - t),\ 0 \leq t \leq 1$$

Answer available at the end of this section.

Exercise 14.2.5

(See Exercise 13 from " Problems On Random Vectors and Joint Distributions", Exercise 23 from "Problems on Mathematical Expectation", and Exercise 24 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t, u) = \frac{1}{8}(t + u)$ for $0 \leq t \leq 2$, $0 \leq u \leq 2$.

The regression line of Y on X is $u = -t/11 + 35/33$.

$$f_X(t) = \frac{1}{4}(t + 1),\ 0 \leq t \leq 2$$

Answer available at the end of this section.

Exercise 14.2.6

(See Exercise 15 from " Problems On Random Vectors and Joint Distributions", Exercise 25 from "Problems on Mathematical Expectation", and Exercise 25 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t,u) = \frac{3}{88}(2t + 3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1 + t$.

The regression line of Y on X is $u = 0.0958t + 1.4876$.

$$f_X(t) = \frac{3}{88}(1 + t)(1 + 4t + t^2) = \frac{3}{88}(1 + 5t + 5t^2 + t^3),\ 0 \le t \le 2$$

Answer available at the end of this section.

Exercise 14.2.7

(See Exercise 16 from " Problems On Random Vectors and Joint Distributions", Exercise 26 from "Problems on Mathematical Expectation", and Exercise 26 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t,u) = 12t^2u$ on the parallelogram with vertices

(-1, 0), (0, 0), (1, 1), (0, 1)

The regression line of Y on X is $u = (4t + 5)/9$.

$$f_X(t) = I_{[-1,0]}(t)6t^2(t + 1)^2 + I_{(0,1]}(t)6t^2(1 - t^2)$$

Answer available at the end of this section.

Exercise 14.2.8

(See Exercise 17 from " Problems On Random Vectors and Joint Distributions", Exercise 27 from "Problems on Mathematical Expectation", and Exercise 27 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t,u) = \frac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \le \min\{1, 2 - t\}$.

The regression line of Y on X is $u = (-124t + 368)/431$

$$f_X(t) = I_{[0,1]}(t)\frac{12}{11}t + I_{(1,2]}(t)\frac{12}{11}t(2 - t)^2$$

Answer available at the end of this section.

Exercise 14.2.9

(See Exercise 18 from " Problems On Random Vectors and Joint Distributions", Exercise 28 from "Problems on Mathematical Expectation", and Exercise 28 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t,u) = \frac{3}{23}(t + 2u)$ for $0 \le t \le 2$, $0 \le u \le \max\{2 - t, t\}$.

The regression line of Y on X is $u = 1.0561t - 0.2603$.

$$f_X(t) = I_{[0,1]}(t)\frac{6}{23}(2 - t) + I_{(1,2]}(t)\frac{6}{23}t^2$$

Answer available at the end of this section.

Exercise 14.2.10

(See Exercise 21 from " Problems On Random Vectors and Joint Distributions", Exercise 31 from "Problems on Mathematical Expectation", and Exercise 29 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}(t,u) = \frac{2}{13}(t+2u)$ for $0 \le t \le 2$, $0 \le u \le \min\{2t, 3-t\}$.

The regression line of Y on X is $u = -0.1359t + 1.0839$.

$$f_X(t) = I_{[0,1]}(t)\frac{12}{13}t^2 + I_{(1,2]}(t)\frac{6}{13}(3-t)$$

Answer available at the end of this section.

Exercise 14.2.11

(See Exercise 22 from " Problems On Random Vectors and Joint Distributions", Exercise 32 from "Problems on Mathematical Expectation", and Exercise 30 from "Problems on Variance, Covariance, Linear Regression"). $f_{XY}9t,u) = I_{[0,1]}(t)\frac{3}{8}(t^2+2u) + I_{(1,2]}(t)\frac{9}{14}t^2u^2$. for $0 \le u \le 1$.

The regression line of Y on X is $u = 0.0817t + 0.5989$.

$$f_X(t) = I_{[0,1]}(t)\frac{3}{8}(t^2+1) + I_{(1,2]}(t)\frac{3}{14}t^2$$

Answer available at the end of this section.

For the distributions in Exercises 12-16 below

a. Determine analytically $E[Z|X=t]$
b. Use a discrete approximation to calculate the same functions.

Exercise 14.2.12

$f_{XY}(t,u) = \frac{3}{88}(2t+3u^2)$ for $0 \le t \le 2$, $0 \le u \le 1+t$, (see Exercise 37 from "Problems on Mathematical Expectation", and Exercise 14.2.6).

$$f_X(t) = \frac{3}{88}(1+t)(1+4t+t^2) = \frac{3}{88}(1+5t+5t^2+t^3),\ 0 \le t \le 2$$

$$Z = I_{[0,1]}(X)4X + I_{(1,2]}(X)(X+Y)$$

Answer available at the end of this section.

Exercise 14.2.13

$f_{XY}(t,u) = \frac{24}{11}tu$ for $0 \le t \le 2$, $0 \le u \min\{1, 2-t\}$ (see Exercise 38 from "Problems on Mathematical Expectaton", Exercise 14.2.8).

$$f_X(t) = I_{[0,1]}(t)\frac{12}{11}t + I_{(1,2]}(t)\frac{12}{11}t(2-t)^2$$

$$Z = I_M(X,Y)\frac{1}{2}X + I_M(X,Y)Y^2,\ M = \{(t,u): u > t\}$$

Answer available at the end of this section.

Exercise 14.2.14

$f_{XY}(t,u) = \frac{3}{23}(t+2u)$ for $0 \le t \le 2$, $0 \le u \le \max\{2-t, t\}$ (see Exercise 39 from "Problems on Mathematical Expectaton", and Exercise 14.2.9).

$$f_X(t) = I_{[0,1]}(t)\frac{6}{23}(2-t) + I_{(1,2]}(t)\frac{6}{23}t^2$$

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y,\ M = \{(t,u): \max(t,u) \le 1\}$$

Answer available at the end of this section.

Exercise 14.2.15

$f_{XY}(t,u) = \frac{2}{13}(t+2u)$, for $0 \le t \le 2$, $0 \le u \le \min\{2t, 3-t\}$. (see Exercise 31 from "Problems on Mathematical Expectaton", and Exercise 14.2.10).

$$f_X(t) = I_{[0,1]}(t)\frac{12}{13}t^2 + I_{(1,2]}(t)\frac{6}{13}(3-t)$$

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y^2,\ M = \{(t,u): t \le 1, u \ge 1\}$$

Answer available at the end of this section.

Exercise 14.2.16

$f_{XY}(t,u) = I_{[0,1]}(t)\frac{3}{8}(t^2+2u) + I_{(1,2]}(t)\frac{9}{14}t^2u^2$, for $0 \le u \le 1$. (see Exercise 32 from "Problems on Mathematical Expectaton", and Exercise 14.2.11).

$$f_X(t) = I_{[0,1]}(t)\frac{3}{8}(t^2+1) + I_{(1,2]}(t)\frac{3}{14}t^2$$

$$Z = I_M(X,Y)X + I_{M^c}(X,Y)XY,\ M = \{(t,u): u \le \min(1, 2-t)\}$$

Answer available at the end of this section.

Exercise 14.2.17

Suppose X ~ uniform on 0 through n and Y ~ conditionally uniform on 0 through i, given $X = i$.

a. Determine $E[Y]$ from $E[Y|X=i]$.
b. Determine the joint distribution for $\{X,Y\}$ for $n = 50$ (see Example 7 from "Conditional Expectation, Regression" for a possible approach). Use jcalc to determine $E[Y]$; compare with the theoretical value.

Answer available at the end of this section.

Exercise 14.2.18

Suppose X ~ uniform on 1 through n and Y ~ conditionally uniform on 1 through i, given $X = i$.

a. Determine $E[Y]$ from $E[Y|X = i]$.
b. Determine the joint distribution for $\{X, Y\}$ for $n = 50$ (see Example 7 from "Conditional Expectation, Regression" for a possible approach). Use jcalc to determine $E[Y]$; compare with the theoretical value.

Answer available at the end of this section.

Exercise 14.2.19

Suppose X ~ uniform on 1 through n and Y ~ conditionally binomial (i, p), given $X = i$.

a. Determine $E[Y]$ from $E[Y|X = k]$.
b. Determine the joint distribution for $\{X, Y\}$ for $n = 50$ and $p = 0.3$. Use jcalc to determine $E[Y]$; compare with the theoretical value.

Answer available at the end of this section.

Exercise 14.2.20

A number X is selected randomly from the integers 1 through 100. A pair of dice is thrown X times. Let Y be the number of sevens thrown on the X tosses. Determine the joint distribution for $\{X, Y\}$ and then determine $E[Y]$.

Answer available at the end of this section.

Exercise 14.2.21

A number X is selected randomly from the integers 1 through 100. Each of two people draw X times, independently and randomly, a number from 1 to 10. Let Y be the number of matches (i.e., both draw ones, both draw twos, etc.). Determine the joint distribution and then determine $E[Y]$.

Answer available at the end of this section.

Exercise 14.2.22

$E[Y|X = t] = 10t$ and X has density function $f_X(t) = 4 - 2t$ for $1 \le t \le 2$. Determine $E[Y]$.

Answer available at the end of this section.

Exercise 14.2.23

$E[Y|X = t] = \frac{2}{3}(1 - t)$ for $0 \le t < 1$ and X has density function $f_X(t) = 30t^2(1 - t)^2$ for $0 \le t \le 1$. Determine $E[Y]$.

Answer available at the end of this section.

Exercise 14.2.24

$E[Y|X=t]=\frac{2}{3}(2-t)$ and X has density function $f_X(t)=\frac{15}{16}t^2(2-t)^2\ 0\le t<2$. Determine $E[Y]$.

Answer available at the end of this section.

Exercise 14.2.25

Suppose the pair $\{X,Y\}$ is independent, with X ~ Poisson (μ) and Y ~ Poisson (λ). Show that X is conditionally binomial $(n,\mu/(\mu+\lambda))$, given $X+Y=n$. That is, show that

$$P(X=k|X+Y=n)=C(n,k)p^k(1-p)^{n-k},\ 0\le k\le n, \text{ for } p=\mu/(\mu+\lambda)$$

Answer available at the end of this section.

Exercise 14.2.26

Use the fact that $g(X,Y)=g^*(X,Y,Z)$, where $g^*(t,u,v)$ does not vary with v. Extend property (CE10) to show

$$E[g(X,Y)|X=t,Z=v]=E[g(t,Y)|X=t,Z=v] \text{ a.s. } [P_{XZ}]$$

Answer available at the end of this section.

Exercise 14.2.27

Use the result of Exercise 14.2.26 and properties (CE9a) and (CE10) to show that

$$E[g(X,Y)|Z=v]=\int E[g(t,Y)|X=t,Z=v]F_{X|Z}(dt|v) \text{ a.s. } [P_Z]$$

Answer available at the end of this section.

Exercise 14.2.28

A shop which works past closing time to complete jobs on hand tends to speed up service on any job received during the last hour before closing. Suppose the arrival time of a job in hours before closing time is a random variable T ~ uniform [0, 1]. Service time Y for a unit received in that period is conditionally exponential $\beta(2-u)$, given $T=u$. Determine the distribution unction for Y.

Answer available at the end of this section.

Exercise 14.2.29

Time to failure X of a manufactured unit has an exponential distribution. The parameter is dependent upon the manufacturing process. Suppose the parameter is the value of random variable H ~ uniform on[0.0005, 0.01], and X is conditionally exponential (u), given $H=u$. Determine $P(X>150)$. Determine $E[X|H=u]$ and use this to determine $E[X]$.

Answer available at the end of this section.

Exercise 14.2.30

A system has n components. Time to failure of the ith component is X_i and the class

$\{X_i : 1 \le i \le n\}$ is iid exponential (λ). The system fails if any one or more of the components fails. Let W be the time to system failure. What is the probability the failure is due to the ith component?

Suggestion. Note that $W = X_i$ iff $X_j > X_i$, for all $j \ne i$. Thus

$$\{W = X_i\} = \{(X_1, X_2, \cdots, X_n) \in Q\}, Q = \{(t_1, t_2, \cdots t_n) : t_k > t_i, \forall k \ne i\}$$

$$P(W = X_i) = E[I_Q(X_1, X_2, \cdots, X_n)] = E\{E[I_Q(X_1, X_2, \cdots, X_n)|X_i]\}$$

Answer available at the end of this section.

Answers

Exercise 14.2.1

The regression line of Y on X is $u = 0.5275t + 0.6924$.

```
npr08_07
Data are in X, Y, P
jcalc
- - - - - - - - - - -
EYx = sum(u.*P)./sum(P);
disp([X;EYx]')
   -3.1000   -0.0290
   -0.5000   -0.6860
    1.2000    1.3270
    2.4000    2.1960
    3.7000    3.8130
    4.9000    2.5700
G = t.^2.*u + abs(t+u);
EZx = sum(G.*P)./sum(P);
disp([X;EZx]')
   -3.1000    4.0383
   -0.5000    3.5345
    1.2000    6.0139
    2.4000   17.5530
    3.7000   59.7130
    4.9000   69.1757
```

Exercise 14.2.2

The regression line of Y on X is $u = -0.2584t + 5.6110$.

```
npr08_08
Data are in X, Y, P
jcalc
- - - - - - - - - - -
EYx = sum(u.*P)./sum(P);
disp([X;EYx]')
    1.0000    5.5350
    3.0000    5.9869
    5.0000    3.6500
    7.0000    2.3100
    9.0000    2.0254
   11.0000    2.9100
```

```
   13.0000    3.1957
   15.0000    0.9100
   17.0000    1.5254
   19.0000    0.9100
M = u<=t;
G = (u-4).*sqrt(t).*M + t.*u.^2.*(1-M);
EZx = sum(G.*P)./sum(P);
disp([X;EZx]')
    1.0000   58.3050
    3.0000  166.7269
    5.0000  175.9322
    7.0000  185.7896
    9.0000  119.7531
   11.0000  105.4076
   13.0000   -2.8999
   15.0000  -11.9675
   17.0000  -10.2031
   19.0000  -13.4690
```

Exercise 14.2.3

The regression line of Y on X is $u = -0.7793t + 4.3051$.

```
npr08_09
Data are in X, Y, P
jcalc
- - - - - - - - - - - -
EYx = sum(u.*P)./sum(P);
disp([X;EYx]')
    1.0000    3.8333
    1.5000    3.1250
    2.0000    2.5175
    2.5000    2.3933
    3.0000    2.3900
G = (u - 2.8)./t;
EZx = sum(G.*P)./sum(P);
disp([X;EZx]')
    1.0000    1.0333
    1.5000    0.2167
    2.0000   -0.1412
    2.5000   -0.1627
    3.0000   -0.1367
```

Exercise 14.2.4

The regression line of Y on X is $u = 1 - t$.

$$f_{Y|X}(u|t) = \frac{1}{2(1-t)}.\ 0 \le t \le 1,\ 0 \le u \le 2(1-t)$$

$$E[Y|X=t] = \frac{1}{2(1-t)} \int_0^{2(1-t)} u\,du = 1 - t,\ 0 \le t \le 1$$

```
tuappr: [0 1] [0 2] 200 400 u<=2*(1-t)
- - - - - - - - - - - - -
EYx = sum(u.*P)./sum(P);
plot(X,EYx)   % Straight line thru  (0,1), (1,0)
```

Exercise 14.2.5

The regression line of Y on X is $u = -t/11 + 35/33$.

$$f_{Y|X}(u|t) = \frac{(t+u)}{2(t+1)}\ 0 \le t \le 2,\ 0 \le u \le 2$$

$$E[Y|X=t] = \frac{1}{2(t+1)} \int_0^2 (tu + u^2)\ du = 1 + \frac{1}{3t+3}\ 0 \le t \le 2$$

```
tuappr: [0 2] [0 2] 200 200 (1/8)*(t+u)
EYx = sum(u.*P)./sum(P);
eyx = 1 + 1./(3*X+3);
plot(X,EYx,X,eyx)            % Plots nearly indistinguishable
```

Exercise 14.2.6

The regression line of Y on X is $u = 0.0958t + 1.4876$.

$$f_{Y|X}(u|t) = \frac{2t + 3u^2}{(1+t)(1+4t+t^2)} \quad 0 \le u \le 1 + t$$

$$E[Y|X = t] = \frac{1}{(1+t)(1+4t+t^2)} \int_0^{1+t} (2tu + 3u^3)\, du$$

$$= \frac{(t+1)(t+3)(3t+1)}{4(1+4t+t^2)}, \quad 0 \le t \le 2$$

```
tuappr:  [0 2] [0 3] 200 300 (3/88)*(2*t + 3*u.^2).*(u<=1+t)
EYx = sum(u.*P)./sum(P);
eyx = (X+1).*(X+3).*(3*X+1)./(4*(1 + 4*X + X.^2));
plot(X,EYx,X,eyx)            % Plots nearly indistinguishable
```

Exercise 14.2.7

The regression line of Y on X is $u = (23t + 4)/18$.

$$f_{Y|X}(u|t) = I_{[-1,0]}(t)\frac{2u}{(t+1)^2} + I_{(0,1]}(t)\frac{2u}{(1-t^2)} \text{ on the parallelogram}$$

$$E[Y|X = t] = I_{[-1,0]}(t)\frac{1}{(t+1)^2}\int_0^{t+1} 2u\, du + I_{(0,1]}(t)\frac{1}{(1-t^2)}\int_t^1 2u\, du$$

$$= I_{[-1,0]}(t)\frac{2}{3}(t+1) + I_{(0,1]}(t)\frac{2}{3}\frac{t^2+t+1}{t+1}$$

```
tuappr: [-1 1] [0 1] 200 100 12*t.^2.*u.*((u<= min(t+1,1))&(u>=max(0,t)))
EYx = sum(u.*P)./sum(P);
M = X<=0;
eyx = (2/3)*(X+1).*M + (2/3)*(1-M).*(X.^2 + X + 1)./(X + 1);
plot(X,EYx,X,eyx)            % Plots quite close
```

Exercise 14.2.8

The regression line of Y on X is $u = (-124t + 368)/431$

$$f_{Y|X}(u|t) = I_{[0,1]}(t)2u + I_{(1,2]}(t)\frac{2u}{(2-t)^2}$$

$$E[Y|X = t] = I_{[0,1]}(t)\int_0^1 2u^2\, du + I_{(1,2]}(t)\frac{1}{(2-t)^2}\int_0^{2-t} 2u^2\, du$$

$$= I_{[0,1]}(t)\frac{2}{3} + I_{(1,2]}(t)\frac{2}{3}(2-t)$$

```
tuappr: [0 2] [0 1] 200 100 (24/11)*t.*u.*(u<=min(1,2-t))
EYx = sum(u.*P)./sum(P);
M = X <= 1;
eyx = (2/3)*M + (2/3).*(2 - X).*(1-M);
plot(X,EYx,X,eyx)            % Plots quite close
```

Exercise 14.2.9

The regression line of Y on X is $u = 1.0561t - 0.2603$.

$$f_{Y|X}(u|t) = I_{[0,1]}(t)\frac{t+2u}{2(2-t)} + I_{(1,2]}(t)\frac{t+2u}{2t^2} \quad 0 \le u \le \max(2-t, t)$$

$$E[Y|X = t] = I_{[0,1]}(t)\frac{1}{2(2-t)}\int_0^{2-t}(tu + 2u^2)\, du + I_{(1,2]}(t)\frac{1}{2t^2}\int_0^t (tu + 2u^2)\, du$$

$$= I_{[0,1]}(t)\frac{1}{12}(t-2)(t-8) + I_{(1,2]}(t)\frac{7}{12}t$$

```
tuappr:  [0 2] [0 2] 200 200 (3/23)*(t+2*u).*(u<=max(2-t,t))
EYx = sum(u.*P)./sum(P);
M = X<=1;
eyx = (1/12)*(X-2).*(X-8).*M + (7/12)*X.*(1-M);
plot(X,EYx,X,eyx)             % Plots quite close
```

Exercise 14.2.10

The regression line of Y on X is $u = -0.1359t + 1.0839$.

$$f_{Y|X}(t|u) = I_{[0,1]}(t)\frac{t+2u}{6t^2} + I_{(1,2]}(t)\frac{t+2u}{3(3-t)} \quad 0 \le u \le \max(2t, 3-t)$$

$$E[Y|X=t] = I_{[0,1]}(t)\frac{1}{6t^2}\int_0^t (tu+2u^2)\,du + I_{(1,2]}(t)\frac{1}{3(3-t)}\int_0^{3-t}(tu+2u^2)\,du$$

$$= I_{[0,1]}(t)\frac{11}{9}t + I_{(1,2]}(t)\frac{1}{18}(t^2-15t+36)$$

```
tuappr: [0 2] [0 2] 200 200 (2/13)*(t+2*u).*(u<=min(2*t,3-t))
EYx = sum(u.*P)./sum(P);
M = X<=1;
eyx = (11/9)*X.*M + (1/18)*(X.^2 - 15*X + 36).*(1-M);
plot(X,EYx,X,eyx)              % Plots quite close
```

Exercise 14.2.11

The regression line of Y on X is $u = 0.0817t + 0.5989$.

$$f_{Y|X}(t|u) = I_{[0,1]}(t)\frac{t^2+2u}{t^2+1} + I_{(1,2]}(t)3u^2 \quad 0 \le u \le 1$$

$$E[Y|X=t] = I_{[0,1]}(t)\frac{1}{t^2+1}\int_0^1 (t^2u+2u^2)\,du + I_{(1,2]}(t)\int_0^1 3u^3\,du$$

$$= I_{[0,1]}(t)\frac{3t^2+4}{6(t^2+1)} + I_{(1,2]}(t)\frac{3}{4}$$

```
tuappr: [0 2] [0 1] 200 100 (3/8)*(t.^2 + 2*u).*(t<=1) + ...
       (9/14)*t.^2.*u.^2.*(t>1)
EYx = sum(u.*P)./sum(P);
M = X<=1;
eyx = M.*(3*X.^2 + 4)./(6*(X.^2 + 1)) + (3/4)*(1 - M);
plot(X,EYx,X,eyx)              % Plots quite close
```

Exercise 14.2.12

$Z = I_M(X)4X + I_N(X)(X+Y)$. Use of linearity, (CE8), and (CE10) gives

$$E[Z|X=t] = I_M(t)4t + I_N(t)(t + E[Y|X=t])$$

$$= I_M(t)4t + I_N(t)(t + \frac{(t+1)(t+3)(3t+1)}{4(1+4t+t^2)})$$

```
% Continuation of Exercise 14.2.6
G = 4*t.*(t<=1) + (t + u).*(t>1);
EZx = sum(G.*P)./sum(P);
M = X<=1;
ezx = 4*X.*M + (X + (X+1).*(X+3).*(3*X+1)./(4*(1 + 4*X + X.^2))).*(1-M);
plot(X,EZx,X,ezx)              % Plots nearly indistinguishable
```

Exercise 14.2.13

$$Z = I_M(X,Y)\frac{1}{2}X + I_{M^c}(X,Y)Y^2,\ M = \{(t,u): u > t\}$$

$$I_M(t,u) = I_{[0,1]}(t)I_{[t,1]}(u)\ \ I_{M^c}(t,u) = I_{[0,1]}(t)I_{[0,t]}(u) + I_{(1,2]}(t)I_{[0,2-t]}(u)$$

$$E[Z|X=t] = I_{[0,1]}(t)[\frac{t}{2}\int_t^1 2u\,du + \int_0^t u^2\cdot 2u\,du] + I_{(1,2]}(t)\int_0^{2-t} u^2\cdot\frac{2u}{(2-t)^2}\,du$$

$$= I_{[0,1]}(t)\frac{1}{2}t(1-t^2+t^3) + I_{(1,2]}(t)\frac{1}{2}(2-t)^2$$

```
% Continuation of Exercise 14.2.8
Q = u>t;
G = (1/2)*t.*Q + u.^2.*(1-Q);
EZx = sum(G.*P)./sum(P);
M = X <= 1;
ezx = (1/2)*X.*(1-X.^2+X.^3).*M + (1/2)*(2-X).^2.*(1-M);
plot(X,EZx,X,ezx)              % Plots nearly indistinguishable
```

Exercise 14.2.14

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y,\ M = \{(t,u): \max(t,u) \le 1\}$$

$$I_M(t,u) = I_{[0,1]}(t)I_{[0,1]}(u)\ \ I_{M^c}(t,u) = I_{[0,1]}(t)I_{[1,2-t]}(u) + I_{(1,2]}(t)I_{[0,1]}(u)$$

$$E[Z|X=t] = I_{[0,1]}(t)\frac{1}{2(2-t)}\int_0^1 (t+u)(t+2u)\,du + \frac{1}{2-t}\int_1^{2-t} u(t+2u)\,du] + I_{(1,2]}(t)2E[Y|X=t]$$

$$= I_{[0,1]}(t)\frac{1}{12}\cdot\frac{2t^3-30t^2+69t-60}{t-2} + I_{(1,2]}(t)\frac{7}{6}2t$$

```
% Continuation of Exercise 14.2.9
M = X <= 1;
Q = (t<=1)&(u<=1);
G = (t+u).*Q + 2*u.*(1-Q);
EZx = sum(G.*P)./sum(P);
ezx = (1/12)*M.*(2*X.^3 - 30*X.^2 + 69*X -60)./(X-2) + (7/6)*X.*(1-M);
plot(X,EZx,X,ezx)
```

Exercise 14.2.15

$$Z = I_M(X,Y)(X+Y) + I_{M^c}(X,Y)2Y^2,\ M = \{(t,u): t \le 1, u \ge 1\}$$
$$I_M(t,u) = I_{[0,1]}(t0I_{[1,2]}(u)\ I_{M^c}(t,u) = I_{[0,1]}(t)I_{[0,1)}(u) + I_{(1,2]}(t)I_{[0,3-t]}(u)$$
$$E[Z|X=t] = I_{[0,1/2]}(t)\frac{1}{6t^2}\int_0^{2t} 2u^2(t+2u)\ du +$$
$$I_{(1/2,1]}(t)[\frac{1}{6t^2}\int_0^1 2u^2(t+2u)\ du + \frac{1}{6t^2}\int_1^{2t}(t+u)(t+2u)\ du] + I_{(1,2]}(t)\frac{1}{3(3-t)}\int_0^{3-t} 2u^2(t+2u)\ du$$
$$= I_{[0,1/2]}(t)\frac{32}{9}t^2 + I_{(1/2,1]}(t)\frac{1}{36}\cdot\frac{80t^3 - 6t^2 - 5t + 2}{t^2} + I_{(1,2]}(t)\frac{1}{9}(-t^3 + 15t^2 - 63t + 81)$$

```
tuappr:  [0 2] [0 2] 200 200 (2/13)*(t + 2*u).*(u<=min(2*t,3-t))
M = (t<=1)&(u>=1);
Q = (t+u).*M + 2*(1-M).*u.^2;
EZx = sum(Q.*P)./sum(P);
N1 = X <= 1/2;
N2 = (X > 1/2)&(X<=1);
N3 = X > 1;
ezx = (32/9)*N1.*X.^2 + (1/36)*N2.*(80*X.^3 - 6*X.^2 - 5*X + 2)./X.^2 ...
         + (1/9)*N3.*(-X.^3 + 15*X.^2 - 63.*X + 81);
plot(X,EZx,X,ezx)
```

Exercise 14.2.16

$$Z = I_M(X,Y)X + I_{M^c}(X,Y)XY,\ M = \{(t,u): u \le \min\ (1, 2-t)\}$$
$$E[|X=t] = I_{[0,1]}(t)\int_0^1 \frac{t^3 + 2tu}{t^2+1}\ du + I_{(1,2]}(t)[\int_0^{2-t} 3tu^2\ du + \int_{2-t}^1 3tu^3\ du]$$
$$= I_{[0,1]}(t)t + I_{(1,2]}(t)(-\frac{13}{4}t + 12t^2 - 12t^3 + 5t^4 - \frac{3}{4}t^5)$$

```
tuappr:  [0 2] [0 1] 200 100  (t<=1).*(t.^2 + 2*u)./(t.^2 + 1) +3*u.^2.*(t>1)
M = u<=min(1,2-t);
G = M.*t + (1-M).*t.*u;
EZx = sum(G.*P)./sum(P);
N = X<=1;
ezx = X.*N + (1-N).*(-(13/4)*X + 12*X.^2 - 12*X.^3 + 5*X.^4 - (3/4)*X.^5);
plot(X,EZx,X,ezx)
```

Exercise 14.2.17

a. $E[Y|X=i] = i/2$, so

$$E[Y] = \sum_{i=0}^n E[Y|X=i]P(X=i) = \frac{1}{n+1}\sum_{i=1}^n i/2 = n/4$$

b. $P(X=i) = 1/(n+1)$, $0 \le i \le n$, $P(Y=k|X=i) = 1/(i+1)$. $0 \le k \le i$; hence
$P(X=i, Y=k) = 1/(n+1)(i+1)$, $0 \le i \le n$, $0 \le k \le i$.

```
n = 50; X = 0:n; Y = 0:n; P0 = zeros(n+1,n+1); for i = 0:n P0(i+1,1:i+1) = (1/((n+1)*(i+1)))*ones(1,i+1); end P = rot90(P0); jcalc: X Y P - - - - - - - - - - - - EY = dot(Y,PY) EY = 12.5000 % Comparison with part (a): 50/4 = 12.5
```

Exercise 14.2.18

a. $E[Y|X=i] = (i+1)/2$, so

$$E[Y] = \sum_{i=0}^n E[Y|X=i]P(X=i) = \frac{1}{n+1}\sum_{i=1}^n \frac{i+1}{2} = \frac{n+3}{4}$$

b. $P(X=i) = 1/n$, $1 \le i \le n$, $P(Y=k|X=i) = 1/i$. $1 \le k \le i$; hence
$P(X=i, Y=k) = 1/ni$, $1 \le i \le n$, $1 \le k \le i$.

```
n = 50; X = 1:n; Y = 1:n; P0 = zeros(n,n); for i = 1:n P0(i,1:i) = (1/(n*i))*ones(1,i); end P = rot90(P0); jcalc: P X Y - - - - - - - - - - - - EY = dot(Y,PY) EY = 13.2500 % Comparison with part (a): 53/4 = 13.25
```

Exercise 14.2.19

a. $E[Y|X=i] = ip$, so

$$E[Y] = \sum_{i=1}^n E[Y|X=i]P(X=i) = \frac{p}{n}\sum_{i=1}^n i = \frac{p(n+1)}{2}$$

b. $P(X=i) = 1/n$, $1 \le i \le n$, $P(Y=k|X=i)$ = ibinom$(i, p, 0:i)$, $0 \le k \le i$.

```
n = 50; p = 0.3; X = 1:n; Y = 0:n; P0 = zeros(n,n+1); % Could use randbern for i = 1:n P0(i,1:i+1) = (1/n)*ibinom(i,p,0:i); end P = rot90(P0); jcalc: X Y P - - - - - - - - - - - EY = dot(Y,PY) EY = 7.6500 % Comparison with part (a): 0.3*51/2 = 0.765
```

Exercise 14.2.20

a. $P(X = i) = 1/n$, $E[Y|X = i] = i/6$, so

$$E[Y] = \frac{1}{6}\sum_{i=0}^{n} i/n = \frac{(n+1)}{12}$$

b.
```
n = 100; p = 1/6; X = 1:n; Y = 0:n; PX = (1/n)*ones(1,n); P0 = zeros(n,n+1); % Could use randbern for i = 1:n P0(i,1:i+1) = (1/n)*ibinom(i,p,0:i); end P = rot90(P0); jcalc EY = dot(Y,PY) EY = 8.4167 % Comparison with part (a): 101/12 = 8.4167
```

Exercise 14.2.21

Same as Exercise 14.2.20, except $p = 1/10$. $E[Y] = (n+1)/20$

```
n = 100; p = 0.1; X = 1:n; Y = 0:n; PX = (1/n)*ones(1,n);
P0 = zeros(n,n+1);         % Could use randbern
for i = 1:n
  P0(i,1:i+1) = (1/n)*ibinom(i,p,0:i);
end
P = rot90(P0);
jcalc
- - - - - - - - - - -
EY = dot(Y,PY)
EY =  5.0500                    % Comparison with part (a): EY = 101/20 = 5.05
```

Exercise 14.2.22

$$E[Y] = \int E[Y|X = t]f_X(t)\, dt = \int_1^2 10t(4-2t)\, dt = 40/3$$

Exercise 14.2.23

$$E[Y] = \int E[Y|X = t]f_X(t)\, dt = \int_0^1 20t^2(1-t)^3\, dt = 1/3$$

Exercise 14.2.24

$$E[Y] = \int E[Y|X = t]f_X(t)\, dt = \frac{5}{8}\int_0^2 t^2(2-t)^3\, dt = 2/3$$

Exercise 14.2.25

X ~ Poisson (μ), Y ~ Poisson (λ). Use of property (T1) and generating functions shows that $X + Y$ ~ Poisson $(\mu + \lambda)$

$$P(X = k|X + Y = n) = \frac{P(X = k, X + Y = n)}{P(X + Y = n)} = \frac{P(X = k, Y = n - k)}{P(X + Y) = n}$$
$$= \frac{e^{-\mu}\frac{\mu^k}{k!}e^{-\lambda}\frac{\lambda^{n-k}}{(n-k)!}}{e^{-(\mu+\lambda)}\frac{(\mu+\lambda)^n}{n!}} = \frac{n!}{k!(n-k)!}\frac{\mu^k\lambda^{n-k}}{(\mu+\lambda)^n}$$

Put $p = \mu/(\mu + \lambda)$ and $q = 1 - p = \lambda/(\mu + \lambda)$ to get the desired result.

Exercise 14.2.26

$$E[g(X,Y)|X = t, Z = v] = E[g^*(X,Z,Y)|(X,Z) = (t,v)] = E[g^*(t,v,Y)|(X,Z) = (t,v)]$$
$$= E[g(t,Y)|X = t, Z = v] \text{ a.s. } [P_{XZ}] \text{ by (CE10)}$$

Exercise 14.2.27

By (CE9), $E[g(X,Y)|Z] = E\{E[g(X,Y)|X,Z]|Z\} = E[e(X,Z)|Z]$ a.s.

By (CE10),

$$E[e(X,Z)|Z = v] = E[e(X,v)|Z = v] =$$
$$\int e(t,v)F_{X|Z}(dt|v) \text{ a.s.}$$

By Exercise 14.2.26,

$$\int E[g(X,Y)|X = t, Z = v]F_{X|Z}(dt|v) =$$
$$\int E[g(t,Y)|X = t, Z = v]F_{X|Z}(dt|v) \text{ a.s. } [P_Z]$$

Exercise 14.2.28

$$F_Y(v) = \int F_{Y|T}(v|u)f_T(u)\, du = \int_0^1 (1 - e^{-\beta(2-u)v})\, du =$$
$$1 - e^{-2\beta v}\frac{e^{\beta v} - 1}{\beta v} = 1 - e^{\beta v}[\frac{1 - e^{-\beta v}}{\beta v}],\ 0 < v$$

Exercise 14.2.29

$$F_{X|H}(t|u) = 1 - e^{ut}\ f_H(u) = \frac{1}{0.05} = 200,\ 0.005 \le u \le 0.01$$
$$F_X(t) = 1 - 200\int_{0.005}^{0.01} e^{-ut}\, du = 1 - \frac{200}{t}[e^{-0.005t} - e^{-0.01t}]$$
$$P(X > 150) = \frac{200}{150}[e^{-0.75} - e^{-1.5}] \approx 0.3323$$
$$E[X|H = u] = 1/u\ E[X] = 200\int_{0.005}^{0.01}\frac{du}{u} = 200\ln 2$$

Exercise 14.2.30

Let $Q = \{(t_1, t_2, \cdots, t_n) : t_k > t_i, k \neq i\}$. Then

$$P(W = X_i) = E[I_Q(X_1, X_2, \cdots, X_n)] = E\{E[I_Q(X_1, X_2, \cdots, X_n)|X_i]\}$$
$$= \int E[I_Q(X_1, X_2, \cdots, t_i, \cdots X_n)] F_X(dt)$$
$$E[I_Q(X_1, X_2, \cdots, t_i, \cdots, X_n)] = \prod_{k \neq i} P(X_k > t) = [1 - F_X(t)]^{n-1}$$

If F_X is continuous, strictly increasing, zero for $t < 0$, put $u = F_X(t)$, $du = f_X(t)\ dt$, $t = 0 \sim u = 0, t = \infty \sim u = 1$. Then

$$P(W = X_i) = \int_0^1 (1 - u)^{n-1}\ du = \int_0^1 u^{n-1}\ du = 1/n$$

This page titled 14.2: Problems on Conditional Expectation, Regression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

15: RANDOM SELECTION

15.1: Random Selection
15.2: Some Random Selection Problems
15.3: Problems on Random Selection

This page titled 15: Random Selection is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

15.1: Random Selection

Introduction

The usual treatments deal with a single random variable or a fixed, finite number of random variables, considered jointly. However, there are many common applications in which we select at random a member of a class of random variables and observe its value, or select a random number of random variables and obtain some function of those selected. This is formulated with the aid of a *counting*or *selecting* random variable N, which is nonegative, integer valued. It may be independent of the class selected, or may be related in some sequential way to members of the class. We consider only the independent case. Many important problems require *optional*random variables, sometimes called *Markov time*s. These involve more theory than we develop in this treatment.

Some common examples:

Total demand of N customers— N independent of the individual demands.
Total service time for N units— N independent of the individual service times.
Net gain in N plays of a game— N independent of the individual gains.
Extreme values of N random variables— N independent of the individual values.
Random sample of size N— N is usually determined by propereties of the sample observed.
Decide when to play on the basis of past results— N dependent on past

A useful model—random sums

As a basic model, we consider the sum of a random number of members of an iid class. In order to have a concrete interpretation to help visualize the formal patterns, we think of the demand of a random number of customers. We suppose the number of customers *N*is independent of the individual demands. We formulate a model to be used for a variety of applications.

A *basic sequence* $\{X_n : 0 \le n\}$ [Demand of n customers]
An incremental sequence $\{Y_n : 0 \le n\}$ [Individual demands]
These are related as follows:

$$X_n = \sum_{k=0}^{n} Y_k \text{ for } n \ge 0 \text{ and } X_n = 0 \text{ for } n < 0 \; Y_n = X_n - X_{n-1} \text{ for all } n$$

A *counting random variable* N. If $N = n$ then n of the Y_k are added to give the *compound demand* D (the random sum)

$$D = \sum_{k=0}^{N} Y_k = \sum_{k=0}^{\infty} I_{[N=k]} X_k = \sum_{k=0}^{\infty} I_{\{k\}}(N) X_k$$

Note. In some applications the counting random variable may take on the idealized value ∞. For example, in a game that is played until some specified result occurs, this may never happen, so that no finite value can be assigned to N. In such a case, it is necessary to decide what value X_∞ is to be assigned. For N independent of the Y_n (hence of the X_n), we rarely need to consider this possibility.

Independent selection from an iid incremental sequence

We assume throughout, unless specifically stated otherwise, that:
$X_0 = Y_0 = 0$
$\{Y_k : 1 \le k\}$ is iid
$\{N, Y_k : 0 \le k\}$ is an independent class

We utilize repeatedly two important propositions:
$E[h(D)|N = n] = E[h(X_n)]$, $n \ge 0$
$M_D(s) = g_N[M_Y(s)]$. If the Y_n are nonnegative integer valued, then so is D and $g_D(s) = g_N[g_Y(s)]$

DERIVATION

We utilize properties of generating functions, moment generating functions, and conditional expectation.

$E[I_{\{n\}}(N)h(D)] = E[h(D)|N = n]P(N = n)$ by definition of conditional expectation, given an event, Now,
$I_{\{n\}}(N)h(D) = I_{\{n\}}(N)h(X_n)$ and $E[I_{\{n\}}(N)h(X_n)] = P(N = n)E[h(X_n)]$. Hence
$E[h(D)|N = n]P(N = n) = P(N = n)E[h(X_n)]$. Division by $P(N = n)$ gives the desired result.
By the law of total probability (CE1b), $M_D(s) = E[e^{sD}] = E\{E[e^{sD}|N]\}$. By proposition 1 and the product rule for moment generating functions,

$$E[e^{sD}|N = n] = E[e^{sX_n}] = \prod_{k=1}^{n} E[e^{sY_k}] = M_Y^n(s)$$

Hence

$$M_D(s) = \sum_{n=0}^{\infty} M_Y^n(s)P(N = n) = g_N[M_Y(s)]$$

A parallel argument holds for g_D

— □

Remark. The result on M_D and g_D may be developed without use of conditional expectation.

in the integer-valued case.

$$M_D(s) = E[e^{sD}] = \sum_{k=0}^{\infty} E[I_{\{N=n\}}e^{sX_n}] = \sum_{k=0}^{\infty} P(N = n)E[e^{sX_n}]$$

$$= \sum_{k=0}^{\infty} P(N = n)M_Y^n(s) = g_N[M_Y(s)]$$

— □

Example 15.1.1 A service shop

Suppose the number N of jobs brought to a service shop in a day is Poisson (8). One fourth of these are items under warranty for which no charge is made. Others fall in one of two categories. One half of the arriving jobs are charged for one hour of shop time; the remaining one fourth are charged for two hours of shop time. Thus, the individual shop hour charges Y_k have the common distribution

$$Y = [0\ 1\ 2] \text{ with probabilities } PY = [1/4\ 1/2\ 1/4]$$

Make the basic assumptions of our model. Determine $P(D \leq 4)$.

Solution

$$g_N(s) = e^{8(s-1)}g_Y(s) = \frac{1}{4}(1 + 2s + s^2)$$

According to the formula developed above,

$$g_D(s) = g_N[g_Y(s)] = \exp((8/4)(1 + 2s + s^2) - 8) = e^{4s}e^{2s^2}e^{-6}$$

Expand the exponentials in power series about the origin, multiply out to get enough terms. The result of straightforward but somewhat tedious calculations is

$$g_D(s) = e^{-6}(1 + 4s + 10s^2 + \frac{56}{3}s^3 + \frac{86}{3}s^4 + \cdots)$$

Taking the coefficients of the generating function, we get

$$P(D \leq 4) \approx e^{-6}(1 + 4 + 10 + \frac{56}{3} + \frac{86}{3}) = e^{-6}\frac{187}{3} \approx 0.1545$$

Example 15.1.2 A result on Bernoulli trials

Suppose the counting random variable N ~ binomial (n, p) and $Y_i = I_{E_i}$, with $P(E_i) = p_0$. Then

$$g_N = (q + ps)^n \text{ and } g_Y(s) = q_0 + p_0 s$$

By the basic result on random selection, we have

$$g_D(s) = g_N[g_Y(s)] = [q + p(q_0 + p_0 s)]^n = [(1 - pp_0) + pp_0 s]^n$$

so that D ~ binomial (n, pp_0).

In the next section we establish useful m-procedures for determining the generating function g_D and the moment generating function M_D for the compound demand for simple random variables, hence for determining the complete distribution. Obviously, these will not work for all problems. It may helpful, if not entirely sufficient, in such cases to be able to determine the mean value $E[D]$ and variance $\text{Var}[D]$. To this end, we establish the following expressions for the mean and variance.

Example 15.1.3 Mean and variance of the compound demand

$E[D] = E[N]E[Y]$ and $\text{Var}[D] = E[N]\text{Var}[Y] + \text{Var}[N]E^2[Y]$

DERIVATION

$$E[D] = E[\sum_{n=0}^{\infty} I_{\{N=n\}} X_n] = \sum_{n=0}^{\infty} P(N = n)E[X_n]$$

$$= E[Y] \sum_{n=0}^{\infty} nP(N = n) = E[Y]E[N]$$

$$E[D^2] = \sum_{n=0}^{\infty} P(N = n)E[X_n^2] = \sum_{n=0}^{\infty} P(N = n)\{\text{Var}[X_n] + E^2[X_n]\}$$

$$= \sum_{n=0}^{\infty} P(N = n)\{n\text{Var}[Y] = n^2 E^2[Y]\} = E[N]\text{Var}[Y] + E[N^2]E^2[Y]$$

Hence

$$\text{Var}[D] = E[N]\text{Var}[Y] + E[N^2]E^2[Y] - E[N]^2 E^2[Y] = E[N]\text{Var}[Y] + \text{Var}[N]E^2[Y]$$

Example 15.1.4 Mean and variance for Example 15.1.1

$E[N] = \text{Var}[N] = 9$. By symmetry $E[Y] = 1$. $\text{Var}[Y] = 0.25(0 + 2 + 4) - 1 = 0.5$. Hence,

$$E[D] = 8 \cdot 1 = 8, \text{Var}[D] = 8 \cdot 0.5 + 8 \cdot 1 = 12$$

Calculations for the compound demand

We have m-procedures for performing the calculations necessary to determine the distribution for a composite demand D when the counting random variable N and the individual demands Y_k are simple random variables with not too many values. In some cases, such as for a Poisson counting random variable, we are able to approximate by a simple random variable.

The procedure gend

If the Y_i are nonnegative, integer valued, then so is D, and there is a generating function. We examine a strategy for computation which is implemented in the m-procedure *gend*. Suppose

$$g_N(s) = p_0 + p_1 s + p_2 s^2 + \cdot \cdot \cdot p_n s^n$$

$$g_Y(s) = \pi_0 + \pi_1 s + \pi_2 s^2 + \cdots \pi_m s^m$$

The coefficients of g_N and g_Y are the probabilities of the values of N and Y, respectively. We enter these and calculate the coefficients for powers of g_Y:

$$\begin{array}{lll} gN = [p_0\ p_1\ \cdots\ p_n] & 1 \times (n+1) & \text{Coefficients of } g_N \\ y = [\pi_0\ \pi_1\ \cdots\ \pi_n] & 1 \times (m+1) & \text{Coefficients of } g_Y \\ \quad \cdots & & \\ y2 = \text{conv}(y, y) & 1 \times (2m+1) & \text{Coefficients of } g_Y^2 \\ y3 = \text{conv}(y, y2) & 1 \times (3m+1) & \text{Coefficients of } g_Y^3 \\ \quad \cdots & & \\ yn = \text{conv}(y, y(n-1)) & 1 \times (nm+1) & \text{Coefficients of } g_Y^n \end{array}$$

We wish to generate a matrix P whose rows contain the joint probabilities. The probabilities in the ith row consist of the coefficients for the appropriate power of g_Y multiplied by the probability N has that value. To achieve this, we need a matrix, each of whose $n+1$ rows has $nm+1$ elements, the length of yn. We begin by "preallocating" zeros to the rows. That is, we set $P = \text{zeros}(n+1, n * m+1)$. We then replace the appropriate elements of the successive rows. The replacement probabilities for the ith row are obtained by the convolution of g_Y and the power of g_Y for the previous row. When the matrix P is completed, we remove zero rows and columns, corresponding to missing values of N and D (i.e., values with zero probability). To orient the joint probabilities as on the plane, we rotate P ninety degrees counterclockwise. With the joint distribution, we may then calculate any desired quantities.

Example 15.1.5 A compound demand

The number of customers in a major appliance store is equally likely to be 1, 2, or 3. Each customer buys 0, 1, or 2 items with respective probabilities 0.5, 0.4, 0.1. Customers buy independently, regardless of the number of customers. First we determine the matrices representing g_N and g_Y. The coefficients are the probabilities that each integer value is observed. Note that the *zero coefficients for any missing powers* must be included.

```
gN = (1/3)*[0 1 1 1];    % Note zero coefficient for missing zero power
gY = 0.1*[5 4 1];        % All powers 0 thru 2 have positive coefficients
gend
 Do not forget zero coefficients for missing powers
Enter the gen fn COEFFICIENTS for gN gN    % Coefficient matrix named gN
Enter the gen fn COEFFICIENTS for gY gY    % Coefficient matrix named gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view distribution for D, call for gD
disp(gD)                  % Optional display of complete distribution
         0    0.2917
    1.0000    0.3667
    2.0000    0.2250
    3.0000    0.0880
    4.0000    0.0243
    5.0000    0.0040
    6.0000    0.0003
EN = N*PN'
EN =   2
EY = Y*PY'
EY =  0.6000
ED = D*PD'
ED =  1.2000                % Agrees with theoretical EN*EY
P3 = (D>=3)*PD'
P3  = 0.1167
[N,D,t,u,PN,PD,PL] = jcalcf(N,D,P);
EDn = sum(u.*P)./sum(P);
```

```
disp([N;EDn]')
    1.0000    0.6000        % Agrees with theoretical E[D|N=n] = n*EY
    2.0000    1.2000
    3.0000    1.8000
VD = (D.^2)*PD' - ED^2
VD =  1.1200                % Agrees with theoretical EN*VY + VN*EY^2
```

Example 15.1.6 A numerical example

$$g_N(s) = \frac{1}{5}(1 + s + s^2 + s^3 + s^4)\ g_Y(s) = 0.1(5s + 3s^2 + 2s^3$$

Note that the zero power is missing from gY. corresponding to the fact that $P(Y = 0) = 0$.

```
gN = 0.2*[1 1 1 1 1];
gY = 0.1*[0 5 3 2];      % Note the zero coefficient in the zero position
gend
Do not forget zero coefficients for missing powers
Enter the gen fn COEFFICIENTS for gN  gN
Enter the gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view distribution for D, call for gD
disp(gD)                % Optional display of complete distribution
         0    0.2000
    1.0000    0.1000
    2.0000    0.1100
    3.0000    0.1250
    4.0000    0.1155
    5.0000    0.1110
    6.0000    0.0964
    7.0000    0.0696
    8.0000    0.0424
    9.0000    0.0203
   10.0000    0.0075
   11.0000    0.0019
   12.0000    0.0003
```

```
p3 = (D == 3)*PD'        % P(D=3)
P3 =  0.1250
P4_12 = ((D >= 4)&(D <= 12))*PD'
P4_12 = 0.4650           % P(4 <= D <= 12)
```

Example 15.1.7 Number of successes for random number N of trials.

We are interested in the number of successes in N trials for a general counting random variable. This is a generalization of the Bernoulli case in Example 15.1.2. Suppose, as in Example 15.1.2, the number of customers in a major appliance store is equally likely to be 1, 2, or 3, and each buys at least one item with probability $p = 0.6$. Determine the distribution for the number D of buying customers.

Solution

We use gN, gY, and gend.

```
gN = (1/3)*[0 1 1 1]; % Note zero coefficient for missing zero power
gY = [0.4 0.6];       % Generating function for the indicator function
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view distribution for D, call for gD
disp(gD)
         0    0.2080
    1.0000    0.4560
    2.0000    0.2640
    3.0000    0.0720
```

The procedure gend is limited to simple N and Y_k, with nonnegative integer values. Sometimes, a random variable with unbounded range may be approximated by a simple random variable. The solution in the following example utilizes such an approximation procedure for the counting random variable N.

Example 15.1.8 Solution of the shop time Example 15.1.1

The number N of jobs brought to a service shop in a day is Poisson (8). The individual shop hour charges Y_k have the common distribution $Y =$ [0 1 2] with probabilities $PY =$ [1/4 1/2 1/4].

Under the basic assumptions of our model, determine $P(D \leq 4)$.

Solution

Since Poisson N is unbounded, we need to check for a sufficient number of terms in a simple approximation. Then we proceed as in the simple case.

```
pa = cpoisson(8,10:5:30)     % Check for sufficient number of terms
pa =   0.2834    0.0173    0.0003    0.0000    0.0000
p25 = cpoisson(8,25)         % Check on choice of n = 25
p25 =  1.1722e-06
gN = ipoisson(8,0:25);       % Approximate gN
gY = 0.25*[1 2 1];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view distribution for D, call for gD
disp(gD(D<=20,:))            % Calculated values to D = 50
         0    0.0025         % Display for D <= 20
    1.0000    0.0099
    2.0000    0.0248
    3.0000    0.0463
    4.0000    0.0711
    5.0000    0.0939
    6.0000    0.1099
    7.0000    0.1165
    8.0000    0.1132
    9.0000    0.1021
```

```
   10.0000    0.0861
   11.0000    0.0684
   12.0000    0.0515
   13.0000    0.0369
   14.0000    0.0253
   15.0000    0.0166
   16.0000    0.0105
   17.0000    0.0064
   18.0000    0.0037
   19.0000    0.0021
   20.0000    0.0012
sum(PD)                      % Check on sufficiency of approximation
ans =  1.0000
P4 = (D<=4)*PD'
P4 =   0.1545                % Theoretical value (4  places) = 0.1545
ED = D*PD'
ED =   8.0000                % Theoretical = 8  (Example 15.1.4)
VD = (D.^2)*PD' - ED^2
VD =  11.9999                % Theoretical = 12 (Example 15.1.4)
```

The m-procedures mgd and jmgd

The next example shows a fundamental limitation of the gend procedure. The values for the individual demands are not limited to integers, and there are considerable gaps between the values. In this case, we need to implement the moment generating function M_D rather than the generating function g_D.

In the generating function case, it is as easy to develop the joint distribution for $\{N, D\}$ as to develop the marginal distribution for D. For the moment generating function, the joint distribution requires considerably more computation. As a consequence, we find it convenient to have two m-procedures: *mgd* for the marginal distribution and *jmgd* for the joint distribution.

Instead of the convolution procedure used in gend to determine the distribution for the sums of the individual demands, the m-procedure mgd utilizes the m-function mgsum to obtain these distributions. The distributions for the various sums are concatenated into two row vectors, to which csort is applied to obtain the distribution for the compound demand. The procedure requires as input the generating function for N and the actual distribution, Y and PY, for the individual demands. For gN, it is necessary to treat the coefficients as in gend. However, the actual values and probabilities in the distribution for *Y* are put into a pair of row matrices. *If Y is integer valued, there are no zeros in the probability matrix for missing values.*

Example 15.1.9 Noninteger values

A service shop has three standard charges for a certain class of warranty services it performs: \$10, \$12.50, and \$15. The number of jobs received in a normal work day can be considered a random variable N which takes on values 0, 1, 2, 3, 4 with equal probabilities 0.2. The job types for arrivals may be represented by an iid class $\{Y_i : 1 \le i \le 4\}$, independent of the arrival process. The Y_i take on values 10, 12.5, 15 with respective probabilities 0.5, 0.3, 0.2. Let C be the total amount of services rendered in a day. Determine the distribution for C.

Solution

```
gN = 0.2*[1 1 1 1 1];        % Enter data
Y = [10 12.5 15];
PY = 0.1*[5 3 2];
mgd                          % Call for procedure
Enter gen fn COEFFICIENTS for gN  gN
Enter VALUES for Y  Y
Enter PROBABILITIES for Y  PY
Values are in row matrix D; probabilities are in PD.
```

```
To view the distribution, call for mD.
disp(mD)                        % Optional display of distribution
         0    0.2000
   10.0000    0.1000
   12.5000    0.0600
   15.0000    0.0400
   20.0000    0.0500
   22.5000    0.0600
   25.0000    0.0580
   27.5000    0.0240
   30.0000    0.0330
   32.5000    0.0450
   35.0000    0.0570
   37.5000    0.0414
   40.0000    0.0353
   42.5000    0.0372
   45.0000    0.0486
   47.5000    0.0468
   50.0000    0.0352
   52.5000    0.0187
   55.0000    0.0075
   57.5000    0.0019
   60.0000    0.0003
```

We next recalculate Example 15.1.6, above, using mgd rather than gend.

Example 15.1.10 Recalculation of Example 15.1.6

In Example 15.1.6, we have

$$g_N(s) = \frac{1}{5}(1 + s + s^2 + s^3 + s^4)\ g_Y(s) = 0.1(5s + 3s^2 + 2s^3)$$

The means that the distribution for Y is $Y =$ [1 2 3] and $PY =$ 0.1 * [5 3 2].

We use the same expression for gN as in Example 15.1.6.

```
gN = 0.2*ones(1,5);
Y = 1:3;
PY = 0.1*[5 3 2];
mgd
Enter gen fn COEFFICIENTS for gN  gN
Enter VALUES for Y  Y
Enter PROBABILITIES for Y  PY
Values are in row matrix D; probabilities are in PD.
To view the distribution, call for mD.
disp(mD)
         0    0.2000
    1.0000    0.1000
    2.0000    0.1100
    3.0000    0.1250
    4.0000    0.1155
    5.0000    0.1110
    6.0000    0.0964
    7.0000    0.0696
    8.0000    0.0424
```

```
    9.0000    0.0203
   10.0000    0.0075
   11.0000    0.0019
   12.0000    0.0003
P3 = (D==3)*PD'
P3 =   0.1250
ED = D*PD'
ED =   3.4000
P_4_12 = ((D>=4)&(D<=12))*PD'
P_4_12 =  0.4650
P7 = (D>=7)*PD'
P7 =   0.1421
```

As expected, the results are the same as those obtained with gend.

If it is desired to obtain the joint distribution for $\{N, D\}$, we use a modification of mgd called *jmgd*. The complications come in placing the probabilities in the P matrix in the desired positions. This requires some calculations to determine the appropriate size of the matrices used as well as a procedure to put each probability in the position corresponding to its D value. Actual operation is quite similar to the operation of mgd, and requires the same data format.

A principle use of the joint distribution is to demonstrate features of the model, such as $E[D|N = n] = nE[Y]$, etc. This, of course, is utilized in obtaining the expressions for $M_D(s)$ in terms of $g_N(s)$ and $M_Y(s)$. This result guides the development of the computational procedures, but these do not depend upon this result. However, it is usually helpful to demonstrate the validity of the assumptions in typical examples.

Remark. In general, if the use of gend is appropriate, it is faster and more efficient than mgd (or jmgd). And it will handle somewhat larger problems. But both m-procedures work quite well for problems of moderate size, and are convenient tools for solving various “compound demand” type problems.

This page titled 15.1: Random Selection is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

15.2: Some Random Selection Problems

In the unit on Random Selection, we develop some general theoretical results and computational procedures using MATLAB. In this unit, we extend the treatment to a variety of problems. We establish some useful theoretical results and in some cases use MATLAB procedures, including those in the unit on random selection.

The Poisson Decomposition

In many problems, the individual demands may be categorized in one of *m* types. If the random variable T_i is the type of the ith arrival and the class $\{T_i : 1 \leq i\}$ is iid, we have *multinomial trials*. For $m = 2$ we have the Bernoulli or binomial case, in which one type is called a success and the other a failure.

Multinomial trials

We analyze such a sequence of trials as follows. Suppose there are *m* types, which we number 1 through m. Let E_{ki} be the event that type k occurs on the ith component trial. For each i, the class $\{E_{ki} : 1 \leq k \leq m\}$ is a partition, since on each component trial exactly one of the types will occur. The type on the ith trial may be represented by the type random variable

$$T_i = \sum_{k=1}^{m} kI_{E_{ki}}$$

we assume

$$\{T_k : 1 \leq i\} \text{ is iid, with } P(T_i = k) = P(E_{ki}) = p_k \text{ invariant with } i$$

In a sequence of n trials, we let N_{kn} be the number of occurrences of type k. Then

$$N_{kn} = \sum_{i=1}^{n} I_{E_{ki}} \text{ with } \sum_{k=1}^{m} N_{kn} = n$$

Now each N_{kn} ~ binomial (n, p_k). The class $\{N_{kn} : 1 \leq k \leq m\}$ cannot be independent, since it sums to n. If the values of $m - 1$ of them are known, the value of the other is determined. If $n_1 + n_2 + \cdots n_m = n$. the event

$$\{N_{1n} = n_1, N_{2n} = n_2, \cdots, N_{mn} = n_m\}$$

is one of the

$$C(n; n_1, n_2, \cdots, n_m) = n!/(n1!n2! \cdots n_m!)$$

ways of arranging n_1 of the E_{1i}, n_2 of the E_{2i}, $\cdots$, n_m of the E_{mi}. Each such arrangement has probability $p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$, so that

$$P(N_{1n} = n_1, N_{2n} = n_2, \cdots N_{mn} = n_m) = n! \prod_{k=1}^{m} \frac{p_k^{n_k}}{n_k!}$$

This set of joint probabilities constitutes the *multinomial distribution*. For $m = 2$, and type 1 a success, this is the binomial distribution with parameter (n, p_1).

A random number of multinomial trials

We consider, in particular, the case of a random number N of multinomial trials, where N ~ Poisson (μ). Let N_k be the number of results of type k in a random number N of multinomial trials.

$$N_k = \sum_{i=1}^{N} I_{E_{ki}} = \sum_{n=1}^{\infty} I_{\{N=n\}} N_{kn} \text{ with } \sum_{k=1}^{m} N_k = N$$

Poisson decomposition

Suppose

N ~ Poisson (μ)

$\{T_i : 1 \le i\}$ is iid with $P(T_i = k) = p_k$, $1 \le k \le m$
$\{N, T_i : 1 \le i\}$ is independent

Then

Each N_k ~ Poisson (μp_k)
$\{N_k : 1 \le k \le m\}$ is independent.

— □

The usefulness of this remarkable result is enhanced by the fact that the sum of independent Poisson random variables is also Poisson, with μ for the sum the sum of the μ_i for the variables added. This is readily established with the aid of the generating function. Before verifying the propositions above, we consider some examples.

Example 15.2.1 A shipping problem

The number N of orders per day received by a mail order house is Poisson (300). Orders are shipped by next day express, by second day priority, or by regular parcel mail. Suppose 4/10 of the customers want next day express, 5/10 want second day priority, and 1/10 require regular mail. Make the usual assumptions on compound demand. What is the probability that fewer than 150 want next day express? What is the probability that fewer than 300 want one or the other of the two faster deliveries?

Solution

Model as a random number of multinomial trials, with three outcome types: Type 1 is next day express, Type 2 is second day priority, and Type 3 is regular mail, with respective probabilities $p_1 = 0.4$, $p_2 = 0.5$, and $p_3 = 0.1$. The N_1 ~ Poisson $(0.4 \cdot 300 = 120)$, N_2 ~ Poisson $(0.5 \cdot 300 = 150)$, and N_3 ~ Poisson $(0.1 \cdot 300 = 30)$. Also $N_1 + N_2$ ~ Poisson (120 + 150 = 270).

```
P1 = 1 - cpoisson(120,150)
P1  =  0.9954
P12 = 1 - cpoisson(270,300)
P12 =  0.9620
```

Example 15.2.2 Message routing

A junction point in a network has two incoming lines and two outgoing lines. The number of incoming messages N_1 on line one in one hour is Poisson (50); on line 2 the number is N_2 ~ Poisson (45). On incoming line 1 the messages have probability $P_{1a} = 0.33$ of leaving on outgoing line a and $1 - p_{1a}$ of leaving on line b. The messages coming in on line 2 have probability $P_{2a} = 0.47$ of leaving on line a. Under the usual independence assumptions, what is the distribution of outgoing messages on line a? What are the probabilities of at least 30, 35, 40 outgoing messages on line a?

Solution

By the Poisson decomposition, N_a ~ Poisson $(50 \cdot 0.33 + 45 \cdot 0.47 = 37.65)$.

```
ma = 50*0.33 + 45*0.47
ma =  37.6500
Pa = cpoisson(ma,30:5:40)
Pa =   0.9119    0.6890    0.3722
```

VERIFICATION of the Poisson decomposition

$N_k = \sum_{i=1}^{N} I_{Eki}$.
This is composite demand with $Y_k = I_{E_{ki}}$, so that $g_{Y_k}(s) = q_k + sp_k = 1 + p_k(s - 1)$. Therefore,

$$g_{N_k}(s) = g_N[g_{Y_k}(s)] = e = e$$

which is the generating function for N_k ~ Poisson (μp_k).
For any $n_1, n_2, \cdots, n_m$, let $n = n_1 + n_2 + \cdots + n_m$, and consider

$$A = \{N_1 = n_1, N_2 = n_2, \cdots, N_m = n_m\} = \{N = n\} \cap \{N_{1n} = N_1, N_{2n} = n_2, \cdots, N_{mn} = n_m\}$$

Since N is independent of the class of $I_{E_{ki}}$, the class

$$\{\{N = n\}, \{N_{1n} = n_1, N_{2n} = n_2, \cdots, N_{mn} = n_m\}\}$$

is independent. By the product rule and the multinomial distribution

$$P(A) = e^{-\mu}\frac{\mu^n}{n!} \cdot n! \prod_{k=1}^m \frac{p_k^{n_k}}{(n_k)!} = \prod_{k=1}^m e^{-\mu p_k}\frac{p_k^{n_k}}{n_k!} = \prod_{k=1}^m P(N_k = n_k)$$

The second product uses the fact that

$$e^{\mu} = e^{\mu(p_1+p_2+\cdots+p_m)} = \prod_{k=1}^m e^{\mu p_k}$$

Thus, the product rule holds for the class

Extreme values

Consider an iid class $\{Y_i : 1 \le i\}$ of nonnegative random variables. For any positive integer n we let

$$V_n = \min\{Y_1, Y_2, \cdots, Y_n\} \text{ and } W_n = \max\{Y_1, Y_2, \cdots, Y_n\}$$

Then

$$P(V_n > t) = P^n(Y > t) \text{ and } P(W_n \le t) = P^n(Y \le t)$$

Now consider a random number N of the Y_i. The minimum and maximum random variables are

$$V_N = \sum_{n=0}^{\infty} I_{\{N=n\}} V_n \text{ and } W_N = \sum_{n=0}^{\infty} I_{\{N=n\}} W_n$$

— □

Computational formulas

If we set $V_0 = W_0 = 0$, then

$F_V(t) = P(V \le t) = 1 + P(N = 0) - g_N[P(Y > t)]$
$F_W(t) = g_N[P(Y \le t)]$

These results are easily established as follows. $\{V_N > t\} = \bigvee_{n=0}^{\infty}\{N = n\}\ \{V_n > t\}$. By additivity and independence of $\{N, V_n\}$ for each n

$$P(V_N > t) = \sum_{n=0}^{\infty} P(N = n)P(V_n > t) = \sum_{n=1}^{\infty} P(N = n)P^n(Y > t), \text{ since } P(V_0 > t) = 0$$

If we add into the last sum the term $P(N = 0)P^0(Y > t) = P(N = 0)$ then subtract it, we have

$$P(V_N > t) = \sum_{n=0}^{\infty} P(N = n)P^n(Y > t) - P(N = 0) = g_N[P(Y > t)] - P(N = 0)$$

A similar argument holds for proposition (b). In this case, we do not have the extra term for $\{N = 0\}$, since $P(W_0 \le t) = 1$.

Special case. In some cases, $N = 0$ does not correspond to an admissible outcome (see Example 14.2.4, below, on lowest bidder and Example 14.2.6). In that case

$$F_V(t) = \sum_{n=1}^{\infty} P(V_n \le t)P(N = n) = \sum_{n=1}^{\infty}[1 - P^n(Y > t)]P(N = n) = \sum_{n=1}^{\infty} P(N = n) - \sum_{n=1}^{\infty} P^n(Y > t)P(N = n)$$

Add $P(N=0)=p^0\ (Y>t)P(N=0)$ to each of the sums to get

$$F_V(t)=1-\sum_{n=0}^{\infty}P^n(Y>t)P(N=n)=1-g_N[P(Y>t)]$$

— □

Example 15.2.3 Maximum service time

The number N of jobs coming into a service center in a week is a random quantity having a Poisson (20) distribution. Suppose the service times (in hours) for individual units are iid, with common distribution exponential (1/3). What is the probability the maximum service time for the units is no greater than 6, 9, 12, 15, 18 hours?

Solution

$$P(W_N\le t)=g_N[P(Y\le t)]=e^{20[F_Y(t)-1]}=\exp(-20e^{-t/3})$$

```
t = 6:3:18;
PW = exp(-20*exp(-t/3));
disp([t;PW]')
    6.0000    0.0668
    9.0000    0.3694
   12.0000    0.6933
   15.0000    0.8739
   18.0000    0.9516
```

Example 15.2.4 Lowest Bidder

A manufacturer seeks bids on a modification of one of his processing units. Twenty contractors are invited to bid. They bid with probability 0.3, so that the number of bids N ~ binomial (20,0.3). Assume the bids Y_i (in thousands of dollars) form an iid class. The market is such that the bids have a common distribution symmetric triangular on (150,250). What is the probability of at least one bid no greater than 170, 180, 190, 200, 210? Note that no bid is *not* a low bid of zero, hence we must use the special case.

Solution

$P(V\le t)=1-g_N[P(Y>t)]=1-(0.7+0.3p)^{20}$ where $p=P(Y>t)$

Solving graphically for $p=P(V>t)$, we get

$p=$ [23/25 41/50 17/25 1/2 8/25] for $t=$ [170 180 190 200 210]

Now $g_N(s)=(0.7+0.3s)^{20}$. We use MATLAB to obtain

```
t = [170 180 190 200 210];
p = [23/25 41/50 17/25 1/2 8/25];
PV = 1 - (0.7 + 0.3*p).^20;
disp([t;p;PV]')
170.0000    0.9200    0.3848
180.0000    0.8200    0.6705
190.0000    0.6800    0.8671
200.0000    0.5000    0.9612
210.0000    0.3200    0.9896
```

Example 15.2.5 Example 15.2.4 with a general counting variable

Suppose the number of bids is 1, 2 or 3 with probabilities 0.3, 0.5, 0.2, respectively.

Determine $P(V \leq t)$ in each case.

Solution

The minimum of the selected Y's is no greater than t if and only if there is at least one Y less than or equal to t. We determine in each case probabilities for the number of bids satisfying $Y \leq t$. For each t, we are interested in the probability of one or more occurrences of the event $Y \leq t$. This is essentially the problem in Example 7 from "Random Selection", with probability $p = P(Y \leq t)$.

```
t = [170 180 190 200 210];
p = [23/25 41/50 17/25 1/2 8/25]; % Probabilities Y <= t are 1 - p
gN = [0 0.3 0.5 0.2];             % Zero for missing value
PV = zeros(1,length(t));
for i=1:length(t)
gY = [p(i),1 - p(i)];
[d,pd] = gendf(gN,gY);
PV(i) = (d>0)*pd';                 % Selects positions for d > 0 and
end                                % adds corresponding probabilities
disp([t;PV]')
  170.0000    0.1451
  180.0000    0.3075
  190.0000    0.5019
  200.0000    0.7000
  210.0000    0.8462
```

Example 15.2.4 may be worked in this manner by using `gN = ibinom(20,0.3,0:20)`. The results, of course, are the same as in the previous solution. The fact that the probabilities in this example are lower for each *t* than in Example 15.2.4 reflects the fact that there are probably fewer bids in each case.

Example 15.2.6 Batch testing

Electrical units from a production line are first inspected for operability. However, experience indicates that a fraction p of those passing the initial operability test are defective. All operable units are subsequenly tested in a batch under continuous operation (a “burn in” test). Statistical data indicate the defective units have times to failure Y_i iid, exponential (λ, whereas good units have very long life (infinite from the point of view of the test). A batch of n units is tested. Let V be the time of the first failure and N be the number of defective units in the batch. If the test goes t units of time with no failure (i.e., $V > t$), what is the probability of no defective units?

Solution

Since no defective units implies no failures in any reasonable test time, we have

$$\{N = 0\} \subset \{V > t\} \text{ so that } P(N = 0|V > t) = \frac{P(N = 0)}{P(V > t)}$$

Since $N = 0$ does not yield a minimum value, we have $P(V > t) = g_N[P(Y > t)]$. Now under the condition above, the number of defective units N ~ binomial (n, p), so that $g_N(s) = (q + ps)^n$. If N is large and p is reasonably small, N is approximately Poisson (np) with $g_N(s) = e^{np(s-1)}$ and $P(N = 0) = e^{-np}$. Now $P(Y > t) = e^{-\lambda t}$; for large n

$$P(N = 0|V > t) = \frac{e^{-np}}{e^{np[P(Y>t)-1]}} = e^{-npP(Y>t)} = e^{-npe^{-lambdat}}$$

For $n = 5000$, $p = 0.001$, $\lambda = 2$, and $t = 1, 2, 3, 4, 5$, MATLAB calculations give

```
t = 1:5;
n = 5000;
p = 0.001;
lambda = 2;
P = exp(-n*p*exp(-lambda*t));
disp([t;P]')
    1.0000    0.5083
    2.0000    0.9125
    3.0000    0.9877
    4.0000    0.9983
    5.0000    0.9998
```

It appears that a test of three to five hours should give reliable results. In actually designing the test, one should probably make calculations with a number of different assumptions on the fraction of defective units and the life duration of defective units. These calculations are relatively easy to make with MATLAB.

Bernoulli trials with random execution times or costs

Consider a Bernoulli sequence with probability p of success on any component trial. Let N be the number of the trial on which the first success occurs. Let Y_i be the time (or cost) to execute the ith trial. Then the total time (or cost) from the beginning to the completion of the first success is

$T = \sum_{i=1}^{N} Y_i$ (composite "demand" with $N - 1$ ~ geometric p)

We suppose the Y_i form an iid class, independent of N. Now $N - 1$ ~ geometric (p) implies $g_N(s) = ps/(1 - qs)$, so that

$$M_T(s) = g_N[M_Y(s)] = \frac{pM_Y(s)}{1 - qM_Y(s)}$$

There are two useful special cases:

Y_i ~ exponential (λ), so that $M_Y(s) =$ -.

$$M_T(s) = - = -$$

which implies T ~ exponential $(p\lambda)$.

$Y_i - 1$ ~ geometric (p_0), so that $g_Y(s) = \dfrac{\lambda}{\lambda - s}$

$$g_T(s) = \frac{p\lambda/(\lambda - s)}{1 - q\lambda/(\lambda - s)} = \frac{p\lambda}{p\lambda - s}$$

so that $T - 1$ ~ geometric (pp_0).

Example 15.2.7 Job interviews

Suppose a prospective employer is interviewing candidates for a job from a pool in which twenty percent are qualified. Interview times (in hours) Y_i are presumed to form an iid class, each exponential (3). Thus, the average interview time is 1/3 hour (twenty minutes). We take the probability for success on any interview to be $p = 0.2$. What is the probability a satisfactory candidate will be found in four hours or less? What is the probability the maximum interview time will be no greater than 0.5, 0.75, 1, 1.25, 1.5 hours?

Solution

T ~ exponential $(0.2 \cdot 3 = 0.6)$, so that $P(T \leq 4) = 1 - e^{-0.6 \cdot 4} = 0.9093$.

$$P(W \leq t) = g_N[P(Y \leq t)] = \frac{0.2(1 - e^{-3t})}{1 - 0.8(1 - e^{-3t})} = \frac{1 - e^{-3t}}{1 + 4e^{-3t}}$$

MATLAB computations give

```
t = 0.5:0.25:1.5;
PWt = (1 - exp(-3*t))./(1 + 4*exp(-3*t));
disp([t;PWt]')
    0.5000    0.4105
    0.7500    0.6293
    1.0000    0.7924
    1.2500    0.8925
    1.5000    0.9468
```

The average interview time is 1/3 hour; with probability 0.63 the maximum is 3/4 hour or less; with probability 0.79 the maximum is one hour or less; etc.

In the general case, solving for the distribution of T requires transform theory, and may be handled best by a program such as Maple or Mathematica.

For the case of *simple* Y_i we may use approximation procedures based on properties of the geometric series. Since $N - 1$ ~ geometric (p).

$$g_N 9s) = \frac{ps}{1 - qs} = ps \sum_{k=0}^{\infty} (qs)^k = ps[\sum_{k=0}^{n} (qs)^k + \sum_{k=m+1}^{\infty} (qs)^k] = ps[\sum_{k=0}^{n} (qs)^k + (qs)^{n+1} \sum_{k=0}^{\infty} (qs)^k]$$

$$= ps[\sum_{k=0}^{n} (qs)^k] + (qs)^{n+1} g_N 9s) = g_n(s) + (qs)^{n+1} g_N(s)$$

Note that $g_n(s)$ has the form of the generating function for a simple approximation N_n which matches values and probabilities with N up to $k = n$. Now

$$g_T(s) = g_n[g_Y(s)] + (qs)^{n+1} g_N[g_Y(s)]$$

The evaluation involves convolution of coefficients which effectively sets $s = 1$. Since $g_N(1) = g_Y(1) = 1$.

$(qs)^{n+1} g_N[g_Y(s)]$ for $s = 1$ reduces to $q^{n+1} = P(N > n)$

which is negligible if n is large enough. Suitable n may be determined in each case. With such an n, if the Y_i are nonnegative, integer-valued, we may use the gend procedure on $g_n[g_Y(s)]$, where

$$g_n(s) = ps + pqs^2 + pq^2 s^3 + \cdots + pq^n s^{n+1}$$

For the integer-valued case, as in the general case of simple Y_i, we could use mgd. However, gend is usually faster and more efficient for the integer-valued case. Unless q is small, the number of terms needed to approximate g_n is likely to be too great.

Example 15.2.8 Approximating the generating function

Let $p = 0.3$ and Y be uniformly distributed on $\{1, 2, \cdots, 10\}$. Determine the distribution for

$$T = \sum_{k=1}^{N} Y_k$$

Solution

```
p = 0.3;
q = 1 - p;
a = [30 35 40];          % Check for suitable n
b = q.^a
b =  1.0e-04 *           % Use n = 40
     0.2254    0.0379    0.0064
n = 40;
k = 1:n;
gY = 0.1*[0 ones(1,10)];
gN = p*[0 q.^(k-1)];     % Probabilities, 0 <= k <= 40
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Values are in row matrix D; probabilities are in PD.
To view the distribution, call for gD.
sum(PD)                % Check sum of probabilities
ans =  1.0000
FD = cumsum(PD);       % Distribution function for D
plot(0:100,FD(1:101))  % See Figure 15.2.1
P50 = (D<=50)*PD'
P50 =  0.9497
P30 = (D<=30)*PD'
P30 =  0.8263
```

Figure one is a graph labeled, execution time distribution function. The horizontal axis is labeled, Time, and the vertical axis is labeled, probability. The values on the horizontal axis range from 0 to 100 in increments of 10. The values on the vertical axis range from 0 to 1 in increments of 0.1. There is one plotted distribution function on this graph. It begins in the bottom-left corner, at the point (0, 0), and moves right at a strong positive slope. As the plot moves from left to right, the slope decreases as the function increases. About midway across the graph horizontally, the plot is nearly at the top, at a probability value above 0.9. The plot continues to increase at a decreasing rate until it tapers off to a horizontal line by the point (80, 1), at which it continues and terminates at the top-right corner.

Figure 15.2.1. Execution Time Distribution Function F_D.

The same results may be achieved with mgd, although at the cost of more computing time. In that case, use gN as in Example 15.2.8, but use the actual distribution for Y.

Arrival times and counting processes

Suppose we have phenomena which take place at discrete instants of time, separated by random waiting or interarrival times. These may be arrivals of customers in a store, of noise pulses on a communications line, vehicles passing a position on a road, the failures of a system, etc. We refer to these occurrences as *arrivals* and designate the times of occurrence as *arrival times*. A stream of arrivals may be described in three equivalent ways.

- *Arrival times*: $\{S_n : 0 \le n\}$, with $0 = S_0 < S_1 < \cdots$ a.s. (basic sequence)
- *Interarrival times*: $\{W_i : 1 \le i\}$, with each $W_i > 0$ a.s. (incremental sequence)

The strict inequalities imply that with probability one there are no simultaneous arrivals. The relations between the two sequences are simply

$$S_0 = 0,\ S_n = \sum_{i=1}^{n} W_i \text{ and } W_n = S_n - S_{n-1} \text{ for all } n \ge 1$$

The formulation indicates the essential equivalence of the problem with that of the compound demand. The notation and terminology are changed to correspond to that customarily used in the treatment of arrival and counting processes.

The stream of arrivals may be described in a third way.

- *Counting processes*: $N_t = N(t)$ is the number of arrivals in time period $(0, t]$. It should be clear that this is a random quantity for each nonnegative t. For a given t, ω the value is $N(t, \omega)$. Such a family of random variables constitutes a *random process*. In this case the random process is a *counting process*.

We thus have three equivalent descriptions for the stream of arrivals.

$$\{S_n : 0 \le n\}\ \{W_n : 1 \le n\}\ \{N_t : 0 \le t\}$$

Several properties of the counting process N should be noted:
$N(t + h) - N(t)$ counts the arrivals in the interval $(t, t + h]$, $h > 0$, so that $N(t + h) \ge N(t)$ for $h > 0$.
$N_0 = 0$ and for $t > 0$ we have

$N_t = \sum_{i=1}^{\infty} I_{(0,t]}(S_i) = \max\ \{n : S_n \le t\} = \min\ \{n : S_{n+1} > t\}$

For any given ω, $N(\cdot, \omega)$ is a nondecreasing, right-continuous, integer-valued function defined on $[0, \infty)$, with $N(0, \omega) = 0$.

The essential relationships between the three ways of describing the stream of arrivals is displayed in

$$W_n = S_n - S_{n-1},\ \{N_t \ge n\} = \{S_n \le t\},\ \{N_t = n\} = \{S_n \le t < S_{n+1}\}$$

This imples

$$P(N_t = n) = P(S_n \le t) - P(S_{n+1} \le t) = P(S_{n+1} > t) - P(S_n > t)$$

Although there are many possibilities for the interarrival time distributions, we assume

$$\{W_i : 1 \le i\} \text{ is iid, with } W_i > 0 \text{ a.s.}$$

Under such assumptions, the counting process is often referred to as a *renewal process* and the interrarival times are called *renewal times*. In the literature on renewal processes, it is common for the random variable to count an arrival at $t = 0$. This requires an adjustment of the expressions relating N_t and the S_i. We use the convention above.

Exponential iid interarrival times

The case of exponential interarrival times is natural in many applications and leads to important mathematical results. We utilize the following propositions about the arrival times S_n, the interarrival times W_i, and the counting process N.

If $\{W_i : 1 \le i\}$ is iid exponential (λ), then S_n ~ gamma (n, λ) for all $n \ge 1$. This is worked out in the unit on TRANSFORM METHODS, in the discussion of the connection between the gamma distribution and the exponential distribution.
S_n ~ gamma (n, λ) for all $n \ge 1$, and $S_0 = 0$, iff N_t ~ Poisson (λt) for all $t > 0$. This follows the result in the unit DISTRIBUTION APPROXI9MATIONS on the relationship between the Poisson and gamma distributions, along with the fact that $\{N_t \ge n\} = \{S_n \le t\}$.

Remark. The counting process is a Poisson process in the sense that N_t ~ Poisson (λt) for all $t > 0$. More advanced treatments show that the process has independent, stationary increments. That is

$N(t + h) - N(t) = N(h)$ for all $t, h > 0$, and
For $t_1 < t_2 \le t_3 < t_4 \le \cdots \le t_{m-1} < t_m$, the class $\{N(t_2) - N(N_1), N(t_4) - N(t_3), \cdots, N(t_m) - N(t_{m-1})\}$ is independent.

In words, the number of arrivals in any time interval depends upon the length of the interval and not its location in time, and the numbers of arrivals in nonoverlapping time intervals are independent.

Example 15.2.9 Emergency calls

Emergency calls arrive at a police switchboard with interarrival times (in hours) exponential (15). Thus, the average interarrival time is 1/15 hour (four minutes). What is the probability the number of calls in an eight hour shift is no more than 100, 120, 140?

```
p = 1 - cpoisson(8*15,[101 121 141])
p  =  0.0347    0.5243    0.9669
```

We develop next a simple computational result for arrival processes for which S_n ~ gamma (n, λ)

Example 15.2.10 Gamma arrival times

Suppose the arrival times S_n ~ gamma (n, λ) and g is such that

$$\int_0^\infty |g| < \infty \text{ and } E[\sum_{n=1}^\infty |g(S_n)|] < \infty$$

Then

$$E[\sum_{n=1}^\infty g(S_n)] = \lambda \int_0^\infty g$$

VERIFICATION

We use the countable sums property (E8b) for expectation and the corresponding property for integrals to assert

$$E[\sum_{n=1}^\infty g(S_n)] = \sum_{n=1}^\infty E[g(S_n)] = \sum_{n=1}^\infty \int_0^\infty g(t) f_n(t)\, dt \text{ where } f_n(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$$

We may apply (E8b) to assert

$$\sum_{n=1}^\infty \int_0^\infty g f_n = \int_0^\infty g \sum_{n=1}^\infty f_n$$

Since

$$\sum_{n=1}^\infty f_n(t) = \lambda e^{-\lambda t} \sum_{n=1}^\infty \frac{(\lambda t)^{n-1}}{(n-1)!} = \lambda e^{-\lambda t} e^{\lambda t} = \lambda$$

the proposition is established.

Example 15.2.11 Discounted replacement costs

A critical unit in a production system has life duration exponential (λ). Upon failure the unit is replaced immediately by a similar unit. Units fail independently. Cost of replacement of a unit is c dollars. If money is discounted at a rate α, then a dollar spent t units of time in the future has a current value $e^{\alpha t}$. If S_n is the time of replacement of the nth unit, then S_n ~ gamma (n, λ) and the present value of all future replacements is

$$C = \sum_{n=1}^\infty c e^{-\alpha S_n}$$

The expected replacement cost is

$$E[C] = E[\sum_{n=1}^\infty g(S_n)] \text{ where } g(t) = ce^{-\infty}$$

Hence

$$E[C] = \lambda \int_0^\infty ce^{-\alpha t}\, dt = \frac{\lambda c}{\alpha}$$

Suppose unit replacement cost $c = 1200$, average time (in years) to failure $1/\lambda = 1/4$, and the discount rate per year $\alpha = 0.08$ (eight percent). Then

$$E[C] = \frac{1200 \cdot 4}{0.08} = 60,000$$

Example 15.2.12 Random costs

Suppose the cost of the nth replacement in Example 15.2.11 is a random quantity C_n, with $\{C_n, S_n\}$ independent and $E[C_n] = c$, invariant with n. Then

$$E[C] = E[\sum_{n=1}^\infty C_n e^{-\alpha S_n}] = \sum_{n=1}^\infty E[C_n]E[e^{-\alpha S_n}] = \sum_{n=1}^\infty cE[e^{-\alpha S_n}] = \frac{\lambda c}{\alpha}$$

The analysis to this point assumes the process will continue endlessly into the future. Often, it is desirable to plan for a specific, finite period. The result of Example 15.2.10 may be modified easily to account for a finite period, often referred to as a *finite horizon*.

Example 15.2.13 Finite horizon

Under the conditions assumed in Example 15.2.10, above, let N_t be the counting random variable for arrivals in the interval $(0, t]$.

$$\text{If } Z_t = \sum_{n=1}^{N_t} g(S_n), \text{ then } E[Z_t] = \lambda \int_0^t g(u)\, du$$

VERIFICATION

Since $N_t \geq n$ iff $S_n \leq t$. $\sum_{n=1}^{N_t} g(S_n) = \sum_{n=0}^\infty I_{(0,t]}(S_n)g(S_n)$. In the result of Example 15.2.10, replace g by $I_{(0,t]}g$ and note that

$$\int_0^\infty I_{(0,t]}(u)g(u)\, du = \int_0^t g(u)\, du$$

Example 15.2.14 Replacement costs, finite horizon

Under the condition of Example 15.2.11, consider the replacement costs over a two-year period.

Solution

$$E[C] = \lambda c \int_0^t e^{-\alpha u}\, du = \frac{\lambda c}{\alpha}(1 - e^{-\alpha t})$$

Thus, the expected cost for the infinite horizon $\lambda c/\alpha$ is reduced by the factor $1 - e^{-\alpha t}$. For $t = 2$ and the number in Example 15.2.11, the reduction factor is $1 - e^{-0.16} = 0.1479$ to give $E[C] = 60000 \cdot 0.1479 = 8871.37$.

In the important special case that $g(u) = ce^{-\alpha u}$, the exporession for $E[\sum_{n=1}^\infty g(S_n)]$ may be put into a form which does not require the interarrival times to be exponential.

Example 15.2.15 General interarrival, exponential g

Suppose $S_0 = 0$ and $S_n = \sum_{i=1}^n W_i$, where $\{W_i : 1 \leq i\}$ is iid. Let $\{V_n : 1 \leq n\}$ be a class such that each $E[V_n] = c$ and each pair $\{V_n, S_n\}$ is independent. Then for $\alpha > 0$

$$E[C] = E[\sum_{n=1}^{\infty} V_n e^{-\alpha S_n}] = c \cdot \frac{M_W(-\alpha)}{1 - M_W(-\alpha)}$$

where M_W is the moment generating function for W.

DERIVATION

First we note that

$$E[V_n e^{-\alpha S_n}] = cM_{S_n}(-\alpha) = cM_W^n(-\alpha)$$

Hence, by properties of expectation and the geometric series

$$E[C] = c\sum_{n=1}^{\infty} M_W^n(-\alpha) = \frac{M_W(-\alpha)}{1 - M_W(-\alpha)}, \text{ provided } |M_W(-\alpha)| < 1$$

Since $\alpha > 0$ and $W > 0$, we have $0 < e^{-\alpha W} < 1$, so that $M_W(-\alpha) = E[e^{-\alpha W}] < 1$

Example 15.2.16 Uniformly distributed interarrival times

Suppose each W_i ~ uniform (a, b). Then (see Appendix C),

$$M_W(-\alpha) = \frac{e^{-a\alpha} - e^{-b\alpha}}{\alpha(b - a)} \text{ so that } E[C] = c \cdot \frac{e^{-a\alpha} - e^{-b\alpha}}{\alpha(b - a) - [e^{-a\alpha} - e^{-b\alpha}]}$$

Let $a = 1$, $b = 5$, $c = 100$ and $\alpha = 0$. Then,

```
a = 1;
b = 5;
c = 100;
A = 0.08;
MW = (exp(-a*A) - exp(-b*A))/(A*(b - a))
MW =    0.7900
EC = c*MW/(1 - MW)
EC =  376.1643
```

This page titled 15.2: Some Random Selection Problems is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

15.3: Problems on Random Selection

Exercise 15.3.1

(See Exercise 3 from "Problems on Random Variables and Joint Distributions") A die is rolled. Let X be the number of spots that turn up. A coin is flipped X times. Let Y be the number of heads that turn up. Determine the distribution for Y.

Answer available at the end of this section.

Exercise 15.3.2

(See Exercise 4 from "Problems on Random Variables and Joint Distributions") As a variation of Exercise 15.3.1, suppose a pair of dice is rolled instead of a single die. Determine the distribution for Y.

Answer available at the end of this section.

Exercise 15.3.3

(See Exercise 5 from "Problems on Random Variables and Joint Distributions") Suppose a pair of dice is rolled. Let X be the total number of spots which turn up. Roll the pair an additional X times. Let Y be the number of sevens that are thrown on the X rolls. Determine the distribution for Y. What is the probability of three or more sevens?

Answer available at the end of this section.

Exercise 15.3.4

(See Example 7 from "Conditional Expectation, Regression") A number X is chosen by a random selection from the integers 1 through 20 (say by drawing a card from a box). A pair of dice is thrown X times. Let Y be the number of "matches" (i.e., both ones, both twos, etc.). Determine the distribution for Y.

Answer available at the end of this section.

Exercise 15.3.5

(See Exercise 20 from "Problems on Conditional Expectation, Regression") A number X is selected randomly from the integers 1 through 100. A pair of dice is thrown X times. Let Y be the number of sevens thrown on the X tosses. Determine the distribution for Y. Determine $E[Y]$ and $P(Y \leq 20)$.

Answer available at the end of this section.

Exercise 15.3.6

(See Exercise 21 from "Problems on Conditional Expectation, Regression") A number X is selected randomly from the integers 1 through 100. Each of two people draw X times independently and randomly a number from 1 to 10. Let Y be the number of matches (i.e., both draw ones, both draw twos, etc.). Determine the distribution for Y. Determine $E[Y]$ and $P(Y \leq 10)$.

Answer available at the end of this section.

Exercise 15.3.7

Suppose the number of entries in a contest is N ~ binomial (20, 0.4). There are four questions. Let Y_i be the number of questions answered correctly by the ith contestant. Suppose the Y_i are iid, with common distribution

$$Y = [1\ 2\ 3\ 4]\ PY = [0.2\ 0.4\ 0.3\ 0.1]$$

Let D be the total number of correct answers. Determine $E[D]$, $\mathrm{Var}[D]$, $P(15 \le D \le 25)$, and $P(10 \le D \le 30)$.

Answer available at the end of this section.

Exercise 15.3.8

Game wardens are making an aerial survey of the number of deer in a park. The number of herds to be sighted is assumed to be a random variable N ~ binomial (20, 0.5). Each herd is assumed to be from 1 to 10 in size, with probabilities

Value	1	2	3	4	5	6	7	8	9	10
Probability	0.05	0.10	0.15	0.20	0.15	0.10	0.10	0.05	0.05	0.05

Let D be the number of deer sighted under this model. Determine $P(D \le t)$ for $t = 25, 50, 75, 100$ and $P(D \ge 90)$.

Answer available at the end of this section.

Exercise 15.3.9

A supply house stocks seven popular items. The table below shows the values of the items and the probability of each being selected by a customer.

Value	12.50	25.00	30.50	40.00	42.50	50.00	60.00
Probability	0.10	0.15	0.20	0.20	0.15	0.10	0.10

Suppose the purchases of customers are iid, and the number of customers in a day is binomial (10,0.5). Determine the distribution for the total demand D.

a. How many different possible values are there? What is the maximum possible total sales?
b. Determine $E[D]$ and $P(D \le t)$ for $t = 100, 150, 200, 250, 300$.
   Determine $P(100 < D \le 200)$.

Answer available at the end of this section.

Exercise 15.3.10

A game is played as follows:

1. A wheel is spun, giving one of the integers 0 through 9 on an equally likely basis.
2. A single die is thrown the number of times indicated by the result of the spin of the wheel. The number of points made is the total of the numbers turned up on the sequence of throws of the die.
3. A player pays sixteen dollars to play; a dollar is returned for each point made.

Let Y represent the number of points made and $X = Y - 16$ be the net gain (possibly negative) of the player. Determine the maximum value of

$X, E[X], \mathrm{Var}[X], P(X > 0), P(X \geq 10), P(X \geq 16)$

Answer available at the end of this section.

Exercise 15.3.11

Marvin calls on four customers. With probability $p_1 = 0.6$ he makes a sale in each case. Geraldine calls on five customers, with probability $p_2 = 0.5$ of a sale in each case. Customers who buy do so on an iid basis, and order an amount Y_i (in dollars) with common distribution:

$$Y = [200\ 220\ 240\ 260\ 280\ 300]\ PY = [0.10\ 0.15\ 0.25\ 0.25\ 0.15\ 0.10]$$

Let D_1 be the total sales for Marvin and D_2 the total sales for Geraldine. Let $D = D_1 + D_2$. Determine the distribution and mean and variance for D_1, D_2, and D. Determine $P(D_1 \geq D_2)$ and $P(D \geq 1500)$, $P(D \geq 1000)$, and $P(D \geq 750)$.

Answer available at the end of this section.

Exercise 15.3.12

A questionnaire is sent to twenty persons. The number who reply is a random number N ~ binomial (20, 0.7). If each respondent has probability $p = 0.8$ of favoring a certain proposition, what is the probability of ten or more favorable replies? Of fifteen or more?

Answer available at the end of this section.

Exercise 15.3.13

A random number N of students take a qualifying exam. A grade of 70 or more earns a pass. Suppose N ~ binomial (20, 0.3). If each student has probability $p = 0.7$ of making 70 or more, what is the probability all will pass? Ten or more will pass?

Answer available at the end of this section.

Exercise 15.3.14

Five hundred questionnaires are sent out. The probability of a reply is 0.6. The probability that a reply will be favorable is 0.75. What is the probability of at least 200, 225, 250 favorable replies?

Answer available at the end of this section.

Exercise 15.3.15

Suppose the number of Japanese visitors to Florida in a week is $N1$ ~ Poisson (500) and the number of German visitors is $N2$ ~ Poisson (300). If 25 percent of the Japanese and 20 percent of the Germans visit Disney World, what is the distribution for the total number D of German and Japanese visitors to the park? Determine $P(D \geq k)$ for $k = 150, 155, \cdots, 245, 250$.

Answer available at the end of this section.

Exercise 15.3.16

A junction point in a network has two incoming lines and two outgoing lines. The number of incoming messages N_1 on line one in one hour is Poisson (50); on line 2 the number is N_2 ~ Poisson (45). On incoming line 1 the messages have probability $P_{1a} = 0.33$ of leaving on outgoing line a and $1 - p_{1a}$ of leaving on line b. The messages coming in on line 2 have probability $p_{2a} = 0.47$ of leaving on line a. Under the usual independence assumptions, what is the distribution of outgoing messages on line a? What are the probabilities of at least 30, 35, 40 outgoing messages on line a?

Answer available at the end of this section.

Exercise 15.3.17

A computer store sells Macintosh, HP, and various other IBM compatible personal computers. It has two major sources of customers:

1. Students and faculty from a nearby university
2. General customers for home and business computing. Suppose the following assumptions are reasonable for monthly purchases.

- The number of university buyers $N1$ ~ Poisson (30). The probabilities for Mac, HP, others are 0.4, 0.2, 0.4, respectively.
- The number of non-university buyers $N2$ ~ Poisson (65). The respective probabilities for Mac, HP, others are 0.2, 0.3, 0.5.
- For each group, the composite demand assumptions are reasonable, and the two groups buy independently.

What is the distribution for the number of Mac sales? What is the distribution for the total number of Mac and Dell sales?

Answer available at the end of this section.

Exercise 15.3.18

The number N of "hits" in a day on a Web site on the internet is Poisson (80). Suppose the probability is 0.10 that any hit results in a sale, is 0.30 that the result is a request for information, and is 0.60 that the inquirer just browses but does not identify an interest. What is the probability of 10 or more sales? What is the probability that the number of sales is at least half the number of information requests (use suitable simple approximations)?

Answer available at the end of this section.

Exercise 15.3.19

The number N of orders sent to the shipping department of a mail order house is Poisson (700). Orders require one of seven kinds of boxes, which with packing costs have distribution

Cost (dollars)	0.75	1.25	2.00	2.50	3.00	3.50	4.00
Probability	0.10	0.15	0.15	0.25	0.20	0.10	0.05

What is the probability the total cost of the \$2.50 boxes is no greater than \$475? What is the probability the cost of the \$2.50 boxes is greater than the cost of the \$3.00 boxes? What is the probability the cost of the \$2.50 boxes is not more than \$50.00 greater than the cost of the \$3.00 boxes? *Suggestion*. Truncate the Poisson distributions at about twice the mean value.

Answer available at the end of this section.

Exercise 15.3.20

One car in 5 in a certain community is a Volvo. If the number of cars passing a traffic check point in an hour is Poisson (130), what is the expected number of Volvos? What is the probability of at least 30 Volvos? What is the probability the number of Volvos is between 16 and 40 (inclusive)?

Answer available at the end of this section.

Exercise 15.3.21

A service center on an interstate highway experiences customers in a one-hour period as follows:

- Northbound: Total vehicles: Poisson (200). Twenty percent are trucks.
- Southbound: Total vehicles: Poisson (180). Twenty five percent are trucks.

- Each truck has one or two persons, with respective probabilities 0.7 and 0.3.
- Each car has 1, 2, 3, 4, or 5 persons, with probabilities 0.3, 0.3, 0.2, 0.1, 0.1, respectively

Under the usual independence assumptions, let D be the number of persons to be served. Determine $E[D]$, $\mathrm{Var}[D]$, and the generating function $g_D(s)$.

Answer available at the end of this section.

Exercise 15.3.22

The number N of customers in a shop in a given day is Poisson (120). Customers pay with cash or by MasterCard or Visa charge cards, with respective probabilties 0.25, 0.40, 0.35. Make the usual independence assumptions. Let N_1, N_2, N_3 be the numbers of cash sales, MasterCard charges, Visa card charges, respectively. Determine $P(N_1 \geq 30)$, $P(N_2 \geq 60)$, $P(N_3 \geq 50$, and $P(N_2 > N_3)$.

Answer available at the end of this section.

Exercise 15.3.23

A discount retail store has two outlets in Houston, with a common warehouse. Customer requests are phoned to the warehouse for pickup. Two items, a and b, are featured in a special sale. The number of orders in a day from store A is N_A ~ Poisson (30); from store B, the nember of orders is N_B ~ Poisson (40).

For store A, the probability an order for a is 0.3, and for b is 0.7.

For store B, the probability an order for a is 0.4, and for b is 0.6. What is the probability the total order for item b in a day is 50 or more?

Answer available at the end of this section.

Exercise 15.3.24

The number of bids on a job is a random variable N ~ binomial (7, 0.6). Bids (in thousands of dollars) are iid with Y uniform on [3, 5]. What is the probability of at least one bid of $3,500 or less? *Note* that "no bid" is not a bid of 0.

Answer available at the end of this section.

Exercise 15.3.25

The number of customers during the noon hour at a bank teller's station is a random number N with distribution

$$N = 1:10,\ PN = 0.01 * [5\ 7\ 10\ 11\ 12\ 13\ 12\ 11\ 10\ 9]$$

The amounts they want to withdraw can be represented by an iid class having the common distribution Y ~ exponential (0.01). Determine the probabilities that the maximum withdrawal is less than or equal to t for $t = 100, 200, 300, 400, 500$.

Answer available at the end of this section.

Exercise 15.3.26

A job is put out for bids. Experience indicates the number N of bids is a random variable having values 0 through 8, with respective probabilities

Value	0	1	2	3	4	5	6	7	8
Probability	0.05	0.10	0.15	0.20	0.20	0.10	0.10	0.07	0.03

The market is such that bids (in thousands of dollars) are iid, uniform [100, 200]. Determine the probability of at least one bid of \$125,000 or less.

Answer available at the end of this section.

Exercise 15.3.27

A property is offered for sale. Experience indicates the number N of bids is a random variable having values 0 through 10, with respective probabilities

Value	0	1	2	3	4	5	6	7	8	9	10
Probability	0.05	0.15	0.15	0.20	0.10	0.10	0.05	0.05	0.05	0.05	0.05

The market is such that bids (in thousands of dollars) are iid, uniform [150, 200] Determine the probability of at least one bid of \$180,000 or more.

Answer available at the end of this section.

Exercise 15.3.28

A property is offered for sale. Experience indicates the number N of bids is a random variable having values 0 through 8, with respective probabilities

Number	0	1	2	3	4	5	6	7	8
Probability	0.05	0.15	0.15	0.20	0.15	0.10	0.10	0.05	0.05

The market is such that bids (in thousands of dollars) are iid symmetric triangular on [150 250]. Determine the probability of at least one bid of \$210,000 or more.

Answer available at the end of this section.

Exercise 15.3.29

Suppose N ~ binomial (10, 0.3) and the Y_i are iid, uniform on [10, 20]. Let V be the minimum of the N values of the Y_i. Determine $P(V > t)$ for integer values from 10 to 20.

Answer available at the end of this section.

Exercise 15.3.30

Suppose a teacher is equally likely to have 0, 1, 2, 3 or 4 students come in during office hours on a given day. If the lengths of the individual visits, in minutes, are iid exponential (0.1), what is the probability that no visit will last more than 20 minutes.

Answer available at the end of this section.

Exercise 15.3.31

Twelve solid-state modules are installed in a control system. If the modules are not defective, they have practically unlimited life. However, with probability $p = 0.05$ any unit could have a defect which results in a lifetime (in hours) exponential (0.0025). Under the usual independence assumptions, what is the probability the unit does not fail because of a defective module in the first 500 hours after installation?

Answer available at the end of this section.

Exercise 15.3.32

The number N of bids on a painting is binomial (10, 0.3). The bid amounts (in thousands of dollars) Y_i form an iid class, with common density function $f_Y(t) = 0.005(37 - 2t), 2 \le t \le 10$. What is the probability that the maximum amount bid is greater than \$5,000?

Answer available at the end of this section.

Exercise 15.3.33

A computer store offers each customer who makes a purchase of \$500 or more a free chance at a drawing for a prize. The probability of winning on a draw is 0.05. Suppose the times, in hours, between sales qualifying for a drawing is exponential (4). Under the usual independence assumptions, what is the expected time between a winning draw? What is the probability of three or more winners in a ten hour day? Of five or more?

Answer available at the end of this section.

Exercise 15.3.34

Noise pulses arrrive on a data phone line according to an arrival process such that for each $t > 0$ the number N_t of arrivals in time interval $(0, t]$, in hours, is Poisson $(7t)$. The ith pulse has an "intensity" Y_i such that the class $\{Y_i : 1 \le i\}$ is iid, with the common distribution function $F_Y(u) = 1 - e^{-2u^2}$ for $u \ge 0$. Determine the probability that in an eight-hour day the intensity will not exceed two.

Answer available at the end of this section.

Exercise 15.3.35

The number N of noise bursts on a data transmission line in a period $(0, t]$ is Poisson (μ). The number of digit errors caused by the ith burst is Y_i, with the class $\{Y_i : 1 \leq i\}$ iid, $Y_i - 1$ ~ geometric (p). An error correcting system is capable or correcting five or fewer errors in any burst. Suppose $\mu = 12$ and $p = 0.35$. What is the probability of no uncorrected error in two hours of operation?

Answer available at the end of this section.

Answers

Exercise 15.3.1

```
PX = [0 (1/6)*ones(1,6)];
PY = [0.5 0.5];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  PX
Enter gen fn COEFFICIENTS for gY  PY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
disp(gD)             % Compare with P8-3
         0    0.1641
    1.0000    0.3125
    2.0000    0.2578
    3.0000    0.1667
    4.0000    0.0755
    5.0000    0.0208
    6.0000    0.0026
```

Exercise 15.3.2

```
PN = (1/36)*[0 0 1 2 3 4 5 6 5 4 3 2 1];
PY = [0.5 0.5];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  PN
Enter gen fn COEFFICIENTS for gY  PY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
disp(gD)
         0    0.0269
    1.0000    0.1025
    2.0000    0.1823
    3.0000    0.2158
    4.0000    0.1954
    5.0000    0.1400
    6.0000    0.0806
    7.0000    0.0375
    8.0000    0.0140     % (Continued next page)
    9.0000    0.0040
   10.0000    0.0008
   11.0000    0.0001
   12.0000    0.0000
```

Exercise 15.3.3

```
PX = (1/36)*[0 0 1 2 3 4 5 6 5 4 3 2 1];
PY = [5/6 1/6];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  PX
Enter gen fn COEFFICIENTS for gY  PY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
disp(gD)
         0    0.3072
    1.0000    0.3660
    2.0000    0.2152
    3.0000    0.0828
    4.0000    0.0230
    5.0000    0.0048
    6.0000    0.0008
    7.0000    0.0001
    8.0000    0.0000
    9.0000    0.0000
   10.0000    0.0000
   11.0000    0.0000
   12.0000    0.0000
   P = (D>=3)*PD'
P =  0.1116
```

Exercise 15.3.4

```
gN = (1/20)*[0 ones(1,20)];
gY = [5/6 1/6];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.

disp(gD)
         0    0.2435
    1.0000    0.2661
    2.0000    0.2113
    3.0000    0.1419
    4.0000    0.0795
    5.0000    0.0370
    6.0000    0.0144
    7.0000    0.0047
    8.0000    0.0013
    9.0000    0.0003
   10.0000    0.0001
   11.0000    0.0000
   12.0000    0.0000
   13.0000    0.0000
   14.0000    0.0000
   15.0000    0.0000
   16.0000    0.0000
   17.0000    0.0000
   18.0000    0.0000
```

```
   19.0000    0.0000
   20.0000    0.0000
```

Exercise 15.3.5

```
gN = 0.01*[0 ones(1,100)];
gY = [5/6 1/6];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
EY = dot(D,PD)
EY =   8.4167
P20 = (D<=20)*PD'
P20 =  0.9837
```

Exercise 15.3.6

```
gN = 0.01*[0 ones(1,100)];
gY = [0.9 0.1];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
EY = dot(D,PD)
EY =  5.0500
P10 = (D<=10)*PD'
P10 = 0.9188
```

Exercise 15.3.7

```
gN = ibinom(20,0.4,0:20);
gY = 0.1*[0 2 4 3 1];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
ED = dot(D,PD)
ED =  18.4000
VD = (D.^2)*PD' - ED^2
VD =  31.8720
P1 = ((15<=D)&(D<=25))*PD'
P1 =   0.6386
P2 = ((10<=D)&(D<=30))*PD'
P2 =   0.9290
```

Exercise 15.3.8

```
gN = ibinom(20,0.5,0:20);
gY = 0.01*[0 5 10 15 20 15 10 10 5 5 5];
gend
```

```
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
k = [25 50 75 100];
P = zeros(1,4);
for i = 1:4
    P(i) = (D<=k(i))*PD';
end
disp(P)
    0.0310    0.5578    0.9725    0.9998
```

Exercise 15.3.9

```
gN = ibinom(10,0.5,0:10);
Y  = [12.5 25 30.5 40 42.5 50 60];
PY = 0.01*[10 15 20 20 15 10 10];
mgd
Enter gen fn COEFFICIENTS for gN  gN
Enter VALUES for Y  Y
Enter PROBABILITIES for Y  PY
Values are in row matrix D; probabilities are in PD.
To view the distribution, call for mD.
s = size(D)
s =    1   839
M = max(D)
M =    590
t = [100 150 200 250 300];
P = zeros(1,5);
for i = 1:5
    P(i) = (D<=t(i))*PD';
end
disp(P)
    0.1012    0.3184    0.6156    0.8497    0.9614
P1 = ((100<D)&(D<=200))*PD'
P1 =   0.5144
```

Exercise 15.3.10

```
gn = 0.1*ones(1,10);
gy = (1/6)*[0 ones(1,6)];
[Y,PY] = gendf(gn,gy);
[X,PX] = csort(Y-16,PY);
M = max(X)
M =  38
EX = dot(X,PX)               % Check EX = En*Ey - 16 = 4.5*3.5
EX  =  -0.2500               % 4.5*3.5 - 16 = -0.25
VX = dot(X.^2,PX) - EX^2
VX =  114.1875
Ppos = (X>0)*PX'
Ppos =  0.4667
P10 = (X>=10)*PX'
P10 =   0.2147
P16 = (X>=16)*PX'
P16 =   0.0803
```

Exercise 15.3.11

```
gnM = ibinom(4,0.6,0:4);
gnG = ibinom(5,0.5,0:5);
Y = 200:20:300;
PY = 0.01*[10 15 25 25 15 10];
[D1,PD1] = mgdf(gnM,Y,PY);
[D2,PD2] = mgdf(gnG,Y,PY);
ED1 = dot(D1,PD1)
ED1 =  600.0000              % Check: ED1 = EnM*EY = 2.4*250
VD1 = dot(D1.^2,PD1) - ED1^2
VD1 =    6.1968e+04
ED2 = dot(D2,PD2)
ED2 =  625.0000              % Check: ED2 = EnG*EY = 2.5*250
VD2 = dot(D2.^2,PD2) - ED2^2
VD2 =    8.0175e+04
[D1,D2,t,u,PD1,PD2,P] = icalcf(D1,D2,PD1,PD2);
Use array opertions on matrices X, Y, PX, PY, t, u, and P
[D,PD] = csort(t+u,P);
ED = dot(D,PD)
ED =   1.2250e+03
eD = ED1 + ED2               % Check: ED = ED1 + ED2
eD =   1.2250e+03            % (Continued next page)

VD = dot(D.^2,PD) - ED^2
VD =   1.4214e+05
vD = VD1 + VD2            % Check: VD = VD1 + VD2
vD =   1.4214e+05
P1g2 = total((t>u).*P)
P1g2 = 0.4612
k = [1500 1000 750];
PDk = zeros(1,3);
for i = 1:3
   PDk(i) = (D>=k(i))*PD';
end
disp(PDk)
    0.2556    0.7326    0.8872
```

Exercise 15.3.12

```
gN = ibinom(20,0.7,0:20);
gY = [0.2 0.8];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
P10 = (D>=10)*PD'
P10 =   0.7788
P15 = (D>=15)*PD'
P15 =   0.0660
pD = ibinom(20,0.7*0.8,0:20);  % Alternate: use D binomial (pp0)
D = 0:20;
p10 = (D>=10)*pD'
p10 =  0.7788
p15 = (D>=15)*pD'
p15 =  0.0660
```

Exercise 15.3.13

```
gN = ibinom(20,0.3,0:20);
gY = [0.3 0.7];
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Results are in N, PN, Y, PY, D, PD, P
May use jcalc or jcalcf on N, D, P
To view the distribution, call for gD.
Pall = (D==20)*PD'
Pall =  2.7822e-14
pall = (0.3*0.7)^20    % Alternate: use D binomial (pp0)
pall =  2.7822e-14
P10 = (D >= 10)*PD'
P10 = 0.0038
```

Exercise 15.3.14

```
n = 500;
p = 0.6;
p0 = 0.75;
D = 0:500;
PD = ibinom(500,p*p0,D);
k = [200 225 250];
P = zeros(1,3);
for i = 1:3
   P(i) = (D>=k(i))*PD';
end
disp(P)
    0.9893    0.5173    0.0140
```

Exercise 15.3.15

JD ~ Poisson (500*0.25 = 125); GD ~ Poisson (300*0.20 = 60); D ~ Poisson (185).

```
k = 150:5:250;
PD = cpoisson(185,k);
disp([k;PD]')
  150.0000    0.9964
  155.0000    0.9892
  160.0000    0.9718
  165.0000    0.9362
  170.0000    0.8736
  175.0000    0.7785
  180.0000    0.6532
  185.0000    0.5098
  190.0000    0.3663
  195.0000    0.2405
  200.0000    0.1435
  205.0000    0.0776
  210.0000    0.0379
  215.0000    0.0167
  220.0000    0.0067
  225.0000    0.0024
  230.0000    0.0008
  235.0000    0.0002
  240.0000    0.0001
  245.0000    0.0000
  250.0000    0.0000
```

Exercise 15.3.16

```
m1a = 50*0.33;  m2a = 45*0.47; ma = m1a + m2a;
PNa = cpoisson(ma,[30 35 40])
PNa =   0.9119    0.6890    0.3722
```

Exercise 15.3.17

Mac sales Poisson (30*0.4 + 65*0.2 = 25); HP sales Poisson (30*0.2 + 65*0.3 = 25.5); total Mac plus HP sales Poisson(50.5).

Exercise 15.3.18

```
X = 0:30;
Y = 0:80;
PX = ipoisson(80*0.1,X);
PY = ipoisson(80*0.3,Y);
icalc:  X  Y  PX  PY
- - - - - - - - - - - -
PX10 = (X>=10)*PX'    % Approximate calculation
PX10 =  0.2834
pX10 = cpoisson(8,10)   % Direct calculation
pX10 =  0.2834
M = t>=0.5*u;
PM = total(M.*P)
PM =    0.1572
```

Exercise 15.3.19

```
X = 0:400;
Y = 0:300;
PX = ipoisson(700*0.25,X);
PY = ipoisson(700*0.20,Y);
icalc
Enter row matrix of X-values  X
Enter row matrix of Y-values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
 Use array operations on matrices X, Y, PX, PY, t, u, and P
P1 = (2.5*X<=475)*PX'
P1 =   0.8785
M = 2.5*t<=(3*u + 50);
PM = total(M.*P)
PM =   0.7500
```

Exercise 15.3.20

```
P1 = cpoisson(130*0.2,30) = 0.2407
P2 = cpoisson(26,16) - cpoisson(26,41) = 0.9819
```

Exercise 15.3.21

T ~ Poisson (200*0.2 = 180*0.25 = 85), P ~ Poisson (200*0.8 + 180*0.75 = 295).

```
a =   85
b = 200*0.8 + 180*0.75
b =  295
YT = [1 2];
PYT = [0.7 0.3];
EYT = dot(YT,PYT)
EYT =  1.3000
VYT = dot(YT.^2,PYT) - EYT^2
VYT =  0.2100
YP = 1:5;
```

```
PYP = 0.1*[3 3 2 1 1];
EYP = dot(YP,PYP)
EYP =  2.4000
VYP = dot(YP.^2,PYP) - EYP^2
VYP =   1.6400
EDT = 85*EYT
EDT =  110.5000
EDP = 295*EYP
EDP =  708.0000
ED = EDT + EDP
ED =  818.5000
VT = 85*(VYT + EYT^2)
VT =  161.5000
VP = 295*(VYP + EYP^2)
VP =    2183
VD = VT + VP
VD =   2.2705e+03

NT = 0:180;                    % Possible alternative
gNT = ipoisson(85,NT);
gYT = 0.1*[0 7 3];
[DT,PDT] = gendf(gNT,gYT);
EDT = dot(DT,PDT)
EDT =  110.5000
VDT = dot(DT.^2,PDT) - EDT^2
VDT =  161.5000
NP = 0:500;
gNP = ipoisson(295,NP);
gYP = 0.1*[0 3 2 2 1 1];
[DP,PDP] = gendf(gNP,gYP);     %  Requires too much memory
```

$$g_{DT}(s) = \exp(85(0.7s + 0.3s^2 - 1))\ g_{DP}(s) = \exp(295(0.1(3s + 3s^2 2s^3 + s^4 + s^5) - 1))$$
$$g_D(s) = g_{DT}(s) g_{DP}(s)$$

Exercise 15.3.22

```
X = 0:120;
PX = ipoisson(120*0.4,X);
Y = 0:120;
PY = ipoisson(120*0.35,Y);
icalc
Enter row matrix of X values  X
Enter row matrix of Y values  Y
Enter X probabilities  PX
Enter Y probabilities  PY
Use array opertions on matrices X, Y, PX, PY, t, u, and P
M = t > u;
PM = total(M.*P)
PM =    0.7190
```

Exercise 15.3.23

P = cpoisson(30*0.7+40*0.6,50) = 0.2468

Exercise 15.3.24

```
% First solution ---  FY(t) = 1 - gN[P(Y>t)]
P = 1-(0.4 + 0.6*0.75)^7
P  =    0.6794
% Second solution --- Positive number of satisfactory bids,
% i.e. the outcome is indicator for event E, with P(E) = 0.25
pN = ibinom(7,0.6,0:7);
```

```
gY = [3/4 1/4];          % Generator function for indicator
[D,PD] = gendf(pN,gY);   % D is number of successes
Pa = (D>0)*PD'           % D>0 means at least one successful bid
Pa =    0.6794
```

Exercise 15.3.25

Use $F_W(t) = g_N[P(Y \le T)]$

```
gN = 0.01*[0 5 7 10 11 12 13 12 11 10 9];
t = 100:100:500;
PY = 1 - exp(-0.01*t);
FW = polyval(fliplr(gN),PY)  % fliplr puts coeficients in
                             % descending order of powers
FW =    0.1330    0.4598    0.7490    0.8989    0.9615
```

Exercise 15.3.26

Probability of a successful bid $PY = (125 - 100)/100 = 0.25$

```
PY =0.25;
gN = 0.01*[5 10 15 20 20 10 10 7 3];
P = 1 - polyval(fliplr(gN),PY)
P =  0.9116
```

Exercise 15.3.27

Consider a sequence of N trials with probabilty $p = (180 - 150)/50 = 0.6$.

```
gN = 0.01*[5 15 15 20 10 10 5 5 5 5 5];
gY = [0.4 0.6];
[D,PD] = gendf(gN,gY);
P = (D>0)*PD'
P =   0.8493
```

Exercise 15.3.28

```
gN = 0.01*[5 15 15 20 15 10 10 5 5];
PY = 0.5 + 0.5*(1 - (4/5)^2)
PY = 0.6800
>> PW = 1 - polyval(fliplr(gN),PY)
PW = 0.6536
%alternate
gY = [0.68 0.32];
[D,PD] = gendf(gN,gY);
P = (D>0)*PD'
P = 0.6536
```

Exercise 15.3.29

```
gN = ibinom(10,0.3,0:10);
t = 10:20;
p = 0.1*(20 - t);
P = polyval(fliplr(gN),p) - 0.7^10
P =
  Columns 1 through 7
    0.9718    0.7092    0.5104    0.3612    0.2503    0.1686    0.1092
  Columns 8 through 11
    0.0664    0.0360    0.0147         0
Pa = (0.7 + 0.3*p).^10 - 0.7^10     % Alternate form of gN
Pa =
  Columns 1 through 7
    0.9718    0.7092    0.5104    0.3612    0.2503    0.1686    0.1092
  Columns 8 through 11
```

```
    0.0664    0.0360    0.0147         0
```

Exercise 15.3.30

```
gN = 0.2*ones(1,5);
p = 1 - exp(-2);
FW = polyval(fliplr(gN),p)
FW =    0.7635
gY = [p 1-p];               % Alternate
[D,PD] = gendf(gN,gY);
PW = (D==0)*PD'
PW =    0.7635
```

Exercise 15.3.31

```
p = 1 - exp(-0.0025*500);
FW = (0.95 + 0.05*p)^12
FW =   0.8410
gN = ibinom(12,0.05,0:12);
gY = [p 1-p];
[D,PD] = gendf(gN,gY);
PW = (D==0)*PD'
PW =   0.8410
```

Exercise 15.3.32

$$P(Y \le 5) = 0.005 \int_2^5 (37 - 2t)\ dt = 0.45$$

```
p = 0.45;
P = 1 - (0.7 + 0.3*p)^10
P =   0.8352
gN = ibinom(10,0.3,0:10);
gY = [p 1-p];
[D,PD] = gendf(gN,gY);  % D is number of "successes"
Pa = (D>0)*PD'
Pa =  0.8352
```

Exercise 15.3.33

N_t ~ Poisson (λt), N_{Dt} ~ Poisson (λpt), W_{Dt} exponential (λp).

```
p = 0.05;
t = 10;
lambda = 4;
EW = 1/(lambda*p)
EW =    5
PND10 = cpoisson(lambda*p*t,[3 5])
PND10 =  0.3233    0.0527
```

Exercise 15.3.34

N_8 is Poisson (7*8 = 56) $g_N(s) = e^{56(s-1)}$.

```
t = 2;
FW2 = exp(56*(1 - exp(-t^2) - 1))
FW2 =   0.3586
```

Exercise 15.3.35

$F_W(k) = g_N[P(Y \le k)]P(Y \le k) - 1 - q^{k-1}$ N_t ~ Poisson (12t)

```
q = 1 - 0.35;
k = 5;
t = 2;
mu = 12;
```

```
FW = exp(mu*t*(1 - q^(k-1) - 1))
FW =  0.0138
```

This page titled 15.3: Problems on Random Selection is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

16: CONDITIONAL INDEPENDENCE, GIVEN A RANDOM VECTOR

16.1: Conditional Independence, Given a Random Vector
16.2: Elements of Markov Sequences
16.3: Problems on Conditional Independence, Given a Random Vector

This page titled 16: Conditional Independence, Given a Random Vector is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

16.1: Conditional Independence, Given a Random Vector

In the unit on Conditional Independence , the concept of conditional independence of events is examined and used to model a variety of common situations. In this unit, we investigate a more general concept of conditional independence, based on the theory of conditional expectation. This concept lies at the foundations of Bayesian statistics, of many topics in decision theory, and of the theory of Markov systems. We examine in this unit, very briefly, the first of these. In the unit on Markov Sequences, we provide an introduction to the third.

The concept

The definition of conditional independence of events is based on a product rule which may be expressed in terms of conditional expectation, given an event. The pair $\{A, B\}$ is conditionally independent, given C, iff

$$E[I_A I_B|C] = P(AB|C) = P(A|C)P(B|C) = E[I_A|C]E[I_B|C]$$

If we let $A = X^{-1}(M)$ and $B = Y^{-1}(N)$, then $I_A = I_M(X)$ and $I_B = I_N(Y)$. It would be reasonable to consider the pair $\{X, Y\}$ conditionally independent, given event C, iff the product rule

$$E[I_M(X)I_N(Y)|C] = E[I_M(X)|C]E[I_N(Y)|C]$$

holds for all reasonable M and N (technically, all Borel M and N). This suggests a possible extension to conditional expectation, given a random vector. We examine the following concept.

Definition

The pair $\{X, Y\}$ is *conditionally independent, givenZ*, designated $\{X, Y\}$ ci $|Z$, iff

$$E[I_M(X)I_N(Y)|Z] = E[I_M(X)|Z]E[I_N(Y)|Z] \text{ for all Borel } M.\ N$$

Remark. Since it is not necessary that X, Y, or Z be real valued, we understand that the sets M and N are on the codomains for X and Y, respectively. For example, if X is a three dimensional random vector, then M is a subset of R^3.

As in the case of other concepts, it is useful to identify some key properties, which we refer to by the numbers used in the table in Appendix G. We note two kinds of equivalences. For example, the following are equivalent.

(CI1) $E[I_M(X)I_N(Y)|Z] = E[I_M(X)|Z][E[I_N(Y)|Z]$ a.s. for all Borel sets M, N

(CI5) $E[g(X, Z)h(Y, Z)|Z] = E[g(X, Z)|Z]E[h(Y, Z)|Z]$ a.s. for all Borel functions g, h

Because the indicator functions are special Borel functions, (CI1) is a special case of (CI5). To show that (CI1) implies (CI5), we need to use linearity, monotonicity, and monotone convergence in a manner similar to that used in extending properties (CE1) to (CE6) for conditional expectation. A second kind of equivalence involves various patterns. The properties (CI1), (CI2), (CI3), and (CI4) are equivalent, with (CI1) being the defining condition for $\{X, Y\}$ ci $|Z$.

(CI1) $E[I_M(X)I_N(Y)|Z] = E[I_M(X)|Z][E[I_N(Y)|Z]$ a.s. for all Borel sets M, N

(CI2) $E[I_M(X)|Z, Y] = E[I_M(X)|Z]$ a.s. for all Borel sets M

(CI3) $E[I_M(X)I_Q(Z)|Z, Y] = E[I_M(X)I_Q(Z)|Z]$ a.s. for all Borel sets M, Q

(CI4) $E[I_M(X)I_Q(Z)|Y] = E\{E[I_M(X)I_Q(Z)|Z]|Y\}$ a.s. for all Borel sets M, Q

As an example of the kinds of argument needed to verify these equivalences, we show the equivalence of (CI1) and (CI2).

- (CI1) implies (CI2). Set $e_1(Y, Z) = E[I_M(X)|Z, Y]$ and $e_2(Y, Z) = E[I_M(X)|Z]$. If we show

$$E[I_N(Y)I_Q(Z)e_1(Y,Z) = E[I_N(Y)I_Q(Z)e_2(Y,Z)] \text{ for all Borel } N, Q$$

then by the uniqueness property (E5b) for expectation we may assert $e_1(Y,Z) = e_2(Y,Z)$ a.s. Using the defining property (CE1) for conditional expectation, we have

$$E\{I_N(Y)I_Q(Z)E[I_M(X)|Z,Y]\} = E[I_N(Y)I_Q(Z)I_M(X)]$$

On the other hand, use of (CE1), (CE8), (CI1), and (CE1) yields

$$E\{I_N(Y)I_Q(Z)E[I_M(X)|Z]\} = E\{I_Q(Z)E[I_N(Y)E[I_M(X)|Z]|Z]\}$$

$$= E\{I_Q(Z)E[I_M(X)|Z]E[I_N(Y)|Z]\} = E\{I_Q(Z0E[I_M(X)I_N(Y)|Z]\}$$

$$= E[I_N(Y)I_Q(Z0I_M(X)$$

which establishes the desired equality.

- (CI2) implies (CI1). Using (CE9), (CE8), (CI2), and (CE8), we have

$$E[I_M(X)I_N(Y)|Z] = E\{E[I_M(X)I_N(Y)|Z,Y]|Z\}$$

$$= E[I_N(Y)E[I_M(X)|Z,Y]|Z\} = E\{I_N(Y)E[I_M(X)|Z]|Z\}$$

$$= E[I_M(X)|Z]E[I_N(Y)|Z]$$

Use of property (CE8) shows that (CI2) and (CI3) are equivalent. Now just as (CI1) extends to (CI5), so also (CI3) is equivalent to

(CI6) $E[g(X,Z)|Z,Y] = E[g(X,Z)|Z]$ a.s. for all Borel functions g

Property (CI6) provides an important *interpretation* of conditional independence:

$E[g(X,Z)|Z]$ is the best mean-square estimator for $g(X,Z)$, given knowledge of Z. The conditon $\{X,Y\}$ ci $|Z$ implies that additional knowledge about *Y* does not modify that best estimate. This interpretation is often the most useful as a modeling assumption.

Similarly, property (CI4) is equivalent to

(CI8) $E[g(X,Z)|Y] = E\{E[g(X,Z)|Z]|Y\}$ a.s. for all Borel functions g

The additional properties in Appendix G are useful in a variety of contexts, particularly in establishing properties of Markov systems. We refer to them as needed.

The Bayesian approach to statistics

In the *classical* approach to statistics, a fundamental problem is to obtain information about the population distribution from the distribution in a simple random sample. There is an inherent difficulty with this approach. Suppose it is desired to determine the population mean μ. Now μ is an unknown quantity about which there is uncertainty. However, since it is a constant, we cannot assign a probability such as $P(a < \mu \leq b)$. This has no meaning.

The *Bayesian* approach makes a fundamental change of viewpoint. Since the population mean is a quantity about which there is uncertainty, it is *modeled as a random variable* whose value is to be determined by experiment. In this view, the population distribution is conceived as randomly selected from a class of such distributions. One way of expressing this idea is to refer to a *state of nature*. The population distribution has been "selected by nature" from a class of distributions. The mean value is thus a random variable whose value is determined by this selection. To implement this point of view, we assume

The value of the parameter (say μ in the discussion above) is a "realization" of a *parameter random variable* H. If two or more parameters are sought (say the mean and variance), they may be considered components of a parameter random vector.
The population distribution is a *conditional distribution*, given the value of H.

The Bayesian model

If X is a random variable whose distribution is the population distribution and H is the parameter random variable, then $\{X, H\}$ have a joint distribution.

For each u in the range of H, we have a *conditional distribution* for X, given $H = u$.
We assume a *prior distribution* for H. This is based on previous experience.
We have a *random sampling process*, given H: i.e., $\{X_i : 1 \le i \le n\}$ is conditionally iid, given H. Let $W = (X_1, X_2, \cdots, X_n)$ and consider the joint conditional distribution function

$$F_{W|H}(t_1, t_2, \cdots, t_n|u) = P(X_1 \le t_1, X_2 \le t_2, \cdots, X_n \le t_n|H = u)$$

$$= E[\prod_{i=1}^n I_{(-\infty,t_i]}(X_i)|H = u]] = \prod_{i=1}^n E[I_{(-\infty,t_i]}(X_i)|H = u] = \prod_{i=1}^n F_{X|H}(t_i|u)$$

If X has conditional density, given *H*, then a similar product rule holds.

Population proportion

We illustrate these ideas with one of the simplest, but most important, statistical problems: that of determining the proportion of a population which has a particular characteristic. Examples abound. We mention only a few to indicate the importance.

The proportion of a population of voters who plan to vote for a certain candidate.
The proportion of a given population which has a certain disease.
The fraction of items from a production line which meet specifications.
The fraction of women between the ages eighteen and fifty five who hold full time jobs.

The parameter in this case is the proportion p who meet the criterion. If sampling is at random, then the sampling process is equivalent to a sequence of Bernoulli trials. If H is the parameter random variable and S_n is the number of "successes" in a sample of size n, then the conditional distribution for S_n, given $H = u$, is binomial (n, u). To see this, consider

$$X_i = I_{E_i}, \text{ with } P(E_i|H = u) = E[X_i|H = u] = e(u) = u$$

Anaysis is carried out for each fixed u as in the ordinary Bernoulli case. If

$$S_n = \sum_{i=1}^n X_i = \sum_{i=1}^n I_{E_i}$$

We have the result

$$E[I_{\{k\}}(S_i)|H = u] = P(S_n = k|H = u) = C(n, k)u^k(1 - u)^{n-k} \text{ and } E[S_n|H = u] = nu$$

The objective

We seek to determine the best mean-square estimate of H, given $S_n = k$.
If $H = u$, we know $E[S_n|H] = nu$. Sampling gives $S_n = k$. We make a *Bayesian reversal* to get an exression for $E[H|S_n = k]$.
To complete the task, we must assume a prior distribution for H on the basis of prior knowledge, if any.

The Bayesian reversal

Since $\{S_n = k\}$ is an event with positive probability, we use the definition of the conditional expectation, given an event, and the law of total probability (CE1b) to obtain

$$E[H|S_n = k] = \frac{E[HI_{\{k\}}(S_n)]}{E[I_{\{k\}}(S_n)]} = \frac{E\{HE[I_{\{k\}}(S_n)|H]\}}{E\{E[I_{\{k\}}(S_n)|H]\}} = \frac{\int uE[I_{\{k\}}(S_n)|H = u]f_H(u)\ du}{\int E[I_{\{k\}}(S_n)|H = u]f_H(u)\ du}$$

$$= \frac{C(n, k)\int u^{k+1}(1 - u)^{n-k}f_H(u)\ du}{C(n, k)\int u^k(1 - u)^{n-k}f_H(u)\ du}$$

A prior distribution for H

The beta (r, s) distribution (see Appendix G), proves to be a "natural" choice for this purpose. Its range is the unit interval, and by

proper choice of parameters r, s, the density function can be given a variety of forms (see Figures 16.1.1 and 16.2.2).

Figure one is a graph labeled Beta (r, s) density -- r = 2. The horizontal axis is labeled as t, and the vertical axis as Density. The values on the horizontal axis range from 0 to 1 in increments of 0.1, and the values on the vertical axis range from 0 to 4.5 in increments of 0.5. There are three plotted shapes on this graph. The most simple is a linear path beginning at the origin in the bottom-left corner of the graph that continues with a constant positive slope across the graph and halfway up the vertical axis, terminating at (1, 2). It is labeled as s = 1. The second is a curve that begins at the origin with a positive, decreasing slope. It continues increasing across the page until it flattens at (0.5, 1.5), where it begins decreasing at an increasing rate. It meets the bottom-right corner. The shape is symmetrical, and looks more like a small portion of a circle than a bell curve. This plot is labeled as s = 2.The third is a distribution, again beginning at the origin, that starts out with a sharply positive slope. The distribution increases at a decreasing rate very quickly, where at (0.1, 4.25) it peaks, and begins decreasing at an increasing rate until approximately (0.25, 2), where it begins decreasing at a decreasing rate until it tails off to a horizontal line at approximately (0.6, 0) where it is essentially a horizontal line that continues to the bottom-right edge of the graph. This plot is labeled as s = 10.

Figure 16.1.1. The Beta(r,s) density for $r = 2$, $s = 1, 2, 10$.

Figure two is a graph labeled Beta (r, s) density -- r = 5. The horizontal axis is labeled as t, and the vertical axis as Density. The values on the horizontal axis range from 0 to 1 in increments of 0.1, and the values on the vertical axis range from 0 to 3.5 in increments of 0.5. There are three plotted distributions on the graph. All three maintain some sort of shape that resembles a distorted bell curve, with tapering tails on both sides and a large peak. One plot, labeled s = 5, resembles a symmetrical bell curve, with tails leading to both the bottom-right and bottom-left of the graph, and a peak at approximately (0.5, 2.5) in the middle of the figure. A second plot, labeled s = 10, is not as centered as the previously described plot, as its peak is located at approximately (0.3, 3.25). This distribution is slightly skewed right as a result, and its bell is more slender in shape than the standard normal distribution. The third plot, labeled s = 2, is strongly skewed left, with its peak located at approximately (0.8, 2.5)

Figure 16.1.2. The Beta(r,s) density for $r = 5$, $s = 2, 5, 10$.

Its analysis is based on the integrals

$$\int_0^1 u^{r-1}(1-u)^{s-1}\,du = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} \text{ with } \Gamma(a+1) = a\Gamma(a)$$

For H ~ beta (r, s), the density is given by

$$f_H(t) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} t^{r-1}(1-t)^{s-1} = A(r,s)t^{r-1}(1-t)^{s-1}\ 0 < t < 1$$

For $r \geq 2$, $s \geq 2$, f_H has a maximum at $(r-1)/(r+s-2)$. For r, s positive integers, f_H is a polynomial on [0, 1], so that determination of the distribution function is easy. In any case, straightforward integration, using the integral formula above, shows

$$E[H] = \frac{r}{r+s} \text{ and } \operatorname{Var}[H] = \frac{rs}{(r+s)^2(r+s+1)}$$

If the prior distribution for H is beta (r, s), we may complete the determination of $E[H|S_n = k]$ as follows.

$$E[H|S_n = k] = \frac{A(r,s)\int_0^1 u^{k+1}(1-u)^{n-k}u^{r-1}(1-u)^{s-1}\,du}{A(r,s)\int_0^1 u^{k}(1-u)^{n-k}u^{r-1}(1-u)^{s-1}\,du} = \frac{\int_0^1 u^{k+r}(1-u)^{n+s-k-1}\,du}{\int_0^1 u^{k+r-1}(1-u)^{n+s-k-1}\,du}$$

$$= \frac{\Gamma(r+k+1)\Gamma(n+s-k)}{\Gamma(r+s+n+1)} \cdot \frac{\Gamma(r+s+n)}{\Gamma(r+k)\Gamma(n+s-k)} = \frac{k+r}{n+r+s}$$

We may adapt the analysis above to show that H is conditionally beta $(r+k, s+n-k)$, given $S_n = k$.

$$F_{H|S}(t|k) = \frac{E[I_t(H)I_{\{k\}}(S_n)}{E[I_{\{k\}}(S_n)} \text{ where } I_t(H) = I_{[0,t]}(H)$$

The analysis goes through exactly as for $E[H|S_n = k]$, except that H is replaced by $I_t(H)$. In the integral expression for the numerator, one factor u is replaced by $I_t(u)$. For H ~ beta (r, s), we get

$$F_{H|S}(t|k) = \frac{\Gamma(r+s+n)}{\Gamma(r+k)\Gamma(n+s-k)} \int_0^t u^{k+r-1}(1-u)^{n+s-k-1}\, du = \int_0^t f_{H|S}(u|k)\, du$$

The integrand is the density for beta $(r+k, n+s-k)$.

Any prior information on the distribution for H can be utilized to select suitable r, s. If there is no prior information, we simply take $r = 1$, $s = 1$, which corresponds to

H ~ uniform on (0, 1). The value is as likely to be in any subinterval of a given length as in any other of the same length. The information in the sample serves to modify the distribution for H, conditional upon that information.

Example 16.1.1 Population proportion with a beta prior

It is desired to estimate the portion of the student body which favors a proposed increase in the student blanket tax to fund the campus radio station. A sample of size $n = 20$ is taken. Fourteen respond in favor of the increase. Assuming prior ignorance (i.e., that H beta (1,1)), what is the conditional distribution given $s_{20} = 14$? After the first sample is taken, a second sample of size $n = 20$ is taken, with thirteen favorable responses. Analysis is made using the conditional distribution for the first sample as the prior for the second. Make a new estimate of H.

Figure three is a graph labeled, Condition densities beta (15, 7) and beta (28, 14). The horizontal axis is labeled as t, and the vertical axis is labeled, conditional density. The values on the horizontal axis range from 0 to 1 in increments of 0.1, and the values on the vertical axis range from 0 to 6 in increments of one. There are two plots in this figure. Both are similar in shape, reflecting two beta distributions, with long tails, relatively symmetric in structure, and reaching only one peak of distribution. The first distribution, labeled beta (15, 7), is centered horizontally at 0.7, and reaches a vertical value of conditional density of four. It is slightly skewed to the left, but there is no visible vertical significance further to the left on the horizontal axis past the value 0.3. Likewise, there is no vertical significance past approximately 0.92 in the tail on the right. The second plot, labeled beta (28, 14), is a stronger distribution, centered at approximately 0.68, but reaching a conditional density in respect to the vertical axis of approximately 5.5. The plot again is slightly skewed left, and no significant portion of the graph can be seen in the tails past 0.4 to the left and 0.88 to the right.

Figure 16.1.3. Conditional densities for repeated sampling, Example 16.1.1.

Solution

For the *first sample* the parameters are $r = s = 1$. According the treatment above, H is conditionally beta $(k+r, n+s-k)$ $= (15, 7)$. The density has a maximum at $(r+k-1)/(r+k+n+s-k-2) = k/n$. The conditional expectation, however, is $(r+k)/(r+s+n) = 15/22 \approx 0.6818$.

For the *second sample*, with the conditional distribution as the new prior, we should expect more sharpening of the density about the new mean-square estimate. For the new sample, $n = 20$, $k = 13$, and the prior H ~ beta (15, 7). The new conditional distribution has parameters $r^* = (28 - 1)/(28 + 14 - 2) = 27/40 = 0.6750$. The best estimate of H is 28/(28 + 14) = 2/3. The conditonal densities in the two cases may be plotted with MATLAB (see Figure 16.1.1).

```
t = 0:0.01:1;
plot(t,beta(15,7,t),'k-',t,beta(28,14,t),'k--')
```

As expected, the maximum for the second is somewhat larger and occurs at a slightly smaller t, reflecting the smaller k. And the density in the second case shows less spread, resulting from the fact that prior information from the first sample is incorporated into the analysis of the second sample.

The same result is obtained if the two samples are combined into one sample of size 40.

It may be well to compare the result of Bayesian analysis with that for classical statistics. Since, in the latter, case prior information is not utilized, we make the comparison with the case of no prior knowledge $(r = s = 1)$. For the classical case, the estimator for μ is the sample average; for the Bayesian case with beta prior, the estimate is the conditional expectation of H, given S_n.

$$\text{If } S_n = k\text{: Classical estimate = } k/n \text{ Bayesian estimate = } (k + 1)/(n + 2)$$

For large sample size n, these do not differ significantly. For small samples, the difference may be quite important. The Bayesian estimate is often referred to as the *small sample* estimate, although there is nothing in the Bayesian procedure which calls for small samples. In any event, the Bayesian estimate seems preferable for small samples, and it has the advantage that *prior information may be utilized*. The sampling procedure upgrades the prior distribution.

The essential *idea* of the Bayesian approach is the view that an unknown parameter about which there is uncertainty is modeled as the value of a random variable. The *name* Bayesian comes from the role of Bayesian reversal in the analysis.

The application of Bayesian analysis to the population proportion required Bayesian reversal in the case of discrete S_n. We consider, next, this reversal process when all random variables are absolutely continuous.

The Bayesian reversal for a joint absolutely continuous pair

In the treatment above, we utilize the fact that the conditioning random variable S_n is discrete. Suppose the pair $\{W, H\}$ is jointly absolutely continuous, and $f_{W|H}(t|u)$ and $f_H(u)$ are specified. To determine

$$E[H|W = t] = \int u f_{H|W}(u|t)\ du$$

we need $f_{H|W}(u|t)$. This requires a Bayesian reversal of the conditional densities. Now by definition

$$f_{H|W}(u|t) = \frac{f_{WH}(t,u)}{f_W(t)} \text{ and } f_{WH}(t,u) = f_{W|H}(t|u) f_H(u)$$

Since by the rule for determining the marginal density

$$f_W(t) = \int f_{WH}(t,u)\ du = \int f_{W|H}(t|u) f_H(u)\ du$$

we have

$$f_{H|W}(u|t) = \frac{f_{W|H}(t|u) f_H(u)}{\int f_{W|H}(t|u) f_H(u)\ du} \text{ and } E[H|W = t] = \frac{\int u f_{W|H}(t|u) f_H(u)\ du}{\int f_{W|H}(t|u) f_H(u)\ du}$$

Example 16.1.2 A Bayesian reversal

Suppose H ~ exponential (λ) and the X_i are conditionally iid, exponential (u), given $H = u$. A sample of size n is taken. Put $W = (X_1, X_2, \cdots, X_n)$, and $t^* = t_1 + t_2 + \cdots + t_n$. Determine the best mean-square estimate of H, given $W = t$.

Solution

$$f_{X|H}(t_i|u) = ue^{-ut_i} \text{ so that } f_{W|H}(t|u) = \prod_{i=1}^{n} ue^{-ut_i} = u^n e^{-ut^*}$$

Hence

$$E[H|W = t] = \int u f_{H|W}(u|t)\, du = \frac{\int_0^\infty u^{n+1} e^{-ut^*} \lambda e^{-\lambda u}\, du}{\int_0^\infty u^n e^{-ut^*} \lambda e^{-\lambda u}\, du}$$

$$= \frac{\int_0^\infty u^{n+1} e^{-(\lambda+t^*)u}\, du}{\int_0^\infty u^n e^{-(\lambda+t^*)u}\, du} = \frac{(n+1)!}{(\lambda+t^*)^{n+2}} \cdot \frac{(\lambda+t^*)^{n+1}}{n!} = \frac{n+1}{(\lambda+t^*)} \text{ where } t^* = \sum_{i=1}^{n} t_i$$

This page titled 16.1: Conditional Independence, Given a Random Vector is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

16.2: Elements of Markov Sequences

Elements of Markov Sequences

Markov sequences (Markov chains) are often studied at a very elementary level, utilizing algebraic tools such as matrix analysis. In this section, we show that the fundamental Markov property is an expression of conditional independence of “past” and “future," given the “present.” The essential Chapman-Kolmogorov equation is seen as a consequence of this conditional independence. In the usual time-homogeneous case with finite state space, the Chapman-Kolmogorov equation leads to the algebraic formulation that is widely studied at a variety of levels of mathematical sophistication. With the background laid, we only sketch some of the more common results. This should provide a probabilistic perspective for a more complete study of the algebraic analysis.

Markov sequences

We wish to model a system characterized by a sequence of *state*s taken on at discrete instants which we call *transition time*s. At each transition time, there is either a change to a new state or a renewal of the state immediately before the transition. Each state is maintained unchanged during the *period* or *stage* between transitions. At any transition time, the move to the next state is characterized by a conditional *transition probability* distribution. We suppose that the system is *memoryless*, in the sense that the transition probabilities are dependent upon the current state (and perhaps the period number), but not upon the manner in which that state was reached. *The past influences the future only through the present*. This is the *Markov property*, which we model in terms of conditional independence.

For period i, the state is represented by a value of a random variable X_i, whose value is one of the members of a set **E**, known as the *state space*. We consider only a finite state space and identify the states by integers from 1 to M. We thus have a sequence

$X_N = \{X_n : n \in N\}$, where $N = \{0, 1, 2, \cdots\}$

We view an observation of the system as a composite trial. Each ω yields a sequence of states $\{X_0(\omega), X_1(\omega), \cdots\}$ which is referred to as a *realization* of the sequence, or a *trajectory*. We suppose the system is evolving in time. At discrete instants of time $t_1, t_2, \cdots$ the system makes a transition from one state to the succeeding one (which may be the same).

Initial period:	$n = 0$, $t \in [0, t_1)$,	state is $X_0(\omega)$; at t_1 the transition is to $X_1(\omega)$
Period one:	$n = 1$, $t \in [t_1, t_2)$,	state is $X_1(\omega)$; at t_2 the transition is to $X_2(\omega)$
.....		
Period k:	$n = k$, $t \in [t_k, t_{k=1})$,	state is $X_k(\omega)$; at t_{k+1} move to $X_{k+1}(\omega)$
.....		

The parameter n indicates the *period* $t \in [t_n, t_{n+1})$. If the periods are of unit length, then $t_n = n$. At t_{n+1}, there is a *transition* from the state $X_n(\omega)$ to the state $X_{n+1}(\omega)$ for the next period. To simplify writing, we adopt the following convention:

$$U_n = (X_0, X_1, \cdots, X_n) \in E_n \; U_{m,n} = (X_m, \cdots, X_n) \text{ and } U^n = (X_n, X_{n+1}, \cdots) \in E^n$$

The random vector U_n is called the *past* at n of the sequence X_N and U^n is the *future* at n. In order to capture the notion that the system is without memory, so that the future is affected by the present, but not by how the present is reached, we utilize the notion of conditional independence, given a random vector, in the following

Definition

The sequence X_N is *Markov* iff

(M) $\{X_{n+1}, U_n\}$ ci $|X_n$ for all $n \geq 0$

Several conditions equivalent to the Markov condition (M) may be obtained with the aid of properties of conditional independence. We note first that (M) is equivalent to

$$P(X_{n+1} = k|X_n = j, U_{n-1} \in Q) = P(X_{n+1} = k|X_n = j) \text{ for each } n \geq 0, j, k \in E, \text{ and } Q \subset E^{n-1}$$

The state in the next period is conditioned by the past only through the present state, and not by the manner in which the present state is reached. The statistics of the process are determined by the *initial state probabilities* and the *transition probabilities*

$$P(X_{n+1} = k|X_n = j) \ \forall j, k \in E, n \geq 0$$

The following examples exhibit a pattern which implies the Markov condition and which can be exploited to obtain the transition probabilities.

Example 16.2.1 One-dimensional random walk

An object starts at a given initial position. At discrete instants $t_1, t_2, \cdots$ the object moves a random distance along a line. The various moves are independent of each other. Let

Y_0 be the initial position
Y_k be the amount the object moves at time $t = t_k$ $\{Y_k : 1 \leq k\}$ iid
$X_n = \sum_{k=0}^{n} Y_k$ be the position after n moves

We note that $X_{n+1} = g(X_n, Y_{n+1})$. Since the position after the transition at t_{n+1} is affected by the past only by the value of the position X_n and not by the sequence of positions which led to this position, it is reasonable to suppose that the process X_N is Markov. We verify this below.

Example 16.2.2 A class of branching processes

Each member of a population is able to reproduce. For simplicity, we suppose that at certain discrete instants the entire next generation is produced. Some mechanism limits each generation to a maximum population of M members. Let

Z_{in} be the number propagated by the ith member of the nth generation.
$Z_{in} = 0$ indicates death and no offspring, $Z_{in} = k$ indicates a net of k propagated by the ith member (either death and k offspring or survival and $k - 1$ offspring).

The population in generation $n + 1$ is given by

$$X_{n+1} = \min \{M, \sum_{i=1}^{X_n} Z_{in}\}$$

We suppose the class $\{Z_{in} : 1 \leq i \leq M, 0 \leq n\}$ is iid. Let $Y_{n+1} = (Z_{in}, Z_{2n}, \cdots, Z_{Mn})$. Thne $\{Y_{n+1}, U_n\}$ is independent. It seems reasonable to suppose the sequence X_N is Markov.

Example 16.2.3 An inventory problem

A certain item is stocked according to an (m, M) inventory policy, as follows:

- If stock at the end of a period is less than m, order up to M.
- If stock at the end of a period is m or greater, do not order.

Let X_0 be the initial stock, and X_n be the stock at the end of the nth period (before restocking), and let D_n be the demand during the nth period. Then for $n \geq 0$,

$$X_{n+1} = \begin{cases} \max \{M - D_{n+1}, 0\} & \text{if } 0 \leq X_n < m \\ \max \{X_n - D_{n+1}, 0\} & \text{if } m \leq X_n \end{cases} = g(X_n, D_{n+1})$$

If we suppose $\{D_n : 1 \le n\}$ is independent, then $\{D_{n=1}, U_n\}$ is independent for each $n \ge 0$, and the Markov condition seems to be indicated.

Remark. In this case, the actual transition takes place throughout the period. However, for purposes of analysis, we examine the state only at the end of the period (before restocking). Thus, the transitions are dispersed in time, but the observations are at discrete instants.

Example 16.2.4 Remaining lifetime

A piece of equipment has a lifetime which is an integral number of units of time. When a unit fails, it is replaced immediately with another unit of the same type. Suppose

- X_n is the remaining lifetime of the unit in service at time n
- Y_{n+1} is the lifetime of the unit installed at time n, with $\{Y_n : 1 \le n\}$ iid

$$\text{Then } X_{n+1} = \begin{cases} X_n - 1 & \text{if } X_n \ge 1 \\ Y_{n+1} - 1 & \text{if } X_n = 0 \end{cases} = g(X_n, Y_{n+1})$$

Remark. Each of these four examples exhibits the pattern

$\{X_0, Y_n : 1 \le n\}$ is independent
$X_{n+1} = g_{n+1}(X_n, Y_{n+1})$, $n \ge 0$

We now verify the Markov condition and obtain a method for determining the transition probabilities.

A pattern yielding Markov sequences

Suppose $\{Y_n : 0 \le n\}$ is independent (call these the *driving random variables*). Set

$$X_0 = g_0(Y_0) \text{ and } X_{n+1} = g_{n+1}(X_n, Y_{n+1}) \ \forall n \ge 0$$

Then

X_N is Markov
$P(X_{n+1} \in Q | X_n = u) = P[g_{n+1}(u, Y_{n+1}) \in Q]$ for all n, u, and any Borel set Q.

VERIFICATION

It is apparent that if $Y_0, Y_1, \cdots, Y_n$ are known, then U_n is known. Thus $U_n = h_n(Y_0, Y_1, \cdots, Y_n)$, which ensures each pair $\{Y_{n+1}, U_n\}$ is independent. By property (CI13), with $X = Y_{n+1}$, $Y = X_n$, and $Z = U_{n-1}$, we have

$$\{Y_{n+1}, U_{n-1}\} \text{ ci} | X_n$$

Since $X_{n+1} = g_{n+1}(Y_{n+1}, X_n)$ and $U_n = h_n(X_n, U_{n-1})$, property (CI9) ensures

$$\{X_{n+1}, U_n\} \text{ ci} | X_n \ \forall n \ge 0$$

which is the Markov property.

$P(X_{n+1} \in Q | X_n = u) = E\{I_Q[g_{n+1}(X_n, Y_{n+1})] | X_n = u\}$ a.s. $= E\{I_Q[g_{n+1}(u, Y_{n+1})]\}$ a.s. $[P_X]$ by (CE10b)
$= P[g_{n+1}(u, Y_{n+1}) \in Q]$ by (E1a)

— □

The application of this proposition, below, to the previous examples shows that the transition probabilities are invariant with n. This case is important enough to warrant separate classification.

Definition

If $P(X_{n+1} \in Q | X_n = u)$ is invariant with n, for all Borel sets Q, all $u \in E$, the Markov process X_N is said to be *homogeneous*.

As a matter of fact, this is the only case usually treated in elementary texts. In this regard, we note the following special case of the proposition above.

Homogenous Markov sequences

If $\{Y_n : 1 \leq n\}$ is iid and $g_{n+1} = g$ for all n, then the process is a homogeneous Markov process, and

$$P(X_{n+1} \in Q | X_n = u) = P[g(u, Y_{n+1}) \in Q], \text{ invariant with } n$$

— □

Remark.

In the homogeneous case, the transition probabilities are invariant with n. In this case, we write

$$P(X_{n+1} = j | X_n = i) = p(i, j) \text{ or } p_{ij} \text{ (invariant with } n)$$

These are called the *(one-step) transition probabilities*.

The transition probabilities may be arranged in a matrix **P** called the *transition probability matrix*, usually referred to as the *transition matrix*,

$$\mathbf{P} = [p(i, j)]$$

The element $p(i, j)$ on row i and column j is the probability $P(X_{n+1} = j | X_n = i)$. Thus, the elements on the ith row constitute the conditional distribution for X_{n+1}, given $X_n = i$. The transition matrix thus has the property that *each row sums to one*. Such a matrix is called a *stochastic matrix*. We return to the examples. From the propositions on transition probabilities, it is apparent that each is Markov. Since the function g is the same for all n and the driving random variables corresponding to the Y_i form an iid class, the sequences must be homogeneous. We may utilize part (b) of the propositions to obtain the one-step transition probabilities.

Example 16.2.5 Random walk continued

$g_n(u, Y_{n+1}) = u + Y_{n+1}$. so that g_n is invariant with n. Since $\{Y_n : 1 \leq n\}$ is iid,

$$P(X_{n+1} = k | X_n = j) = P(j + Y = k) = P(Y = k - j) = p_{k-j} \text{ where } p_k = P(Y = k)$$

Example 16.2.6 Branching process continued

$g(j, Y_{n+1}) = \min\{M, \sum_{i=1}^{j} Z_{in}\}$ and **E** = $\{0, 1, \cdots, M\}$. If $\{Z_{in} : 1 \leq i \leq M\}$ is iid, then

$$W_{jn} = \sum_{i=1}^{j} Z_{in} \text{ ensures } \{W_{jn} : 1 \leq n\} \text{ is iid for each } j \in \mathbf{E}$$

We thus have

$$P(X_{n+1} = k | X_n = j) = \begin{cases} P(W_{jn} = k) & \text{for } 0 \leq k < M \\ P(W_{jn} \geq M) & \text{for } k \geq M \end{cases} \quad 0 \leq j \leq M$$

With the aid of moment generating functions, one may determine distributions for

$$W_1 = Z_1, W_2 = Z_1 + Z_2, \cdots, W_M = Z_1 + \cdots + Z_M$$

These calculations are implemented in an m-procedure called *branchp*. We simply need the distribution for the iid Z_{in}.

```
% file branchp.m
% Calculates transition matrix for a simple branching
% process with specified maximum population.
disp('Do not forget zero probabilities for missing values of Z')
PZ = input('Enter PROBABILITIES for individuals  ');
M  = input('Enter maximum allowable population  ');
mz = length(PZ) - 1;
EZ = dot(0:mz,PZ);
disp(['The average individual propagation is ',num2str(EZ),])
P  = zeros(M+1,M+1);
Z  = zeros(M,M*mz+1);
k  = 0:M*mz;
a  = min(M,k);
z  = 1;
P(1,1) = 1;
for i = 1:M                 % Operation similar to genD
  z = conv(PZ,z);
  Z(i,1:i*mz+1) = z;
  [t,p] = csort(a,Z(i,:));
  P(i+1,:) = p;
end
disp('The transition matrix is P')
disp('To study the evolution of the process, call for branchdbn')

PZ = 0.01*[15 45 25 10 5];    % Probability distribution for individuals
branchp                       % Call for procedure
Do not forget zero probabilities for missing values of Z
Enter PROBABILITIES for individuals  PZ
Enter maximum allowable population  10
The average individual propagation is 1.45
The transition matrix is P
To study the evolution of the process, call for branchdbn
disp(P)                       % Optional display of generated P
  Columns 1 through 7
    1.0000         0         0         0         0         0         0
    0.1500    0.4500    0.2500    0.1000    0.0500         0         0
    0.0225    0.1350    0.2775    0.2550    0.1675    0.0950    0.0350
    0.0034    0.0304    0.1080    0.1991    0.2239    0.1879    0.1293
    0.0005    0.0061    0.0307    0.0864    0.1534    0.1910    0.1852
    0.0001    0.0011    0.0075    0.0284    0.0702    0.1227    0.1623
    0.0000    0.0002    0.0017    0.0079    0.0253    0.0579    0.1003
    0.0000    0.0000    0.0003    0.0020    0.0078    0.0222    0.0483
    0.0000    0.0000    0.0001    0.0005    0.0021    0.0074    0.0194
    0.0000    0.0000    0.0000    0.0001    0.0005    0.0022    0.0068
    0.0000    0.0000    0.0000    0.0000    0.0001    0.0006    0.0022
  Columns 8 through 11
         0         0         0         0
         0         0         0         0
    0.0100    0.0025         0         0
    0.0705    0.0315    0.0119    0.0043
    0.1481    0.0987    0.0559    0.0440
    0.1730    0.1545    0.1179    0.1625
    0.1381    0.1574    0.1528    0.3585
    0.0832    0.1179    0.1412    0.5771
    0.0406    0.0698    0.1010    0.7591
```

```
    0.0169    0.0345    0.0590    0.8799
    0.0062    0.0147    0.0294    0.9468
```

Note that $p(0,0) = 1$. If the population ever reaches zero, it is extinct and no more births can occur. Also, if the maximum population (10 in this case) is reached, there is a high probability of returning to that value and very small probability of becoming extinct (reaching zero state).

Example 16.2.7 Inventory problem (continued)

In this case,

$$g(j, D_{n+1}) = \begin{cases} \max\{M - D_{n+1}, 0\} & \text{for } 0 \le j < m \\ \max\{j - D_{n+1}, 0\} & \text{for } m \le j \le M \end{cases}$$

Numerical example

$$m = 1 \; M = 3 \; D_n \text{ is Poisson (1)}$$

To simplify writing, use D for D_n. Because of the invariance with n, set

$$P(X_{n+1} = k | X_n = j) = p(j,k) = P(g(j, D_{n+1}) = k]$$

The various cases yield

$g(0, D) = \max\{3 - D, 0\}$

$g(0, D) = 0$ iff $D \ge 3$ imples $p(0,0) = P(D \ge 3)$
$g(0, D) = 1$ iff $D = 2$ imples $p(0,1) = P(D = 2)$
$g(0, D) = 2$ iff $D = 1$ imples $p(0,2) = P(D = 1)$
$g(0, D) = 3$ iff $D = 0$ imples $p(0,3) = P(D = 0)$

$g(1, D) = \max\{1 - D, 0\}$

$g(1, D) = 0$ iff $D \ge 1$ imples $p(1,0) = P(D \ge 1)$
$g(1, D) = 1$ iff $D = 0$ imples $p(1,1) = P(D = 0)$
$g(1, D) = 2, 3$ is impossible

$g(2, D) = \max\{2 - D, 0\}$

$g(2, D) = 0$ iff $D \ge 2$ imples $p(2,0) = P(D \ge 2)$
$g(2, D) = 1$ iff $D = 1$ imples $p(2,1) = P(D = 1)$
$g(2, D) = 2$ iff $D = 0$ imples $p(2,2) = P(D = 0)$
$g(2, D) = 3$ is impossible

$g(3, D) = \max\{3 - D, 0\} = g(0, D)$ so that $p(3,k) = p(0,k)$

The various probabilities for D may be obtained from a table (or may be calculated easily with cpoisson) to give the transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0.0803 & 0.1839 & 0.3679 & 0.3679 \\ 0.6321 & 0.3679 & 0 & 0 \\ 0.2642 & 0.3679 & 0.3679 & 0 \\ 0.0803 & 0.1839 & 0.3679 & 0.3679 \end{bmatrix}$$

The calculations are carried out “by hand” in this case, to exhibit the nature of the calculations. This is a standard problem in inventory theory, involving costs and rewards. An m-procedure *inventory1* has been written to implement the function g.

```
% file inventory1.m
% Version of 1/27/97
% Data for transition probability calculations
% for (m,M) inventory policy
M  = input('Enter value M of maximum stock  ');
m  = input('Enter value m of reorder point  ');
Y  = input('Enter row vector of demand values  ');
PY = input('Enter demand probabilities  ');
states = 0:M;
ms = length(states);
my = length(Y);
% Calculations for determining P
[y,s] = meshgrid(Y,states);
T  =  max(0,M-y).*(s < m) + max(0,s-y).*(s >= m);
P  = zeros(ms,ms);
for i = 1:ms
   [a,b] = meshgrid(T(i,:),states);
   P(i,:) = PY*(a==b)';
end
P
```

We consider the case $M = 5$, the reorder point $m = 3$. and demand is Poisson (3). We approximate the Poisson distribution with values up to 20.

```
inventory1
Enter value M of maximum stock  5             % Maximum stock
Enter value m of reorder point  3             % Reorder point
Enter row vector of demand values  0:20       % Truncated set of demand values
Enter demand probabilities  ipoisson(3,0:20)  % Demand probabilities
P =
    0.1847    0.1680    0.2240    0.2240    0.1494    0.0498
    0.1847    0.1680    0.2240    0.2240    0.1494    0.0498
    0.1847    0.1680    0.2240    0.2240    0.1494    0.0498
    0.5768    0.2240    0.1494    0.0498         0         0
    0.3528    0.2240    0.2240    0.1494    0.0498         0
    0.1847    0.1680    0.2240    0.2240    0.1494    0.0498
```

Example 16.2.8 Remaining lifetime (continued)

$g(0, Y) = Y - 1$, so that $p(0, k) = P(Y - 1 = k) = P(Y = k + 1)$

$g(j, Y) = j - 1$ for $j \geq 1$, so that $p(j, k) = \delta_{j-1,k}$ for $j \geq 1$

The resulting transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} p_1 & p_2 & p_3 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ \cdots & & & \cdots \\ \cdots & & & \\ \cdots & & & \cdots \end{bmatrix}$$

The matrix is an infinite matrix, unless Y is simple. If the range of Y is $\{1, 2, \cdots, M\}$ then the state space **E** is $\{0, 1, \cdots, M - 1\}$.

Various properties of conditional independence, particularly (CI9), (CI10), and (CI12), may be used to establish the following. The immediate future X_{n+1} may be replaced by any *finite future* $U_{n,n+p}$ and the present X_n may be replaced by any *extended present* $U_{m,n}$. Some results of abstract measure theory show that the finite future $U_{n,n+p}$ may be replaced by the *entire future* U^n. Thus, we may assert

Extended Markov property

X_N is Markov iff

$$\text{(M*) } \{U^n, U_m\} \text{ ci} | U_{m,n} \ \forall 0 \leq m \leq n$$

— □

The Chapman-Kolmogorov equation and the transition matrix

As a special case of the extended Markov property, we have

$$\{U^{n+k}, U_n\} \text{ ci} | X_{n+k} \text{ for all } n \geq 0, k \geq 1$$

Setting $g(U^{n+k}, X_{n+k}) = X_{n+k+m}$ and $h(U_n, X_{n+k}) = X_n$ in (CI9), we get

$$\{X_{n+k+m}, X_n\} \text{ ci} | X_{n+k} \text{ for all } n \geq 0, k, m \geq 1$$

This is the *Chapman-Kolmogorov equation*, which plays a central role in the study of Markov sequences. For a discrete state space **E**, with

$P(X_n = j | X_m = i) = p_{m,n}(i, j)$

this equation takes the form

$$(CK') \ p_{m,q}(i,k) = \sum_{j \in E} p_{m,n}(i,j) p_{n,q}(j,k) \ 0 \leq m < n < q$$

To see that this is so, consider

$$P(X_q = k | X_m = i) = E[I_{\{k\}}(X_q) | X_m = i] = E\{E[I_{\{k\}}(X_q) | X_n] | X_m = i\}$$

$$= \sum_j E[I_{\{k\}}(X_q) | X_n = j] p_{m,n}(i,j) = \sum_j p_{n,q}(j,k) p_{m,n}(i,j)$$

Homogeneous case

For this case, we may put (CK') in a useful matrix form. The conditional probabilities p^m of the form

$$p^m(i,k) = P(X_{n+m} = k | X_n = i) \text{ invariant in } n$$

are known as the *m-step transition probabilities*. The Chapman-Kolmogorov equation in this case becomes

$$(CK'') \ p^{m+n}(i,k) = \sum_{j \in E} p^m(i,j) p^n(j,k) \ \forall i, j \in \mathbf{E}$$

In terms of the *m-step transition matrix* $\mathbf{P}^{(m)} = [p^m(i,k)]$, this set of sums is equivalent to the matrix product

$$(CK'') \ \mathbf{P}^{(m+n)} = \mathbf{P}^{(m)} \mathbf{P}^{(n)}$$

Now

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1)} \mathbf{P}^{(1)} = \mathbf{PP} = \mathbf{P}^2, \ \mathbf{P}^{(3)} = \mathbf{P}^{(2)} \mathbf{P}^{(1)} = \mathbf{P}^3, \text{ etc.}$$

A simple inductive argument based on (CK'') establishes

The product rule for transition matrices

The *m*-step probability matrix $\mathbf{P}^{(m)} = \mathbf{P}^m$, the mth power of the transition matrix **P**

— □

Example 16.2.9 The inventory problem (continued)

For the inventory problem in Example 16.2.7, the three-step transition probability matrix $\mathbf{P}^{(3)}$ is obtained by raising **P** to the third power to get

$$\mathbf{P}^{(3)} = \mathbf{P}^3 = \begin{bmatrix} 0.2930 & 0.2917 & 0.2629 & 0.1524 \\ 0.2619 & 0.2730 & 0.2753 & 0.1898 \\ 0.2993 & 0.2854 & 0.2504 & 0.1649 \\ 0.2930 & 0.2917 & 0.2629 & 0.1524 \end{bmatrix}$$

— □

We consider next the state probabilities for the various stages. That is, we examine the distributions for the various X_n, letting $p_k(n) = P(X_n = k)$ for each $k \in \mathbf{E}$. To simplify writing, we consider a finite state space $\mathbf{E} = \{1, \cdots, M\}$, We use $\pi(n)$ for the *row*matrix

$$\pi(n) = [p_1(n) p_2(n) \cdots p_M(n)]$$

As a consequence of the product rule, we have

Probability distributions for any period

For a homogeneous Markov sequence, the distribution for any X_n is determined by the initial distribution (i.e., for X_0) and the transition probability matrix **P**

VERIFICATION

Suppose the homogeneous sequence X_N has finite state-space $\mathbf{E} = \{1, 2, \cdots, M\}$. For any $n \geq 0$, let $p_j(n) P(X_n = j)$ for each $j \in \mathbf{E}$. Put

$$\pi(n) = [p_1(n) p_2(n) \cdots p_M(n)]$$

Then

$\pi(0) =$ the initial probability distribution
$\pi(1) = \pi(0)\mathbf{P}$
.....
$\pi(n) = \pi(n-1)\mathbf{P} = \pi(0)\ \mathbf{P}^{(n)} = \pi(0)\ \mathbf{P}^n =$ the nth-period distribution

The last expression is an immediate consequence of the product rule.

Example 16.2.10 Inventory problem (continued)

In the inventory system for Examples 3, 7 and 9, suppose the initial stock is $M = 3$. This means that

$$\pi(0) = [0\ 0\ 0\ 1]$$

The product of $\pi(0)$ and P^3 is the fourth row of P^3, so that the distribution for X_3 is

$$\pi(3) = [p_0(3)\ \ p_1(3)\ \ p_2(3)\ \ p_3(3)] = [0.2930\ \ 0.2917\ \ 0.2629\ \ 0.1524]$$

Thus, given a stock of $M = 3$ at startup, the probability is 0.2917 that $X_3 = 1$. This is the probability of one unit in stock at the end of period number three.

Remarks

- A similar treatment shows that for the nonhomogeneous case the distribution at any stage is determined by the initial distribution and the class of one-step transition matrices. In the nonhomogeneous case, transition probabilities $p_{n,n+1}(i,j)$ depend on the stage n.
- A discrete-parameter Markov process, or Markov sequence, is characterized by the fact that each member X_{n+1} of the sequence is conditioned by the value of the previous member of the sequence. This one-step stochastic linkage has made it customary to refer to a Markov sequence as a *Markov chain*. In the discrete-parameter Markov case, we use the terms *process, sequence,* or *chain* interchangeably.

The transition diagram and the transition matrix

The previous examples suggest that a Markov chain is a dynamic system, evolving in time. On the other hand, the stochastic behavior of a homogeneous chain is determined completely by the probability distribution for the initial state and the one-step transition probabilities $p(i,j)$ as presented in the transition matrix **P**. The time-invariant transition matrix may convey a static impression of the system. However, a simple geometric representation, known as the *transition diagram*, makes it possible to link the unchanging structure, represented by the transition matrix, with the dynamics of the evolving system behavior.

Definition

A *transition diagram* for a homogeneous Markov chain is a linear graph with one node for each state and one directed edge for each possible one-step transition between states (nodes).

We ignore, as essentially impossible, any transition which has zero transition probability. Thus, the edges on the diagram correspond to positive one-step transition probabilities between the nodes connected. Since for some pair (i,j) of states, we may have $p(i,j) > 0$ but $p(j,i) = 0$ we may have a connecting edge between two nodes in one direction, but none in the other. *The system can be viewed as an object jumping from state to state (node to node) at the successive transition times*. As we follow the *trajectory* of this object, we achieve a sense of the evolution of the system.

Example 16.2.11 Transition diagram for inventory example

Consider, again, the transition matrix **P** for the inventory problem (rounded to three decimals).

$$\mathbf{P} = \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix}$$

Figure 16.2.1 shows the transition diagram for this system. At each node corresponding to one of the possible states, the *state value* is shown. In this example, the state value is one less than the state number. For convenience, we refer to the node for state $k+1$. which has state value k, as node k. If the state value is zero, there are four possibilities: remain in that condition with probability 0.080; move to node 1 with probability 0.184; move to node 2 with probability 0.368; or move to node 3 with probability 0.368. These are represented by the "self loop" and a directed edge to each of the nodes representing states. Each of these directed edges is marked with the (conditional) transition probability. On the other hand, probabilities of reaching state value 0 from each of the others is represented by directed edges into the node for state value 0. A similar situation holds for each other node. Note that the probabilities on edges leaving a node (including a self loop) must total to one, since these correspond to the transition probability distribution from that node. There is no directed edge from the node 2 to node 3, since the probability of a transition from value 2 to value 3 is zero. Similary, there is no directed edge from node 1 to either node 2 or node 3.

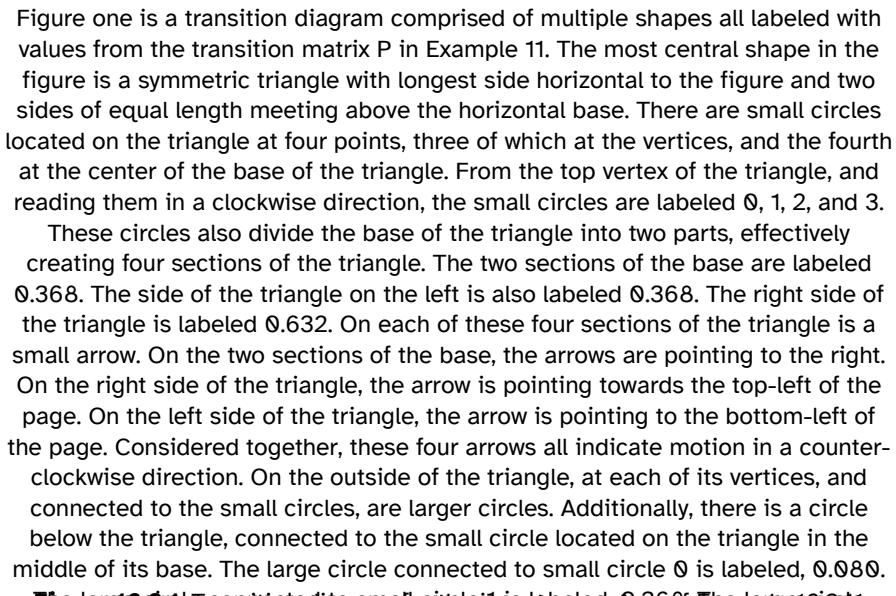


Figure 16.2.1. Transition diagram for the inventory system of Example 16.2.1

There is a one-one relation between the transition diagram and the transition matrix **P**. The transition diagram not only aids in visualizing the dynamic evolution of the chain, but also displays certain structural properties. Often a chain may be decomposed usefully into subchains. Questions of *communication*, and *recurrence* may be answered in terms of the transition diagram. Some subsets of states are essentially closed, in the sense that if the system arrives at any one state in the subset it can never reach a state outside the subset. Periodicities sometimes exist and can usually be easily seen using the diagram to show that periodicities cannot occur.

Classification of states

Many important characteristics of a Markov chain can be studied by considering the number of visits to an arbitrarily chosen, but fixed, state.

Definition

For a fixed state j, let

T_1 = the time (stage number) of the first visit to state j (after the initial period).
$F_k(i,j) = P(T_i = k | X_0 = i)$, the probability of reaching state j for the first time from state i in k steps.
$F(i,j) = P(T_i < \infty | X_0 = i) = \sum_{k=1}^{\infty} F_k(i,j)$, the probability of ever reaching state j from state i.

A number of important theorems may be developed for F_k and F, although we do not develop them in this treatment. We simply quote them as needed. An important classification of states is made in terms of F.

Definition

State j is said to be *transient* iff $F(j,j) < 1$,

and is said to be *recurrent* iff $F(j,j) = 1$.

Remark. If the state space **E** is infinite, recurrent states fall into one of two subclasses: positive or null. Only the positive case is common in the infinite case, and that is the only possible case for systems with finite state space.

Sometimes there is a regularity in the structure of a Markov sequence that results in periodicities.

Definition

For state j, let

$\delta = \text{greatest common denominator of } \{n : p^n(j,j) > 0\}$

If $\delta > 1$, then state j is *periodic* with *period* δ; otherwise, state j is *aperiodic*.

Usually if there are any self loops in the transition diagram (positive probabilities on the diagonal of the transition matrix **P**) the system is aperiodic. *Unless stated otherwise, we limit consideration to the aperiodic case.*

Definition

A state j is called *ergodic* iff it is positive, recurrent, and aperiodic.

It is called *absorbing* iff $F(j,j) = 1$.

A recurrent state is one to which the system eventually returns, hence is visited an infinity of times. If it is absorbing, then once it is reached it returns each step (i.e., never leaves).

An arrow notation is used to indicate important relations between states.

Definition

We say

State i reaches j, denoted $i \to j$, iff $p^n(i,j) > 0$ for some $n > 0$.
State i and j communicate, denoted $i \leftrightarrow j$ iff both i reaches j and j reaches i.

By including j reaches j in all cases, the relation $\leftrightarrow$ is an equivalence relation (i.e., is reflexive, transitive, and idempotent). With this relationship, we can define important classes.

Definition

A class of states is *communicating* iff every state in the class may be reached from every other state in the class (i.e. every pair communicates). A class is *closed* if no state outside the class can be reached from within the class.

The following important conditions are intuitive and may be established rigorously:

$i \leftrightarrow j$ implies i is recurrent iff j is recurrent
$i \to j$ and i recurrent implies $i \leftrightarrow j$
$i \to j$ and i recurrent implies j recurrent

Limit theorems for finite state space sequences

The following propositions may be established for Markov sequences with finite state space:

- There are no null states, and not all states are transient.
- If a class of states is irreducible (i.e.,has no proper closed subsets), then
  - All states are recurrent
  - All states are aperiodic *or* all are periodic with the same period.
  - If a class *C* is closed, irreducible, and *i* is a transient state (necessarily not in C).
    then $F(i,j) = F(i,k)$ for all $j,k \in C$

A limit theorem

If the states in a Markov chain are ergodic (i.e., positive, recurrent, aperiodic), then

$$\lim_n p^n(i,j) = \pi_j > 0 \; \sum_{j=1}^M \pi_j = 1 \; \pi_j = \sum_{i=1}^M \pi_i p(i,j)$$

If, as above, we let

$$\pi(n) = [p_1(n) \; p_2(n) \cdots p_M(n)] \text{ so that } \pi(n) = \pi(0) \; \mathbf{P}^n$$

the result above may be written

$\pi(n) = \pi(0) \; \mathbf{P}^n \to \pi(0) \; \mathbf{P}_0$

where

$$\mathbf{P}_0 = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \pi_1 & \pi_2 & \cdots & \pi_m \\ \cdots & \cdots & \cdots & \cdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{bmatrix}$$

Each row of $\mathbf{P}_0 = \lim_n \mathbf{P}^n$ is the long run distribution $\pi = \lim_n \pi(n)$.

Definition

A distribution is **stationary** iff

$$\pi = \pi \mathbf{P}$$

The result above may be stated by saying that the long-run distribution is the stationary distribution. A generating function analysis shows the convergence is exponential in the following sense

$$|\mathbf{P}^n - \mathbf{P}_0| \leq \alpha |\lambda|^n$$

where $|\lambda|$ is the largest absolute value of the eigenvalues for **P** other than $\lambda = 1$.

Example 16.2.12 The long run distribution for the inventory example

We use MATLAB to check the eigenvalues for the transition probability **P** and to obtain increasing powers of **P**. The convergence process is readily evident.

```
P =
    0.0803    0.1839    0.3679    0.3679
    0.6321    0.3679         0         0
    0.2642    0.3679    0.3679         0
    0.0803    0.1839    0.3679    0.3679
E = abs(eig(P))
E =
   1.0000
   0.2602
   0.2602
   0.0000
   format long
N = E(2).^[4 8 12]
N = 0.00458242348096   0.00002099860496   0.00000009622450
>> P4 = P^4
```

```
P4 =
   0.28958568915950   0.28593792666752   0.26059678211310   0.16387960205989
   0.28156644866011   0.28479107531968   0.26746979455342   0.16617268146679
   0.28385952806702   0.28250048636032   0.26288737107246   0.17075261450021
   0.28958568915950   0.28593792666752   0.26059678211310   0.16387960205989

>> P8 = P^8
P8 =
   0.28580046500309   0.28471421248816   0.26315895715219   0.16632636535655
   0.28577030590344   0.28469190218618   0.26316681807503   0.16637097383535
   0.28581491438224   0.28471028095839   0.26314057837998   0.16633422627939
   0.28580046500309   0.28471421248816   0.26315895715219   0.16632636535655

>> P12 = P^12
P12 =
   0.28579560683438   0.28470680858266   0.26315641543927   0.16634116914369
   0.28579574073314   0.28470680714781   0.26315628010643   0.16634117201261
   0.28579574360207   0.28470687626748   0.26315634631961   0.16634103381085
   0.28579560683438   0.28470680858266   0.26315641543927   0.16634116914369
>> error4 = max(max(abs(P^16 - P4)))    % Use P^16 for P_0
error4 =  0.00441148012334              % Compare with 0.0045824...
>> error8 = max(max(abs(P^16 - P8)))
error8 = 2.984007206519035e-05          % Compare with  0.00002099
>> error12 = max(max(abs(P^16 - P12)))
error12 = 1.005660185959822e-07         % Compare with 0.00000009622450
```

The convergence process is clear, and the agreement with the error is close to the predicted. We have not determined the factor a, and we have approximated the long run matrix P_0 with P^{16}. This exhibits a *practical criterion* for sufficient convergence. If the rows of P^n agree within acceptable precision, then n is sufficiently large. For example, if we consider agreement to four decimal places sufficient, then

```
P10 = P^10
P10 =
    0.2858    0.2847    0.2632    0.1663
    0.2858    0.2847    0.2632    0.1663
    0.2858    0.2847    0.2632    0.1663
    0.2858    0.2847    0.2632    0.1663
```

shows that $n = 10$ is quite sufficient.

Simulation of finite homogeneous Markov sequences

In the section, "The Quantile Function", the quantile function is used with a random number generator to obtain a simple random sample from a given population distribution. In this section, we adapt that procedure to the problem of simulating a trajectory for a homogeneous Markov sequences with finite state space.

Elements and terminology

1. **States and state numbers**. We suppose there are *m* states, usually carrying a numerical value. For purposes of analysis and simulation, we number the states 1 through *m*. Computation is carried out with state numbers; if desired, these can be translated into the actual state values after computation is completed.
2. **Stages, transitions, period numbers, trajectories and time**. We use the term stage and period interchangeably. It is customary to number the periods or stages beginning with zero for the initial stage. The *period number* is the number of transitions to reach that stage from the initial one. Zero transitions are required to reach the original stage (period zero), one transition to reach the next (period one), two transitions to reach period two, etc. We call the sequence of states encountered

as the system evolves a *trajectory* or a *chain*. The terms "sample path" or "realization of the process" are also used in the literature. Now if the periods are of equal time length, the number of transitions is a measure of the elapsed time since the chain originated. We find it convenient to refer to *time* in this fashion. At time k the chain has reached the period numbered k. The trajectory is $k + 1$ stages long, so time or period number is one less than the number of stages.

3. **The transition matrix and the transition distributions**. For each state, there is a conditional transition probability distribution for the next state. These are arranged in a *transition matrix*. The ith row consists of the transition distribution for selecting the next-period state when the current state number is i. The transition matrix P thus has nonnegative elements, with each row summing to one. Such a matrix is known as a *stochastic matrix*.

The fundamental simulation strategy

1. A fundamental strategy for sampling from a given population distribution is developed in the unit on the Quantile Function. If Q is the quantile function for the population distribution and U is a random variable distributed uniformly on the interval [0, 1], then $X = Q(U)$ has the desired distribution. To obtain a sample from the uniform distribution use a random number generator. This sample is "transformed" by the quantile function into a sample from the desired distribution.
2. For a homogeneous chain, if we are in state k, we have a distribution for selecting the next state. If we use the quantile function for that distribution and a number produced by a random number generator, we make a selection of the next state based on that distribution. A succession of these choices, with the selection of the next state made in each case from the distribution for the current state, constitutes a valid simulation of a trajectory.

Arrival times and recurrence times

The basic simulation produces one or more trajectories of a specified length. Sometimes we are interested in continuing until first arrival at (or visit to) a specific target state or any one of a set of target states. The time (in transitions) to reach a target state is one less than the number of stages in the trajectory which begins with the initial state and ends with the target state reached.

- If the initial state is not in the target set, we speak of the *arrival time*.
- If the initial state is in the target set, the arrival time would be zero. In this case, we do not stop at zero but continue until the next visit to a target state (possibly the same as the initial state). We call the number of transitions in this case the *recurrence time*.
- In some instances, it may be desirable to know the time to complete visits to a prescribed number of the target states. Again there is a choice of treatment in the case the initial set is in the target set.

Data files

For use of MATLAB in simulation, we find it convenient to organize the appropriate data in an m-file.

- In every case, we need the transition matrix P. Its size indicates the number of states (say by the length of any row or column).
- If the states are to have values other than the state numbers, these may be included in the data file, although they may be added later, in response to a prompt.
- If long trajectories are to be produced, it may be desirable to determine the fraction of times each state is realized. A comparison with the long-run probabilities for the chain may be of interest. In this case, the data file may contain the long-run probability distribution. Usually, this is obtained by taking one row of a sufficiently large power of the transition matrix. This operation may be performed after the data file is called for but before the simulation procedure begins.

An example data file used to illustrate the various procedures is shown below. These data were generated artificially and have no obvious interpretations in terms of a specific systems to be modeled. However, they are sufficiently complex to provide nontrivial illustrations of the simulation procedures.

```
% file markovp1.m
% Artificial data for a Markov chain, used to
% illustrate the operation of the simulation procedures.
P = [0.050 0.011 0.155 0.155 0.213 0.087 0.119 0.190 0.008 0.012
     0.103 0.131 0.002 0.075 0.013 0.081 0.134 0.115 0.181 0.165
     0.103 0.018 0.128 0.081 0.137 0.180 0.149 0.051 0.009 0.144
     0.051 0.098 0.118 0.154 0.057 0.039 0.153 0.112 0.117 0.101
     0.016 0.143 0.200 0.062 0.099 0.175 0.108 0.054 0.062 0.081
```

```
     0.029 0.085 0.156 0.158 0.011 0.156 0.088 0.090 0.055 0.172
     0.110 0.059 0.020 0.212 0.016 0.113 0.086 0.062 0.204 0.118
     0.084 0.171 0.009 0.138 0.140 0.150 0.023 0.003 0.125 0.157
     0.105 0.123 0.121 0.167 0.149 0.040 0.051 0.059 0.086 0.099
     0.192 0.093 0.191 0.061 0.094 0.123 0.106 0.065 0.040 0.035];
states = 10:3:37;
PI = [0.0849 0.0905 0.1125 0.1268 0.0883 0.1141 ...
      0.1049 0.0806 0.0881 0.1093];         % Long-run distribution
```

The largest absolute value of the eigenvalues (other than one) is 0.1716. Since $0.1716^{16} \approx 5.6 \cdot 10^{-13}$, we take any row of P^{16} as the long-run probabilities. These are included in the matrix PI in the m-file, above. The examples for the various procedures below use this set of artificial data, since the purpose is to illustrate the operation of the procedures.

The setup and the generating m-procedures

The m-procedure *chainset* sets up for simulation of Markov chains. It prompts for input of the transition matrix *P*, the states (if different from the state numbers), the long-run distribution (if available), and the set of target states if it is desired to obtain arrival or recurrence times. The procedure determines the number of states from the size of *P* and calculates the information needed for the quantile function. It then prompts for a call for one of the generating procedures.

The m-procedure *mchain*, as do the other generating procedures below, assumes chainset has been run, so that commonly used data are available in appropriate form. The procedure prompts for the number of stages (length of the trajectory to be formed) and for the initial state. When the trajectory is produced, the various states in the trajectory and the fraction or relative frequency of each is displayed. If the long-run distribution has been supplied by chainset, this distribution is included for comparison. In the examples below, we reset the random number generator (set the "seed" to zero) for purposes of comparison. However, in practice, it may be desirable to make several runs without resetting the seed, to allow greater effective "randomness."

Example 16.2.13

```
markovp1                              % Call for data
chainset                              % Call for setup procedure
Enter the transition matrix  P
Enter the states if not 1:ms  states  % Enter the states
States are
     1    10
     2    13
     3    16
     4    19
     5    22
     6    25
     7    28
     8    31
     9    34
    10    37
Enter the long-run probabilities  PI  % Enter the long-run distribution
Enter the set of target states [16 22 25]   % Not used with mchain
Call for for appropriate chain generating procedure
rand('seed',0)
mchain                                % Call for generating procedure
Enter the number n of stages   10000  % Note the trajectory length
Enter the initial state  16
     State     Frac       P0          % Statistics on the trajectory
   10.0000    0.0812    0.0849
   13.0000    0.0952    0.0905
   16.0000    0.1106    0.1125
   19.0000    0.1226    0.1268
```

```
   22.0000    0.0880    0.0883
   25.0000    0.1180    0.1141
   28.0000    0.1034    0.1049
   31.0000    0.0814    0.0806
   34.0000    0.0849    0.0881
   37.0000    0.1147    0.1093
To view the first part of the trajectory of states, call for TR
disp(TR')
     0     1     2     3     4     5     6     7     8     9    10
    16    16    10    28    34    37    16    25    37    10    13
```

The fact that the fractions or *relative frequencies* approximate the long-run probabilities is an expression of a fundamental limit property of probability theory. This limit property, which requires somewhat sophisticated technique to establish, justifies a relative frequency interpretation of probability.

The procedure *arrival* assumes the setup provided by chainset, including a set E of target states. The procedure prompts for the number *r* of repetitions and the initial state. Then it produces r succesive trajectories, each starting with the prescribed initial state and ending on one of the target states. The arrival times vary from one run to the next. Various statistics are computed and displayed or made available. In the single-run case ($r = 1$), the trajectory may be displayed. An auxiliary procedure *plotdbn* may be used in the multirun case to plot the distribution of arrival times.

Example 16.2.14 Arrival time to a target set of states

```
rand('seed',0)
arrival                                % Assumes chainset has been run, as above
Enter the number of repetitions  1     % Single run case
The target state set is:
    16    22    25
Enter the initial state  34            % Specified initial state
 The arrival time is 6                 % Data on trajectory
The state reached is 16
To view the trajectory of states, call for TR
disp(TR')                              % Optional call to view trajectory
     0     1     2     3     4     5     6
    34    13    10    28    34    37    16
rand('seed',0)
arrival
Enter the number of repetitions  1000  % Call for 1000 repetitions
The target state set is:
    16    22    25
Enter the initial state  34            % Specified initial state
 The result of 1000 repetitions is:    % Run data (see optional calls below)
Term state  Rel Freq   Av time
   16.0000    0.3310    3.3021
   22.0000    0.3840    3.2448
   25.0000    0.2850    4.3895
The average arrival time is 3.59
The standard deviation is 3.207
The minimum arrival time is 1
The maximum arrival time is 23
To view the distribution of arrival times, call for dbn
To plot the arrival time distribution, call for plotdbn
plotdbn                              % See Figure 16.2.2
```

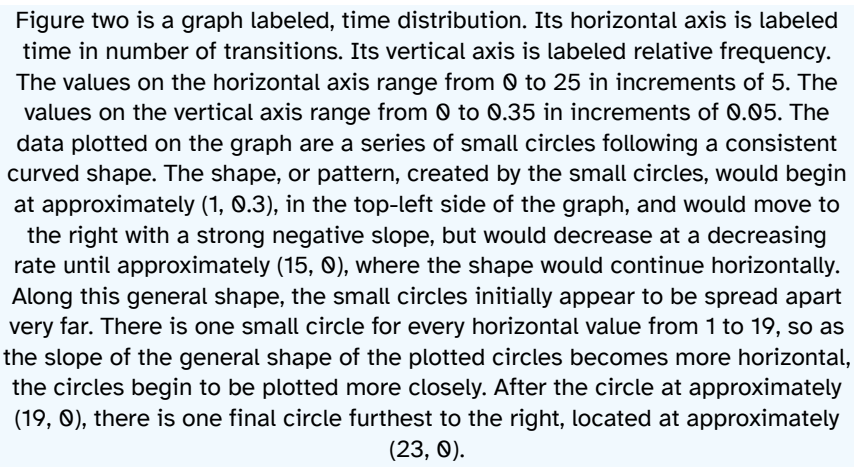


Figure 16.2.2. Time distribution for Example 16.2.14

It would be difficult to establish analytically estimates of arrival times. The simulation procedure gives a reasonable "feel" for these times and how they vary.

The procedure *recurrence* is similar to the procedure arrival. If the initial state is not in the target set, it behaves as does the procedure arrival and stops on the first visit to the target set. However, if the initial state is in the target set, the procedures are different. The procedure arrival stops with zero transitions, since it senses that it has "arrived." We are usually interested in having at least one transition– back to the same state or to another state in the target set. We call these times *recurrence times*.

Example 16.2.15

```
rand('seed',0)
recurrence
Enter the number of repititions  1
The target state set is:
     16     22     25
Enter the initial state  22
```

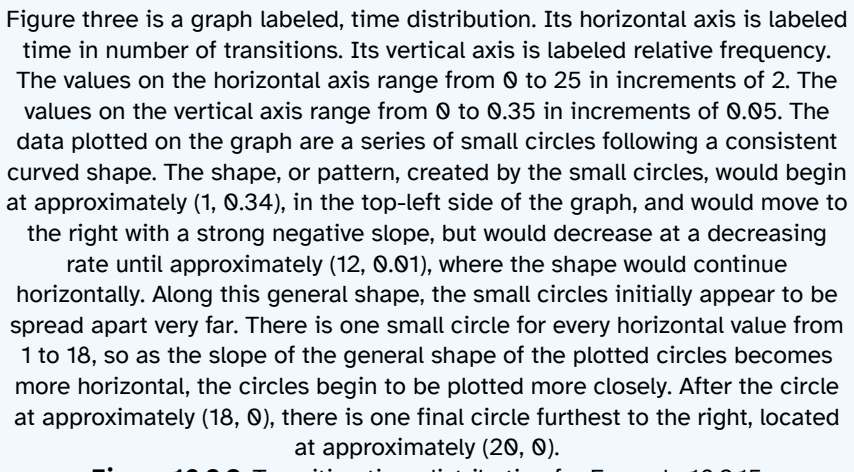


Figure 16.2.3. Transition time distribution for Example 16.2.15

```
The recurrence time is 1
The state reached is 16
To view the trajectory of state numbers, call for TR
```

```
disp(TR')    0     1
            22    16
recurrence
Enter the number of repititions  1000
The target state set is:
     16     22     25
Enter the initial state  25
The result of 1000 repetitions is:
Term state  Rel Freq   Av time
   16.0000    0.3680    2.8723
   22.0000    0.2120    4.6745
   25.0000    0.4200    3.1690
   The average recurrence time is 3.379
The standard deviation is 3.0902
The minimum recurrence time is 1
The maximum recurrence time is 20
To view the distribution of recurrence times, call for dbn
To plot the recurrence time distribution, call for plotdbn
% See Figure 16.2.3
```

The procedure *kvis* stops when a designated number k of states are visited. If k is greater than the number of target states, or if no k is designated, the procedure stops when all have been visited. For $k = 1$, the behavior is the same as arrival. However, that case is better handled by the procedure arrival, which provides more statistics on the results.

Example 16.2.16

```
rand('seed',0)
kvis                % Assumes chainset has been run
Enter the number of repetitions  1
The target state set is:
     16     22     25
Enter the number of target states to visit  2
Enter the initial state  34
The time for completion is 7
To view the trajectory of states, call for TR
disp(TR')
      0     1     2     3     4     5     6     7
     34    13    10    28    34    37    16    25
rand('seed',0)
kvis
Enter the number of repetitions  100
The target state set is:
     16     22     25
Enter the number of target states to visit    % Default-- visit all three
Enter the initial state  31
The average completion time is 17.57
The standard deviation is 8.783
The minimum completion time is 5
The maximum completion time is 42
To view a detailed count, call for D.
The first column shows the various completion times;
the second column shows the numbers of trials yielding those times
```

The first goal of this somewhat sketchy introduction to Markov processes is to provide a general setting which gives insight into the essential character and structure of such systems. The important case of homogenous chains is introduced in such a way that

their algebraic structure appears as a logical consequence of the Markov propertiy. The general theory is used to obtain some tools for formulating homogeneous chains in practical cases. Some MATLAB tools for studying their behavior are applied to an artificial example, which demonstrates their general usefulness in studying many practical, applied problems.

This page titled 16.2: Elements of Markov Sequences is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

16.3: Problems on Conditional Independence, Given a Random Vector

Exercise 16.3.1

The pair $\{X, Y\}$ ci $|H$. X ~ exponential $(u/3)$, given $H = u$; Y ~ exponential $(u/5)$, given $H = u$; and H ~ uniform [1, 2]. Determine a general formula for $P(X > r, Y > s)$, then evaluate for $r = 3$, $s = 10$.

Answer available at the end of this section.

Exercise 16.3.2

A small random sample of size $n = 12$ is taken to determine the proportion of the student body which favors a proposal to expand the student Honor Council by adding two additional members "at large." Prior information indicates that this proportion is about 0.6 = 3/5. From a Bayesian point of view, the population proportion is taken to be the value of a random variable H. It seems reasonable to assume a *prior* distribution H ~ beta (4,3), giving a maximum of the density at (4 - 1)/(4 + 3 - 2) = 3/5. Seven of the twelve interviewed favor the proposition. What is the best mean-square estimate of the proportion, given this result? What is the conditional distribution of H, given this result?

Answer available at the end of this section.

Exercise 16.3.3

Let $\{X_i : 1 \le i \le n\}$ be a random sample, given H. Set $W = (X_1, X_2, \cdots, X_n)$. Suppose X conditionally geometric (u), given $H = u$; i.e., suppose $P(X = k|H = u) = u(1 - u)^k$ for all $k \ge 0$. If H ~ uniform on [0, 1], determine the best mean square estimator for H, given W.

Answer available at the end of this section.

Exercise 16.3.4

Let $\{X_i : 1 \le i \le n\}$ be a random sample, given H. Set $W = (X_1, X_2, \cdots, X_n)$. Suppose X conditionally Poisson (u), given $H = u$; i.e., suppose $P(X = k|H = u) = e^{-u}u^k/k!$. If H ~ gamma (m, λ), determine the best mean square estimator for H, given W.

Answer available at the end of this section.

Exercise 16.3.5

Suppose $\{N, H\}$ is independent and $\{N, Y\}$ ci $|H$. Use properties of conditional expectation and conditional independence to show that

$$E[g(N)h(Y)|H] = E[g(N)]E[h(Y)|H] \text{ a.s.}$$

Answer available at the end of this section.

Exercise 16.3.6

Consider the composite demand D introduced in the section on Random Sums in "Random Selecton"

$$D = \sum_{n=0}^{\infty} I_{\{k\}}(N) X_n \text{ where } X_n = \sum_{k=0}^{n} Y_k,\ Y_0 = 0$$

Suppose $\{N, H\}$ is independent, $\{N, Y_i\}$ ci $|H$ for all i, and $E[Y_i|H] = e(H)$, invariant with i. Show that $E[D|H] = E[N]E[Y|H]$ a.s..

Answer available at the end of this section.

Exercise 16.3.7

The transition matrix P for a homogeneous Markov chain is as follows (in m-file npr16_07.m):

$$P = \begin{bmatrix} 0.23 & 0.32 & 0.02 & 0.22 & 0.21 \\ 0.29 & 0.41 & 0.10 & 0.08 & 0.12 \\ 0.22 & 0.07 & 0.31 & 0.14 & 0.26 \\ 0.32 & 0.15 & 0.05 & 0.33 & 0.15 \\ 0.08 & 0.23 & 0.31 & 0.09 & 0.29 \end{bmatrix}$$

a. Obtain the absolute values of the eigenvalues, then consider increasing powers of P to observe the convergence to the long run distribution.
b. Take an arbitrary initial distribution $p0$ (as a row matrix). The product $p0 * p^k$ is the distribution for stage k. Note what happens as k becomes large enough to give convergence to the long run transition matrix. Does the end result change with change of initial distribution $p0$?

Answer available at the end of this section.

Exercise 16.3.8

The transition matrix P for a homogeneous Markov chain is as follows (in m-file **npr16_08.m**):

$$P = \begin{bmatrix} 0.2 & 0.5 & 0.3 & 0 & 0 & 0 & 0 \\ 0.6 & 0.1 & 0.3 & 0 & 0 & 0 & 0 \\ 0.2 & 0.7 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0.1 & 0.3 & 0 & 0.2 & 0.1 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.2 & 0.2 & 0.2 & 0 \end{bmatrix}$$

a. Note that the chain has two subchains, with states {1, 2, 3} and {4, 5}. Draw a transition diagram to display the two separate chains. Can any state in one subchain be reached from any state in the other?
b. Check the convergence as in part (a) of Exercise 16.3.7. What happens to the state probabilities for states 6 and 7 in the long run? What does that signify for these states? Can these states be reached from any state in either of the subchains? How would you classify these states?

Answer available at the end of this section.

Exercise 16.3.9

The transition matrix P for a homogeneous Markov chain is as follows (in m-file npr16_09.m):

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.1 & 0.3 & 0.2 & 0 & 0.1 \\ 0 & 0.6 & 0 & 0 & 0 & 0 & 0.4 \\ 0 & 0 & 0.2 & 0.5 & 0 & 0.3 & 0 \\ 0 & 0 & 0.6 & 0.1 & 0 & 0.3 & 0 \\ 0.2 & 0.2 & 0.1 & 0.2 & 0 & 0.1 & 0.2 \\ 0 & 0 & 0.2 & 0.7 & 0 & 0.1 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}$$

a. Check the transition matrix P for convergence, as in part (a) of Exercise 16.3.7. How many steps does it take to reach convergence to four or more decimal places? Does this agree with the theoretical result?
b. Examine the long run transition matrix. Identify transient states.
c. The convergence does not make all rows the same. Note, however, that there are two subgroups of similar rows. Rearrange rows and columns in the long run Matrix so that identical rows are grouped. This suggests subchains. Rearrange the rows and columns in the transition matrix P and see that this gives a pattern similar to that for the matrix in Exercise 16.7.8. Raise the rearranged transition matrix to the power for convergence.

Answer available at the end of this section.

Exercise 16.3.10

Use the m-procedure inventory1 (in m-file **inventory1.m**) to obtain the transition matrix for maximum stock $M = 8$, reorder point $m = 3$, and demand D ~ Poisson(4).

a. Suppose initial stock is six. What will the distribution for X_n, $n = 1, 3, 5$ (i.e., the stock at the end of periods 1, 3, 5, before restocking)?

b. What will the long run distribution be?

Answer available at the end of this section.

Answers

Exercise 16.3.1

$$P(X > r, Y > s|H = u) = e^{-ur/3}e^{us/5} = e^{-au},\ a = \frac{r}{3} + \frac{s}{5}$$
$$P(X > r, Y > s) = \int e^{-au} f_H(u)\ du = \int_1^2 e^{-au}\ du = \frac{1}{a}[e^{-a} - e^{-2a}]$$
For $r = 3$, $s = 10$, $a = 3$, $P(X > 3, Y > 10) = \frac{1}{3}(e^{-3} - e^{-6}) = 0.0158$

Exercise 16.3.2

H ~ Beta (r, s), $r = 4$, $s = 3$, $n = 12$, $k = 7$

$$E[H|S = k] = \frac{k + r}{n + r + s} = \frac{7 + 4}{12 + 4 + 3} = \frac{11}{19}$$

Exercise 16.3.3

$$E[H|W = k] = \frac{E[HI_{\{k\}}(W)]}{E[I_{\{k\}}(W)]} = \frac{E[HI_{\{k\}}(W)|H]}{E[I_{\{k\}}(W)|H}$$
$$= \frac{\int uP(W = k|H = u) f_H(u)\ du}{\int P(W = k|H = u) f_H(u)\ du},\ k = (k_1, k_2, \cdots, k_n)$$
$$P(W = k|H = u) = \prod_{i=1}^n u(1 - u)^{k_i} = u^n(1 - u)^{k^*}\ k^* = \sum_{i=1}^n k_i$$
$$E[H|W = k] = \frac{\int_0^1 u^{n+1}(1 - u)^{k^*}\ du}{\int_0^1 u^n(1 - u)^{k^*}\ du} = \frac{\Gamma(n + 2)\Gamma(k^* + 1)}{\Gamma(n + 1 + k^* + 2)} \cdot \frac{\Gamma(n + k^* + 2)}{\Gamma(n + 1)\Gamma(k^* + 1)} =$$
$$\frac{n + 1}{n + k^* + 2}$$

Exercise 16.3.4

$$E[H|W = k] = \frac{\int uP(W = k|H = u) f_H(u)\ du}{\int P(W = k|H = u) f_H(u)\ du}$$
$$P(W = k|H = u) = \prod_{i=1}^n e^{-u}\frac{u^{k_i}}{k_i!} = e^{-nu}\frac{u^{k^*}}{A} k^* = \sum_{i=1}^n k_i$$
$$f_H(u) = \frac{\lambda^m u^{m-1} e^{-\lambda u}}{\Gamma(m)}$$

$$E[H|W=k]=\frac{\int_0^\infty u^{k^*+m}e^{-(\lambda+n)u}\,du}{\int_0^\infty u^{k^*+m-1}e^{-(\lambda+n)u}\,du}=\frac{\Gamma(m+k^*+1)}{(\lambda+n)^{k^*+m+1}}\cdot\frac{(\lambda+n)^{k^*+m}}{\Gamma(m+k^*)}=\frac{m+k^*}{\lambda+n}$$

Exercise 16.3.5

$E[g(N)h(H)|H]=E[g(N)|H]E[h(Y)|H]$ a.s. by (CI6) and
$E[g(N)|H]=E[g(N)]$ a.s. by (CE5).

Exercise 16.3.6

$$E[D|H]=\textstyle\sum_{n=1}^\infty E[I_{\{n\}}(N)X_n|H] \text{ a.s.}$$
$$E[I_{\{n\}}(N)X_n|H]=\textstyle\sum_{k=1}^n E[I_{\{n\}}(N)Y_k|H]=\sum_{k=1}^n P(N=n)E[Y|H]=P(N=n)nE[Y|H] \text{ a.s.}$$
$$E[D|H]=\textstyle\sum_{n=1}^\infty nP(N=n)E[Y|H]=E[N]E[Y|H] \text{ a.s.}$$

Exercise 16.3.7

```
ev = abs(eig(P))'
ev = 1.0000    0.0814    0.0814    0.3572    0.2429
a = ev(4).^[2 4 8 16 24]
a = 0.1276    0.0163    0.0003    0.0000    0.0000
% By P^16 the rows agree to four places
p0 = [0.5 0 0 0.3 0.2];     % An arbitrarily chosen p0
p4 = p0*P^4
p4 =    0.2297    0.2622    0.1444    0.1644    0.1992
p8 = p0*P^8
p8 =    0.2290    0.2611    0.1462    0.1638    0.2000
p16 = p0*P^16
p16 =   0.2289    0.2611    0.1462    0.1638    0.2000
p0a = [0 0 0 0 1];          % A second choice of p0
p16a = p0a*P^16
p16a =  0.2289    0.2611    0.1462    0.1638    0.2000
```

Exercise 16.3.8

Increasing power p^n show the probability of being in states 6, 7 go to zero. These states cannot be reached from any of the other states.

Exercise 16.3.9

Examination of p^{16} suggests set {2, 7} and {3, 4, 6} of states form subchains. Rearrangement of P may be done as follows:

```
PA = P([2 7 3 4 6 1 5], [2 7 3 4 6 1 5])
PA =
    0.6000    0.4000         0         0         0         0         0
    0.5000    0.5000         0         0         0         0         0
         0         0    0.2000    0.5000    0.3000         0         0
         0         0    0.6000    0.1000    0.3000         0         0
         0         0    0.2000    0.7000    0.1000         0         0
    0.2000    0.1000    0.1000    0.3000         0    0.1000    0.2000
    0.2000    0.2000    0.1000    0.2000    0.1000    0.2000         0
PA16 = PA^16
PA16 =
    0.5556    0.4444         0         0         0         0         0
    0.5556    0.4444         0         0         0         0         0
         0         0    0.3571    0.3929    0.2500         0         0
         0         0    0.3571    0.3929    0.2500         0         0
         0         0    0.3571    0.3929    0.2500         0         0
    0.2455    0.1964    0.1993    0.2193    0.1395    0.0000    0.0000
    0.2713    0.2171    0.1827    0.2010    0.1279    0.0000    0.0000
```

It is clear that original states 1 and 5 are transient.

Exercise 16.3.10

```
inventory1
Enter value M of maximum stock  8
Enter value m of reorder point  3
Enter row vector of demand values  0:20
Enter demand probabilities  ipoisson(4,0:20)
```

```
Result is in matrix P
p0 = [0 0 0 0 0 0 1 0 0];
p1 = p0*P
p1 =
  Columns 1 through 7
    0.2149    0.1563    0.1954    0.1954    0.1465    0.0733    0.0183
  Columns 8 through 9
         0         0
p3 = p0*P^3
p3 =
  Columns 1 through 7
    0.2494    0.1115    0.1258    0.1338    0.1331    0.1165    0.0812
  Columns 8 through 9
    0.0391    0.0096
p5 = p0*P^5
p5 =
  Columns 1 through 7
    0.2598    0.1124    0.1246    0.1311    0.1300    0.1142    0.0799
  Columns 8 through 9
    0.0386    0.0095
a = abs(eig(P))'
a =
  Columns 1 through 7
    1.0000    0.4427    0.1979    0.0284    0.0058    0.0005    0.0000
  Columns 8 through 9
    0.0000    0.0000
a(2)^16
ans =
   2.1759e-06       % Convergence to at least five decimals for P^16
pinf = p0*P^16      % Use arbitrary p0,  pinf approx p0*P^16
pinf =  Columns 1 through 7
    0.2622    0.1132    0.1251    0.1310    0.1292    0.1130    0.0789
  Columns 8 through 9
    0.0380    0.0093
```

This page titled 16.3: Problems on Conditional Independence, Given a Random Vector is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17: APPENDICES

17.1: Appendix A to Applied Probability- Directory of m-functions and m-procedures
17.2: Appendix B to Applied Probability- some mathematical aids
17.3: Appendix C- Data on some common distributions
17.4: Appendix D to Applied Probability- The standard normal distribution
17.5: Appendix E to Applied Probability - Properties of Mathematical Expectation
17.6: Appendix F to Applied Probability- Properties of conditional expectation, given a random vector
17.7: Appendix G to Applied Probability- Properties of conditional independence, given a random vector
17.8: Matlab files for "Problems" in "Applied Probability"

This page titled 17: Appendices is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.1: Appendix A to Applied Probability- Directory of m-functions and m-procedures

We use the term *m-function* to designate a user-defined function as distinct from the basic MATLAB functions which are part of the MATLAB package. For example, the m-function *minterm* produces the specified minterm vector. An *m-procedure* (or sometimes a *procedure*) is an m-file containing a set of MATLAB commands which carry out a prescribed set of operations. Generally, these will prompt for (or assume) certain data upon which the procedure is carried out. We use the term *m-program* to refer to either an m-function or an m-procedure.

In addition to the m-programs there is a collection of m-files with properly formatted data which can be entered into the workspace by calling the file.

Although the m-programs were written for MATLAB version 4.2, they work for versions 5.1, 5.2, and 7.0⁄4. The latter versions offer some new features which may make more efficient implementation of some of the m-programs, and which make possible some new ones. With one exception (so noted), these are not explored in this collection.

MATLAB features

Utilization of MATLAB resources is made possible by a systematic analysis of some features of the basic probability model. In particular, the minterm analysis of logical (or Boolean) combinations of events and the analysis of the structure of simple random variables with the aid of indicator functions and minterm analysis are exploited.

A number of standard features of MATLAB are utilized extensively. In addition to standard matrix algebra, we use:

Array arithmetic. This involves element by element calculations. For example, if `a, b` are matrices of the same size, then `a.*b` is the matrix obtained by multiplying corresponding elements in the two matrices to obtain a new matrix of the same size.
Relational operations, such as less than, equal, etc. to obtain zero-one matrices with ones at element positions where the conditions are met.
Logical operations on zero-one matrices utilizing logical operators *and*, *or*, and *not*, as well as certain related functions such as *any*, *all*, *not*, *find*, etc. *Note*. Relational operations and logical operations produce zero-one arrays, called *logical arrays*, which MATLAB treats differently from zero-one *numeric arrays*. A rectangular array in which some rows are logical arrays but others are not is treated as a numeric array. Any zero-one rectangular array can be converted to a numeric array (matrix) by the command `A = ones(size(A)).*A`,
Certain MATLAB functions, such as *meshgrid, sum, cumsum, prod, cumprod* are used repeatedly. The function *dot* for dot product does not work if either array is a logical array. If one of the pair is numeric, the command C = A*B' will work.

Auxiliary user-defined building blocks

csort.m

Description of Code:

One of the most useful is a special sorting and consolidation operation implemented in the m-function *csort*. A standard problem arises when each of a non distinct set of values has an associated probability. To obtain the distribution, it is necessary to sort the values and add the probabilities associated with each distinct value. The following m-function achieves these operations: function [t,p] = csort(T,P). **T** and **P** are matrices with the same number of elements. Values of **T** are sorted and identical values are consolidated; values of **P** corresponding to identical values of **T** are added. A number of derivative functions and procedures utilize csort. The following two are useful.

Answer available at the end of this section.

distinct.m

Description of Code:

distinct.m `function y = distinct(T)` determines and sorts the distinct members of matrix T.

Answer available at the end of this section.

freq.m

Description of Code:

freq.m sorts the distinct members of a matrix, counts the number of occurrences of each value, and calculates the cumulative relative frequencies.

Answer available at the end of this section.

dsum.m

Description of Code:

dsum.m`function y = dsum(v,w)` determines and sorts the distinct elements among the sums of pairs of elements of row vectors **v** and **w**.

Answer available at the end of this section.

rep.m

Description of Code:

rep.m`function y = rep(A,m,n)` replicates matrix **A**, *m* times vertically and *n* times horizontally. Essentially the same as the function *repmat* in MATLAB version 5, released December, 1996.

Answer available at the end of this section.

elrep.m

Description of Code:

elrep.m`function y = elrep(A,m,n)` replicates each element of **A**, m times vertically and n times horizontally.

Answer available at the end of this section.

kronf.m

Description of Code:

kronf.m`function y = kronf(A,B)` determines the Kronecker product of matrices **A**,**B** Achieves the same result for full matrices as the MATLAB function *kron*.

Answer available at the end of this section.

colcopy.m

Description of Code:

colcopy.m`function y = colcopy(v,n)` treats row or column vector **v** as a column vector and makes a matrix with n columns of **v**.

Answer available at the end of this section.

colcopyi.m

Description of Code:

colcopyi.m`function y = colcopyi(v,n)` treats row or column vector **v** as a column vector, reverses the order of the elements, and makes a matrix with *n* columns of the reversed vector.

Answer available at the end of this section.

rowcopy.m

Description of Code:

rowcopy.m`function y = rowcopy(v,n)` treats row or column vector **v** as a row vector and makes a matrix with n rows of **v**.

Answer available at the end of this section.

repseq.m

Description of Code:

repseq.m`function y = repseq(V,n)` replicates vector V n times—horizontally if V is a row vector and vertically if V is a column vector.

Answer available at the end of this section.

total.m

Description of Code:

total.m Total of all elements in a matrix, calculated by: `total(x) = sum(sum(x))`.

Answer available at the end of this section.

dispv.m

Description of Code:

dispv.m Matrices A, B are transposed and displayed side by side.

Answer available at the end of this section.

roundn.m

Description of Code:

roundn.m`function y = roundn(A,n)` rounds matrix **A** to n decimal places.

Answer available at the end of this section.

arrep.m

Description of Code:

arrep.m`function y = arrep(n,k)` forms all arrangements, with repetition, of k elements from the sequence $1:n$.

Answer available at the end of this section.

Minterm vectors and probabilities

The analysis of logical combinations of events (as sets) is systematized by the use of the minterm expansion. This leads naturally to the notion of minterm vectors. These are zero-one vectors which can be combined by logical operations. Production of the basic minterm patterns is essential to a number of operations. The following m-programs are key elements of various other programs.

minterm.m

Description of Code:

minterm.m`function y = minterm(n,k)` generates the kth minterm vector in a class of n.

Answer available at the end of this section.

mintable.m

Description of Code:

mintable.m`function y = mintable(n)` generates a table of minterm vectors by repeated use of the m-function *minterm*.

Answer available at the end of this section.

minvec3.m

Description of Code:

minvec3.m sets basic minterm vectors **A**, **B**, **C**, $\mathbf{A}^c$, $\mathbf{B}^c$, $\mathbf{C}^c$, for the class $\{A, B, C\}$. (Similarly for *minvec4.m, minvec5.m*, etc.)

Answer available at the end of this section.

minmap

Description of Code:

minmap`function y = minmap(pm)` reshapes a row or column vector **pm** of minterm probabilities into minterm map format.

Answer available at the end of this section.

binary.m

Description of Code:

binary.m`function y = binary(d,n)` converts a matrix *d* of floating point nonnegative integers to a matrix of binary equivalents, one on each row. Adapted from m-functions written by Hans Olsson and by Simon Cooke. Each matrix row may be converted to an unspaced string of zeros and ones by the device ys = setstr(y + '0').

Answer available at the end of this section.

mincalc.m

Description of Code:

mincalc.m The m-procedure *mincalc* determines minterm probabilities from suitable data. For a discussion of the data formatting and certain problems, see 2.6.

Answer available at the end of this section.

mincalct.m

Description of Code:

mincalct.m Modification of *mincalc*. Assumes mincalc has been run, calls for new target vectors and performs same calculations as mincalc.

Answer available at the end of this section.

Independent events

minprob.m

Description of Code:

minprob.m`function y = minprob(p)` calculates minterm probabilities for the basic probabilities in row or column vector **p**. Uses the m-functions *mintable, colcopy*.

Answer available at the end of this section.

imintest.m

Description of Code:

imintest.m`function y = imintest(pm)` checks minterm probabilities for independence.

Answer available at the end of this section.

ikn.m

Description of Code:

ikn.m`function y = ikn(P,k)` determines the probability of the occurrence of exactly k of the n independent events whose probabilities are in row or column vector **P**
(*k* may be a row or column vector of nonnegative integers less than or equal to n).

Answer available at the end of this section.

ckn.m

Description of Code:

ckn.m`function y = ckn(P,k)` determines the probability of the occurrence of k or more of the n independent events whose probabilities are in row or column vector **P** (k may be a row or column vector)

Answer available at the end of this section.

parallel.m

Description of Code:

parallel.m`function y = parallel(p)` determines the probability of a parallel combination of the independent events whose probabilities are in row or column vector **p**.

Answer available at the end of this section.

Conditional probability and conditional idependence

bayes.m

Description of Code:

bayes.m produces a Bayesian reversal of conditional probabilities. The input consists of $P(E|A_i)$ and $P(A_i)$ for a disjoint class $\{A_i : 1 \le i \le n\}$ whose union contains E. The procedure calculates $P(A_i|E)$ and $P(A_i|E^c)$ for $1 \le i \le n$.

Answer available at the end of this section.

odds.m

Description of Code:

odds.m The procedure calculates posterior odds for for a specified profile E. Assumes data have been entered by the procedure *oddsf* or *oddsp*.

Answer available at the end of this section.

oddsdf.m

Description of Code:

oddsdf.m Sets up calibrating frequencies for calculating posterior odds.

Answer available at the end of this section.

oddsdp.m

Description of Code:

oddsdp.m Sets up conditional probabilities for odds calculations.

Answer available at the end of this section.

Bernoulli and multinomial trials

btdata.m

Description of Code:

btdata.m Sets parameter p and number n of trials for generating Bernoulli sequences. Prompts for *bt* to generate the trials.

Answer available at the end of this section.

bt.m

Description of Code:

bt.m Generates Bernoulli sequence for parameters set by btdata. Calculates relative frequency of “successes.”

Answer available at the end of this section.

binomial.m

Description of Code:

binomial.m Uses ibinom and cbinom to generate *tables* of the individual and cumulative binomial probabilities for specified parameters. *Note* that for calculation in MATLAB it is usually much more convenient and efficient to use *ibinom* and/or *cbinom*.

Answer available at the end of this section.

multinom.m

Description of Code:

multinom.m Multinomial distribution (small N, m).

Answer available at the end of this section.

Some matching problems

Cardmatch.m

Description of Code:

Cardmatch.m Sampling to estimate the probability of one or more matches when one card is drawn from each of nd identical decks of c cards. The number nsns of samples is specified.

Answer available at the end of this section.

trialmatch.m

Description of Code:

trialmatch.m Estimates the probability of matches in n independent trials from identical distributions. The sample size and number of trials must be kept relateively small to avoid exceeding available memory.

Answer available at the end of this section.

Distributions

comb.m

Description of Code:

comb.m`function y = comb(n,k)` Calculates binomial coefficients. k may be a matrix of integers between 0 and n. The result y is a matrix of the same dimensions.

Answer available at the end of this section.

ibinom.m

Description of Code:

ibinom.m Binomial distribution — individual terms. We have two m-functions *ibinom* and *cbinom* for calculating individual and cumulative terms $P(S_n = k)$ and $P(S_n \geq k)$, respectively.

$P(S_n = k) = C(n,k)p^k(1-p)^{n-k}$ and $P(S_n \geq k) = \sum_{r=k}^{n} P(S_n = r)$ $0 \leq k \leq n$

For these m-functions, we use a modification of a computation strategy employed by S. Weintraub: Tables of the Cumulative Binomial Probability Distribution for Small Values of p, 1963. The book contains a particularly helpful error analysis, written by Leo J. Cohen. Experimentation with sums and expectations indicates a precision for ibinom and cbinom calculations that

is better than 10^{-10} for $n = 1000$ and p from 0.01 to 0.99. A similar precision holds for values of n up to 5000, provided np or nq are limited to approximately 500. Above this value for np or nq, the computations break down. For individual terms, `function y = ibinom(n,p,k)` calculates the probabilities for n a positive integer, k a matrix of integers between 0 and n. The output is a matrix of the corresponding binomial probabilities.

Answer available at the end of this section.

ipoisson.m

Description of Code:

ipoisson.m Poisson distribution — individual terms. As in the case of the binomial distribution, we have an m-function for the individual terms and one for the cumulative case. The m-functions *ipoisson* and *cpoisson* use a computational strategy similar to that used for the binomial case. Not only does this work for large μ, but the precision is at least as good as that for the binomial m-functions. Experience indicates that the m-functions are good for $\mu \leq 700$. They breaks down at about 710, largely because of limitations of the MATLAB exponential function. For individual terms, `function y = ipoisson(mu,k)`calculates the probabilities for mu a positive integer, k a row or column vector of nonnegative integers. The output is a row vector of the corresponding Poisson probabilities.

Answer available at the end of this section.

cpoisson.m

Description of Code:

cpoisson.m Poisson distribution—cumulative terms. `function y = cpoisson(mu,k)`, calculates $P(X \geq k)$, where k may be a row or a column vector of nonnegative integers. The output is a row vector of the corresponding probabilities.

Answer available at the end of this section.

nbinom.m

Description of Code:

nbinom.m Negative binomial — `function y = nbinom(m, p, k)` calculates the probability that the mth success in a Bernoulli sequence occurs on the kth trial.

Answer available at the end of this section.

gaussian.m

Desciption of Code:

gaussian.m`function y = gaussian(m, v, t)` calculates the Gaussian (Normal) distribution function for mean value m, variance v, and matrix t of values. The result $y = P(X \leq t)$ is a matrix of the same dimensions as t.

Answer available at the end of this section.

gaussdensity.m

Description of Code:

gaussdensity.m`function y = gaussdensity(m,v,t)` calculates the Gaussian density function $f_X(t)$ for mean value m, variance t, and matrix t of values.

Answer available at the end of this section.

norminv.m

Description of Code:

norminv.m`function y = norminv(m,v,p)` calculates the inverse (the quantile function) of the Gaussian distribution function for mean value m, variance v, and p a matrix of probabilities.

Answer available at the end of this section.

gammadbn.m

Description of Code:

gammadbn.m`function y = gammadbn(alpha, lambda, t)` calculates the distribution function for a gamma distribution with parameters alpha, lambda. t is a matrix of evaluation points. The result is a matrix of the same size.

Answer available at the end of this section.

beta.m

Description of Code:

beta.m`function y = beta(r,s,t)` calculates the density function for the beta distribution with parameters r, s, t is a matrix of numbers between zero and one. The result is a matrix of the same size.

Answer available at the end of this section.

betadbn.m

Description of Code:

betadbn.m`function y = betadbn(r,s,t)` calculates the distribution function for the beta distribution with parameters r, s, t is a matrix of evaluation points. The result is a matrix of the same size.

Answer available at the end of this section.

weibull.m

Description of Code:

weibull.m`function y = weibull(alpha,lambda,t)` calculates the density function for the Weibull distribution with parameters alpha, lambda. t is a matrix of evaluation points. The result is a matrix of the same size.

Answer available at the end of this section.

weibulld.m

Description of Code:

weibulld.m`function y = weibulld(alpha, lambda, t)` calculates the distribution function for the Weibull distribution with parameters alpha, lambda. t is a matrix of evaluation points. The result is a matrix of the same size.

Answer available at the end of this section.

Binomial, Poisson, and Gaussian dstributions

bincomp.m

Description of Code:

bincomp.m Graphical comparison of the binomial, Poisson, and Gaussian distributions. The procedure calls for binomial parameters n, p, determines a reasonable range of evaluation points and plots on the same graph the binomial distribution function, the Poisson distribution function, and the gaussian distribution function with the adjustment called the "continuity correction."

Answer available at the end of this section.

poissapp.m

Desciption of Code:

poissapp.m Graphical comparison of the Poisson and Gaussian distributions. The procedure calls for a value of the Poisson parameter mu, then calculates and plots the Poisson distribution function, the Gaussian distribution function, and the adjusted Gaussian distribution function.

Answer available at the end of this section.

Setup for simple random variables

If a simple random variable X is in canonical form, the distribution consists of the coefficients of the indicator funtions (the values of X) and the probabilities of the corresponding events. If X is in a primitive form other than canonical, the csort operation is applied to the coefficients of the indicator functions and the probabilities of the corresponding events to obtain the distribution. If $Z = g(X)$ and X is in a primitive form, then the value of Z on the event in the partition associated with t_i is $g(t_i)$. The distribution for Z is obtained by applying csort to the $g(t_i)$ and the p_i. Similarly, if $Z = g(X, Y)$ and the joint distribution is available, the value $g(t_i, u_j)$ is associated with $P(X = t_i, Y = u_j)$. The distribution for Z is obtained by applying csort to the matrix of values and the corresponding matrix of probabilities.

canonic.m

Description of Code:

canonic.m The procedure determines the distribution for a simple random variable in affine form, when the minterm probabilities are available. Input data are a row vector of coefficients for the indicator functions in the affine form (with the constant value last) and a row vector of the probabilities of the minterm generated by the events. Results consist of a row vector of values and a row vector of the corresponding probabilities.

Answer available at the end of this section.

canonicf.m

Description of Code:

canonicf.m`function [x,px] = canonicf(c,pm)` is a function version of canonic, which allows arbitrary naming of variables.

Answer available at the end of this section.

jcalc.m

Description of Code:

jcalc.m Sets up for calculations for joint simple random variables. The matrix P of $P(X = t_i, Y = u_j)$ is arranged as on the plane (i.e., values of Y increase upward). The MATLAB function meshgrid is applied to the row matrix X and the reversed row matrix for Y to put an appropriate X-value and Y-value at each position. These are in the "calculating matrices" t and u, respectively, which are used in determining probabilities and expectations of various functions of t, u.

Answer available at the end of this section.

jcalcf.m

Description of Code:

jcalcf.m`function [x,y,t,u,px,py,p] = jcalcf(X,Y,P)` is a function version of jcalc, which allows arbitrary naming of variables.

Answer available at the end of this section.

jointzw.m

Description of Code:

jointzw.m Sets up joint distribution for $Z = g(X, Y)$ and $W = h(X, Y)$ and provides calculating matrices as in jcalc. Inputs are P, X and Y as well as array expressions for $g(t, u)$ and $h(t, u)$. Outputs are matrices Z, W, PZW for the joint distribution, marginal probabilities PZ, PW, and the calculating matrices v, w.

Answer available at the end of this section.

jdtest.m

Description of Code:

jdtest.m Tests a joint probability matrix P for negative entries and unit total probability..

Answer available at the end of this section.

Setup for general random variables

tappr.m

Description of Code:

tappr.m Uses the density function to set up a discrete approximation to the distribution for absolutely continuous random variable X.

Answer available at the end of this section.

tuappr.m

Description of Code:

tuappr.m Uses the joint density to set up discrete approximations to X, Y, t, u, and density.

Answer available at the end of this section.

dfappr.m

dfappr.m Approximate discrete distribution from distribution function entered as a function of t.

Answer available at the end of this section.

acsetup.m

Description of Code:

acsetup.m Approximate distribution for absolutely continuous random variable X. Density is entered as a *string variable*function of t.

Answer available at the end of this section.

dfsetup.m

Description of Code:

dfsetup.m Approximate discrete distribution from distribution function entered as a *string variable* function of t.

Answer available at the end of this section.

Setup for independent simple random variables

MATLAB version 5.1 has provisions for multidimensional arrays, which make possible more direct implementation of icalc3 and icalc4.

icalc.m

Description of Code:

icalc.m Calculation setup for an independent pair of simple random variables. Input consists of marginal distributions for X, Y. Output is joint distribution and calculating matrices t, u.

Answer available at the end of this section.

icalcf.m

icalcf.m`[x,y,t,u,px,py,p] = icalcf(X,Y,PX,PY)` is a function version of icalc, which allows arbitrary naming of variables.

Answer available at the end of this section.

icalc3.m

Description of Code:

icalc3.m Calculation setup for an independent class of three simple random variables.

Answer available at the end of this section.

icalc4.m

Description of Code:

icalc4.m Calculation setup for an independent class of four simple random variables.

Answer available at the end of this section.

Calculations for random variables

ddbn.m

Description of Code:

ddbn.m Uses the distribution of a simple random variable (or simple approximation) to plot a step graph for the distribution function F_X

Answer available at the end of this section.

cdbn.m

Description of Code:

cdbn.m Plots a continuous graph of a distribution function of a simple random variable (or simple approximation).

Answer available at the end of this section.

simple.m

Description of Code:

simple.m Calculates basic quantites for simple random variables from the distribution, input as row matrices X and PX.

Answer available at the end of this section.

jddbn.m

Description of Code:

jddbn.m Representation of joint distribution function for simple pair by obtaining the value of F_{XY} at the lower left hand corners of each grid cell.

Answer available at the end of this section.

jsimple.m

Description of Code:

jsimple.m Calculates basic quantities for a joint simple pair $\{X, Y\}$ from the joint distrsibution X, Y, P as in jcalc. Calculated quantities include means, variances, covariance, regression line, and regression curve (conditional expectation $E[Y|X = t]$

Answer available at the end of this section.

japprox.m

Description of Code:

japprox.m Assumes discrete setup and calculates basic quantities for a pair of random variables as in jsimple. Plots the regression line and regression curve.

Answer available at the end of this section.

Calculations and tests for independent random variables

mgsum.m

Description of Code:

mgsum.m`function [z,pz] = mgsum(x,y,px,py)` determines the distribution for the sum of an independent pair of simple random variables from their distributions.

Answer available at the end of this section.

mgsum3.m

Description of Code:

mgsum3.m`function [w,pw] = mgsum3(x,y,z,px,py,pz)` extends mgsum to three random variables by repeated application of mgsum. Similarly for mgsum4.m.

Answer available at the end of this section.

mgnsum.m

Description of Code:

mgnsum.m`function [z,pz] = mgnsum(X,P)` determines the distribution for a sum of n independent random variables. X an n-row matrix of X-values and n-row matrix of P-values (padded with zeros, if necessary, to make all rows the same length.

Answer available at the end of this section.

mgsumn.m

Description of Code:

mgsumn.m`function [z,pz] = mgsumn(varargin)` is an alternate to mgnsum, utilizing *varargin* in MATLAB version 5.1. The call is of the form `[z,pz] = mgsumn([x1;p1],[x2;p2], ..., [xn;pn])`.

Answer available at the end of this section.

diidsum.m

Description of Code:

diidsum.m`function [x,px] = diidsum(X,PX,n)` determines the sum of n iid simple random variables, with the common distribution X, PX

Answer available at the end of this section.

itest.m

Description of Code:

itest.m Tests for independence the matrix P of joint probabilities for a simple pair $\{X, Y\}$ of random variables.

Answer available at the end of this section.

idbn.m

Description of Code:

idbn.m`function p = idbn(px,py)` uses marginal probabilities to determine the joint probability matrix (arranged as on the plane) for an independent pair of simple random variables.

Answer available at the end of this section.

isimple.m

Description of Code:

isimple.m Takes as inputs the marginal distributions for an independent pair $\{X, Y\}$ of simple random variables. Sets up the joint distribution probability matrix P as in *idbn*, and forms the calculating matrices t, u as in *jcalc*. Calculates basic quantities and makes available matrices X, Y, PX, PY, P, t, u, for additional calculations.

Answer available at the end of this section.

Quantile functions for bounded distributions

dquant.m

Description of Code:

dquant.m`function t = dquant(X,PX,U)` determines the values of the quantile function for a simple random variable with distribution X, PX at the probability values in row vector U. The probability vector U is often determined by a random number generator.

Answer available at the end of this section.

dquanplot.m

Description of Code:

dquanplot.m Plots as a stairs graph the quantile function for a simple random variable X. The plot is the values of X versus the distribution function F_X.

Answer available at the end of this section.

dsample.m

Description of Code:

dsample.m Calculates a sample from a discrete distribution, determines the relative frequencies of values, and compares with actual probabilities. Input consists of value and probability matrices for X and the sample size n. A matrix U is determined by a random number generator, and the m-function dquant is used to calculate the corresponding sample values. Various data on the sample are calculated and displayed.

Answer available at the end of this section.

quanplot.m

Description of Code:

quanplot.m Plots the quantile function for a distribution function F_X. Assumes the procedure *dfsetup* or *acsetup* has been run. A suitable set U of probability values is determined and the m-function *dquant* is used to determine corresponding values of the quantile function. The results are plotted.

Answer available at the end of this section.

qsample.m

Description of Code:

qsample.m Simulates a sample for a given population density. Determines sample parameters and approximate population parameters. Assumes dfsetup or acsetup has been run. Takes as input the distribution matrices X, PX and the sample size n. Uses a random number generator to obtain the probability matrix U and uses the m-function *dquant* to determine the sample. Assumes *dfsetup* or *acsetup* has been run.

Answer available at the end of this section.

targetset.m

Description of Code:

targetset.m Setup for arrival at a target set of values. Used in conjunction with the m-procedure *targetrun* to determine the number of trials needed to visit k of a specified set of target values. Input consists of the distribution matrices X, PX and the specified set E of target values.

Answer available at the end of this section.

targetrun.m

Description of Code:

targetrun.m Assumes the m-file *targetset* has provided the basic data. Input consists of the number r of repetitions and the number k of the target states to visit. Calculates and displays various results.

Answer available at the end of this section.

Compound demand

The following pattern provides a useful model in many situations. Consider

$$D = \sum_{k=0}^{N} Y_k$$

where $Y_0 = 0$, and the class $\{Y_k : 1 \le k\}$ is iid, independent of the counting random variable N. One natural interpretation is to consider N to be the number of customers in a store and Y_k the amount purchased by the kth customer. Then D is the total demand of the actual customers. Hence, we call D the *compound demand*.

gend.m

Description of Code:

gend.m Uses coefficients of the generating functions for N and Y to calculate, in the integer case, the marginal distribution for the compound demand D and the joint distribution for $\{N, D\}$

Answer available at the end of this section.

gendf.m

Description of Code:

gendf.m`function [d,pd] = gendf(gn,gy)` is a function version of *gend*, which allows arbitrary naming of the variables. Calculates the distribution for D, but not the joint distribution for $\{N, D\}$

Answer available at the end of this section.

mgd.m

Description of Code:

mgd.m Uses coefficients for the generating function for N and the distribution for simple Y to calculate the distribution for the compound demand.

Answer available at the end of this section.

Exercise 17.1.1

Description of Code:

mgdf.m`function [d,pd] = mgdf(pn,y,py)` is a function version of *mgd*, which allows arbitrary naming of the variables. The input matrix pn is the coefficient matrix for the counting random variable generating function. Zeros for the missing powers must be included. The matrices y, py are the actual values and probabilities of the demand random variable.

Answer available at the end of this section.

Exercise 17.1.1

Description of Code:

randbern.m Let S be the number of successes in a random number N of Bernoulli trials, with probability p of success on each trial. The procedure randbern takes as inputs the probability p of success and the distribution matrices N, PN for the counting random variable N and calculates the joint distribution for $\{N, S\}$ and the marginal distribution for S.

Answer available at the end of this section.

Simulation of Markov systems

Exercise 17.1.1

Description of Code:

inventory1.m Calculates the transition matrix for an (m, M) inventory policy. At the end of each period, if the stock is less than a reorder point m, stock is replenished to the level M. Demand in each period is an integer valued random variable Y. Input consists of the parameters m, M and the distribution for *Y* as a simple random variable (or a discrete approximation).

Answer available at the end of this section.

branchp.m

Description of Code:

branchp.m Calculates the transition matrix for a simple branching process with a specified maximum population. Input consists of the maximum population value M and the coefficient matrix for the generating function for the individual propagation random variables Z_i. The latter matrix must include zero coefficients for missing powers.

Answer available at the end of this section.

chainset.m

Description of Code:

chainset.m Sets up for simulation of Markov chains. Inputs are the transition matrix **P** the set of states, and an optional set of target states. The chain generating procedures listed below assume this procedure has been run.

Answer available at the end of this section.

mchain.m

Description of Code:

mchain.m Assumes *chainset* has been run. Generates trajectory of specified length, with specified initial state.

Answer available at the end of this section.

arrival.m

Description of Code:

arrival.m Assumes *chainset* has been run. Calculates repeatedly the arrival time to a prescribed set of states.

Answer available at the end of this section.

recurrence.m

Description of Code:

recurrence.m Assumes *chainset* has been run. Calculates repeatedly the recurrence time to a prescribed set of states, if initial state is in the set; otherwise calculates the arrival time.

Answer available at the end of this section.

kvis.m

Description of Code:

kvis.m Assumes *chainset* has been run. Calculates repeatedly the time to complete visits to a specified k of the states in a prescribed set.

Answer available at the end of this section.

plotdbn

Description of Code:

plotdbn Used after m-procedures *arrival* or *recurrence* to plot arrival or recurrence time distribution.

Answer available at the end of this section.

Answers

csort.m

```
function [t,p] = csort(T,P)
% CSORT  [t,p] = csort(T,P) Sorts T, consolidates P
% Version of 4/6/97
% Modified to work with Versions 4.2 and 5.1, 5.2
% T and P matrices with the same number of elements
% The vector T(:)' is sorted:
%   * Identical values in T are consolidated;
%   * Corresponding values in P are added.
T = T(:)';
n = length(T);
[TS,I] = sort(T);
d  = find([1,TS(2:n) - TS(1:n-1) >1e-13]); % Determines distinct values
t  = TS(d);                                % Selects the distinct values
m  = length(t) + 1;
P  = P(I);                                 % Arranges elements of P
F  = [0 cumsum(P(:)')];
Fd = F([d length(F)]);                     % Cumulative sums for distinct values
p  = Fd(2:m) - Fd(1:m-1);                  % Separates the sums for these values
```

distinct.m

```
function y = distinct(T)
% DISTINCT y = distinct(T) Disinct* members of T
% Version of 5/7/96  Rev 4/20/97 for version 4 & 5.1, 5.2
% Determines distinct members of matrix T.
% Members which differ by no more than 10^{-13}
% are considered identical.  y is a row
% vector of the distinct members.
TS = sort(T(:)');
n  = length(TS);
d  = [1  abs(TS(2:n) - TS(1:n-1)) >1e-13];
y  = TS(find(d));
```

freq.m

```
% FREQ file freq.m Frequencies of members of matrix
% Version of 5/7/96
% Sorts the distinct members of a matrix, counts
% the number of occurrences of each value, and
% calculates the cumulative relative frequencies.
T  = input('Enter matrix to be counted  ');
[m,n] = size(T);
[t,f] = csort(T,ones(m,n));
```

```
p = cumsum(f)/(m*n);
disp(['The number of entries is  ',num2str(m*n),])
disp(['The number of distinct entries is  ',num2str(length(t)),] )
disp(' ')
dis = [t;f;p]';
disp('    Values     Count   Cum Frac')
disp(dis)
```

dsum.m

```
function y = dsum(v,w)
% DSUM y = dsum(v,w) Distinct pair sums of elements
% Version of 5/15/97
% y is a row vector of distinct
% values among pair sums of elements
% of matrices v, w.
% Uses m-function distinct
[a,b] = meshgrid(v,w);
t = a+b;
y = distinct(t(:)');
```

rep.m

```
function y = rep(A,m,n)
% REP y = rep(A,m,n) Replicates matrix A
% Version of 4/21/96
% Replicates A,
% m times vertically,
% n times horizontally
% Essentially the same as repmat in version 5.1, 5.2
[r,c] = size(A);
R = [1:r]';
C = [1:c]';
v = R(:,ones(1,m));
w = C(:,ones(1,n));
y = A(v,w);
```

elrep.m

```
function y = elrep(A,m,n)
% ELREP y = elrep(A,m,n) Replicates elements of A
% Version of 4/21/96
% Replicates each element,
% m times vertically,
% n times horizontally
[r,c] = size(A);
R = 1:r;
C = 1:c;
v = R(ones(1,m),:);
w = C(ones(1,n),:);
y = A(v,w);
```

kronf.m

```
function y = kronf(A,B)
% KRONF y = kronf(A,B) Kronecker product
% Version of 4/21/96
% Calculates Kronecker product of full matrices.
% Uses m-functions elrep and rep
% Same result for full matrices as kron for version 5.1, 5.2
```

```
[r,c] = size(B);
[m,n] = size(A);
y = elrep(A,r,c).*rep(B,m,n);
```

colcopy.m

```
function y = colcopy(v,n)
% COLCOPY y = colcopy(v,n)  n columns of v
% Version of 6/8/95 (Arguments reversed 5/7/96)
% v a row or column vector
% Treats v as column vector
% and makes n copies
% Procedure based on "Tony's trick"
[r,c] = size(v);
if r == 1
  v = v';
end
y = v(:,ones(1,n));
```

colcopyi.m

```
function y = colcopyi(v,n)
% COLCOPYI y = colcopyi(v,n) n columns in reverse order
% Version of 8/22/96
% v a row or column vector.
% Treats v as column vector,
% reverses the order of the
% elements, and makes n copies.
% Procedure based on "Tony's trick"
N = ones(1,n);
[r,c] = size(v);
if r == 1
  v = v(c:-1:1)';
else
  v = v(r:-1:1);
end
y = v(:,N);
```

rowcopy.m

```
function y = rowcopy(v,n)
% ROWCOPY y = rowcopy(v,n)  n rows of v
% Version of 5/7/96
% v a row or column vector
% Treats v as row vector
% and makes n copies
% Procedure based on "Tony's trick"
[r,c] = size(v);
if c == 1
  v = v';
end
y = v(ones(1,n),:);
```

repseq.m

```
function y = repseq(V,n);
% REPSEQ y = repseq(V,n) Replicates vector V n times
% Version of 3/27/97
% n replications of vector V
% Horizontally if V a row vector
```

```
% Vertically if V a column vector
m = length(V);
s = rem(0:n*m-1,m)+1;
y = V(s);
```

total.m

```
function y = total(x)
% TOTAL y = total(x)
% Version of 8/1/93
% Total of all elements in matrix x.
y = sum(sum(x));
```

dispv.m

```
function y = dispv(A,B)
% DISPV y = dispv(A,B) Transpose of A, B side by side
% Version of 5/3/96
% A, B are matrices of the same size
% They are transposed and displayed
% side by side.
y  = [A;B]';
```

roundn.m

```
function y = roundn(A,n);
% ROUNDN y = roundn(A,n)
% Version of 7/28/97
% Rounds matrix A to n decimals
y = round(A*10^n)/10^n;
```

arrep.m

```
function y = arrep(n,k);
% ARREP  y = arrep(n,k);
% Version of 7/28/97
% Computes all arrangements of k elements of 1:n,
% with repetition allowed. k may be greater than n.
% If only one input argument n, then k = n.
% To get arrangements of column vector V, use
% V(arrep(length(V),k)).
N = 1:n;
if nargin == 1
  k = n;
end
y = zeros(k,n^k);
for i = 1:k
  y(i,:) = rep(elrep(N,1,n^(k-i)),1,n^(i-1));
end
```

minterm.m

```
function y = minterm(n,k)
% MINTERM y = minterm(n,k) kth minterm of class of n
% Version of 5/5/96
% Generates the kth minterm vector in a class of n
% Uses m-function rep
y = rep([zeros(1,2^(n-k)) ones(1,2^(n-k))],1,2^(k-1));
```

mintable.m

```
function y = mintable(n)
% MINTABLE y = mintable(n)  Table of minterms vectors
% Version of 3/2/93
% Generates a table of minterm vectors
% Uses the m-function minterm
y = zeros(n,2^n);
for i = 1:n
    y(i,:) = minterm(n,i);
end
```

minvec3.m

```
% MINVEC3  file minvec3.m Basic minterm vectors
% Version of 1/31/95
A = minterm(3,1);
B = minterm(3,2);
C = minterm(3,3);
Ac = ~A;
Bc = ~B;
Cc = ~C;
disp('Variables are A, B, C, Ac, Bc, Cc')
disp('They may be renamed, if desired.')
```

minmap

```
function y = minmap(pm)
% MINMAP y = minmap(pm) Reshapes vector of minterm probabilities
% Version of 12/9/93
% Reshapes a row or column vector pm of minterm
% probabilities into minterm map format
m = length(pm);
n = round(log(m)/log(2));
a = fix(n/2);
if m ~= 2^n
  disp('The number of minterms is incorrect')
else
  y = reshape(pm,2^a,2^(n-a));
end
```

binary.m

```
function y = binary(d,n)
% BINARY y = binary(d,n) Integers to binary equivalents
% Version of 7/14/95
% Converts a matrix d of floating point, nonnegative
% integers to a matrix of binary equivalents. Each row
% is the binary equivalent (n places) of one number.
% Adapted from the programs dec2bin.m, which shared
% first prize in an April 95 Mathworks contest.
% Winning authors: Hans Olsson from Lund, Sweden,
% and Simon Cooke from Glasgow, UK.
% Each matrix row may be converted to an unspaced string
% of zeros and ones by the device:  ys = setstr(y + '0').
if nargin < 2,  n = 1; end     % Allows omission of argument n
[f,e] = log2(d);
n = max(max(max(e)),n);
y = rem(floor(d(:)*pow2(1-n:0)),2);
```

mincalc.m

```
% MINCALC file mincalc.m Determines minterm probabilities
% Version of 1/22/94 Updated for version 5.1 on  6/6/97
% Assumes a data file which includes
%  1. Call for minvecq to set q basic minterm vectors, each (1 x 2^q)
%  2. Data vectors DV = matrix of md data Boolean combinations of basic sets--
%     Matlab produces md minterm vectors-- one on each row.
%     The first combination is always A|Ac (the whole space)
%  3. DP = row matrix of md data probabilities.
%     The first probability is always 1.
%  4. Target vectors TV = matrix of mt target Boolean combinations.
%     Matlab produces a row minterm vector for each target combination.
%     If there are no target combinations, set TV = [];
[md,nd] = size(DV);
ND = 0:nd-1;
ID = eye(nd);                % Row i is minterm vector i-1
[mt,nt] = size(TV);
MT = 1:mt;
rd = rank(DV);
if rd < md
   disp('Data vectors are NOT linearly independent')
  else
   disp('Data vectors are linearly independent')
end
% Identification of which minterm probabilities can be determined from the data
% (i.e., which minterm vectors are not linearly independent of data vectors)
AM = zeros(1,nd);
for i = 1:nd
  AM(i) = rd == rank([DV;ID(i,:)]);  % Checks for linear dependence of each
  end
am = find(AM);                             % minterm vector
CAM = ID(am,:)/DV;     % Determination of coefficients for the available minterms
pma = DP*CAM';               % Calculation of probabilities of available minterms
PMA = [ND(am);pma]';
if sum(pma < -0.001) > 0      % Check for data consistency
   disp('Data probabilities are INCONSISTENT')
else
% Identification of which target probabilities are computable from the data
CT = zeros(1,mt);
for j = 1:mt
  CT(j) = rd == rank([DV;TV(j,:)]);
  end
ct  = find(CT);
CCT = TV(ct,:)/DV;           % Determination of coefficients for computable targets
ctp = DP*CCT';               % Determination of probabilities
disp(' Computable target probabilities')
disp([MT(ct); ctp]')
end                          % end for "if sum(pma < -0.001) > 0"
disp(['The number of minterms is ',num2str(nd),])
disp(['The number of available minterms is ',num2str(length(pma)),])
disp('Available minterm probabilities are in vector pma')
disp('To view available minterm probabilities, call for PMA')
```

mincalct.m

```
% MINCALCT file mincalct.m  Aditional target probabilities
% Version of 9/1/93  Updated for version 5 on 6/6/97
% Assumes a data file which includes
%  1. Call for minvecq to set q basic minterm vectors.
```

```
%  2. Data vectors DV.  The first combination is always A|Ac.
%  3. Row matrix DP of data probabilities. The first entry is always 1.
TV = input('Enter matrix of target Boolean combinations  ');
[md,nd] = size(DV);
[mt,nt] = size(TV);
MT = 1:mt;
rd = rank(DV);
CT = zeros(1,mt);   % Identification of computable target probabilities
for j = 1:mt
  CT(j) = rd == rank([DV;TV(j,:)]);
end
ct  = find(CT);
CCT = TV(ct,:)/DV;  % Determination of coefficients for computable targets
ctp = DP*CCT';      % Determination of probabilities
disp(' Computable target probabilities')
disp([MT(ct); ctp]')
```

minprob.m

```
function y = minprob(p)
% MINPROB y = minprob(p) Minterm probs for independent events
% Version of 4/7/96
% p is a vector [P(A1) P(A2) ... P(An)], with
% {A1,A2, ... An} independent.
% y is the row vector of minterm probabilities
% Uses the m-functions mintable, colcopy
n = length(p);
M = mintable(n);
a = colcopy(p,2^n);          % 2^n columns, each the vector p
m = a.*M + (1 - a).*(1 - M); % Puts probabilities into the minterm
                             % pattern on its side (n by 2^n)
y = prod(m);                 % Product of each column of m
```

imintest.m

```
function y = imintest(pm)
% IMINTEST y = imintest(pm) Checks minterm probs for independence
% Version of 1/25//96
% Checks minterm probabilities for independence
% Uses the m-functions mintable and minprob
m = length(pm);
n = round(log(m)/log(2));
if m ~= 2^n
  y = 'The number of minterm probabilities is incorrect';
else
P = mintable(n)*pm';
pt = minprob(P');
a = fix(n/2);
s = abs(pm - pt) > 1e-7;
  if sum(s) > 0
    disp('The class is NOT independent')
    disp('Minterms for which the product rule fails')
    y = reshape(s,2^a,2^(n-a));
  else
    y = 'The class is independent';
  end
end
```

ikn.m

```
function y = ikn(P,k)
% IKN y = ikn(P,k) Individual probabilities of k of n successes
% Version of 5/15/95
% Uses the m-functions mintable, minprob, csort
n = length(P);
T = sum(mintable(n));  % The number of successes in each minterm
pm = minprob(P);       % The probability of each minterm
[t,p] = csort(T,pm);   % Sorts and consolidates success numbers
                       % and adds corresponding probabilities
y = p(k+1);
```

ckn.m

```
function y = ckn(P,k)
% CKN y = ckn(P,k) Probability of k or more successes
% Version of 5/15/95
% Probabilities of k or more of n independent events
% Uses the m-functions mintable, minprob, csort
n = length(P);
m = length(k);
T = sum(mintable(n));  % The number of successes in each minterm
pm = minprob(P);       % The probability of each minterm
[t,p] = csort(T,pm);   % Sorts and consolidates success numbers
                       % and adds corresponding probabilities
for i = 1:m            % Sums probabilities for each k value
  y(i) = sum(p(k(i)+1:n+1));
end
```

parallel.m

```
function y = parallel(p)
% PARALLEL y = parallel(p) Probaaability of parallel combination
% Version of 3/3/93
% Probability of parallel combination.
% Individual probabilities in row matrix p.
y = 1 - prod(1 - p);
```

bayes.m

```
% BAYES file bayes.m Bayesian reversal of conditional probabilities
% Version of 7/6/93
% Input P(E|Ai) and P(Ai)
% Calculates P(Ai|E) and P(Ai|Ec)
disp('Requires input PEA = [P(E|A1) P(E|A2) ... P(E|An)]')
disp(' and PA = [P(A1) P(A2) ... P(An)]')
disp('Determines PAE  = [P(A1|E) P(A2|E) ... P(An|E)]')
disp('       and PAEc = [P(A1|Ec) P(A2|Ec) ... P(An|Ec)]')
PEA  = input('Enter matrix PEA of conditional probabilities  ');
PA   = input('Enter matrix  PA of probabilities  ');
PE   = PEA*PA';
PAE  = (PEA.*PA)/PE;
PAEc = ((1 - PEA).*PA)/(1 -  PE);
disp(' ')
disp(['P(E) = ',num2str(PE),])
disp(' ')
disp('    P(E|Ai)   P(Ai)     P(Ai|E)   P(Ai|Ec)')
disp([PEA; PA; PAE; PAEc]')
disp('Various quantities are in the matrices PEA, PA, PAE, PAEc, named above')
```

odds.m

```
% ODDS file odds.m  Posterior odds for profile
% Version of 12/4/93
% Calculates posterior odds for profile E
% Assumes data has been entered by oddsdf or oddsdp
E = input('Enter profile matrix E  ');
C =  diag(a(:,E))';       % aa = a(:,E) is an n by n matrix whose ith column
D =  diag(b(:,E))';       % is the E(i)th column of a.  The elements on the
                          % diagonal are b(i, E(i)), 1 <= i <= n
                          % Similarly for b(:,E)

R = prod(C./D)*(p1/p2);   % Calculates posterior odds for profile
disp(' ')
disp(['Odds favoring Group 1:   ',num2str(R),])
if R > 1
  disp('Classify in Group 1')
else
  disp('Classify in Group 2')
end
```

oddsdf.m

```
% ODDSDF file oddsdf.m Frequencies for calculating odds
% Version of 12/4/93
% Sets up calibrating frequencies
% for calculating posterior odds
A = input('Enter matrix A of frequencies for calibration group 1  ');
B = input('Enter matrix B of frequencies for calibration group 2  ');
n = length(A(:,1));       % Number of questions (rows of A)
m = length(A(1,:));       % Number of answers to each question
p1 = sum(A(1,:));         % Number in calibration group 1
p2 = sum(B(1,:));         % Number in calibration group 2
a = A/p1;
b = B/p2;
disp(' ')                 % Blank line in presentation
disp(['Number of questions = ',num2str(n),])  % Size of profile
disp(['Answers per question = ',num2str(m),]) % Usually 3: yes, no, uncertain
disp(' Enter code for answers and call for procedure "odds"  ')
disp(' ')
```

oddsdp.m

```
  % ODDSDP file oddsdp.m Conditional probs for calculating posterior odds
% Version of 12/4/93
% Sets up conditional probabilities
% for odds calculations
a  = input('Enter matrix A of conditional probabilities for Group 1  ');
b  = input('Enter matrix B of conditional probabilities for Group 2  ');
p1 = input('Probability p1 an individual is from Group 1  ');
n = length(a(:,1));
m = length(a(1,:));
p2 = 1 - p1;
disp(' ')                 % Blank line in presentation
disp(['Number of questions = ',num2str(n),])  % Size of profile
disp(['Answers per question = ',num2str(m),]) % Usually 3: yes, no, uncertain
disp(' Enter code for answers and call for procedure "odds"  ')
disp(' ')
```

btdata.m

```
  % BTDATA file btdata.m Parameters for Bernoulli trials
% Version of 11/28/92
% Sets parameters for generating Bernoulli trials
% Prompts for bt to generate the trials
n = input('Enter n, the number of trials  ');
p = input('Enter p, the probability of success on each trial  ');
disp(' ')
disp(' Call for bt')
disp(' ')
```

bt.m

```
% BT file bt.m Generates Bernoulli sequence
% version of 8/11/95  Revised 7/31/97 for version 4.2 and 5.1, 5.2
% Generates Bernoulli sequence for parameters set by btdata
% Calculates relative frequency of 'successes'
clear SEQ;
B  = rand(n,1) <= p;         %  ones for random numbers <= p
F  = sum(B)/n;               % relative frequency of ones
N  = [1:n]';                 % display details
disp(['n = ',num2str(n),'   p = ',num2str(p),])
disp(['Relative frequency = ',num2str(F),])
SEQ = [N B];
clear N;
clear B;
disp('To view the sequence, call for SEQ')
disp(' ')
```

binomial.m

```
% BINOMIAL file binomial.m Generates binomial tables
% Version of 12/10/92  (Display modified 4/28/96)
% Calculates a TABLE of binomial probabilities
% for specified n, p, and row vector k,
% Uses the m-functions ibinom and cbinom.
n = input('Enter n, the number of trials ');
p = input('Enter p, the probability of success ');
k = input('Enter k, a row vector of success numbers ');
y = ibinom(n,p,k);
z = cbinom(n,p,k);
disp(['    n = ',int2str(n),'    p = ' num2str(p)])
H = ['    k         P(X = k)  P(X >= k)'];
disp(H)
disp([k;y;z]')
```

multinom.m

```
% MULTINOM file multinom.m  Multinomial distribution
% Version of 8/24/96
% Multinomial distribution (small N, m)
N = input('Enter the number of trials  ');
m = input('Enter the number of types   ');
p = input('Enter the type probabilities  ');
M = 1:m;
T = zeros(m^N,N);
for i = 1:N
  a = rowcopy(M,m^(i-1));
  a = a(:);
  a = colcopy(a,m^(N-i));
```

```
  T(:,N-i+1) = a(:);        % All possible strings of the types
end
MT = zeros(m^N,m);
for i = 1:m
 MT(:,i) = sum(T'==i)';
end
clear T                     % To conserve memory
disp('String frequencies for type k are in column matrix MT(:,k)')
P = zeros(m^N,N);
for i = 1:N
  a = rowcopy(p,m^(i-1));
  a = a(:);
  a = colcopy(a,m^(N-i));
  P(:,N-i+1) = a(:);        % Strings of type probabilities
end
PS = prod(P');              % Probability of each string
clear P                     % To conserve memory
disp('String probabilities are in row matrix PS')
```

Cardmatch.m

```
% CARDMATCH file cardmatch.m Prob of matches in cards from identical decks
% Version of 6/27/97
% Estimates the probability of one or more matches
% in drawing cards from nd decks of c cards each
% Produces a supersample of size n = nd*ns, where
% ns is the number of samples
% Each sample is sorted, and then tested for differences
% between adjacent elements.  Matches are indicated by
% zero differences between adjacent elements in sorted sample
c  = input('Enter the number c  of cards in a deck ');
nd = input('Enter the number nd of decks ');
ns = input('Enter the number ns of sample runs ');
X  = 1:c;                   % Population values
PX = (1/c)*ones(1,c);       % Population probabilities
N  = nd*ns;                 % Length of supersample
U  = rand(1,N);             % Matrix of n random numbers
T  = dquant(X,PX,U);        % Supersample obtained with quantile function;
                            % the function dquant determines quantile
                            % function values of random number sequence U
ex = sum(T)/N;              % Sample average
EX = dot(X,PX);             % Population mean
vx = sum(T.^2)/N - ex^2;    % Sample variance
VX = dot(X.^2,PX) - EX^2;   % Population variance
A  = reshape(T,nd,ns);      % Chops supersample into ns samples of size nd
DS = diff(sort(A));         % Sorts each sample
m  = sum(DS==0)>0;          % Differences between elements in each sample
                            % Zero difference iff there is a match
pm = sum(m)/ns;             % Fraction of samples with one or more matches
Pm = 1 - comb(c,nd)*gamma(nd + 1)/c^(nd);  % Theoretical probability of match
disp('The sample is in column vector T')   % Displays of results
disp(['Sample average ex = ', num2str(ex),])
disp(['Population mean E(X) = ',num2str(EX),])
disp(['Sample variance vx = ',num2str(vx),])
disp(['Population variance V(X) = ',num2str(VX),])
disp(['Fraction of samples with one or more matches   pm = ', num2str(pm),])
disp(['Probability of one or more matches in a sample Pm = ', num2str(Pm),])
```

trialmatch.m

```
% TRIALMATCH file trialmatch.m  Estimates probability of matches
% in n independent trials from identical distributions
% Version of 8/20/97
% Estimates the probability of one or more matches
% in a random selection from n identical distributions
% with a small number of possible values
% Produces a supersample of size N = n*ns, where
% ns is the number of samples.  Samples are separated.
% Each sample is sorted, and then tested for differences
% between adjacent elements.  Matches are indicated by
% zero differences between adjacent elements in sorted sample.
X  = input('Enter the VALUES in the distribution ');
PX = input('Enter the PROBABILITIES  ');
c  = length(X);
n  = input('Enter the SAMPLE SIZE n ');
ns = input('Enter the number ns of sample runs ');
N  = n*ns;                  % Length of supersample
U  = rand(1,N);             % Vector of N random numbers
T  = dquant(X,PX,U);        % Supersample obtained with quantile function;
                            %   the function dquant determines quantile
                            %   function values for random number sequence U
ex = sum(T)/N;              % Sample average
EX = dot(X,PX);             % Population mean
vx = sum(T.^2)/N - ex^2;    % Sample variance
VX = dot(X.^2,PX) - EX^2;   % Population variance
A  = reshape(T,n,ns);       % Chops supersample into ns samples of size n
DS = diff(sort(A));         % Sorts each sample
m  = sum(DS==0)>0;          % Differences between elements in each sample
                            % -- Zero difference iff there is a match
pm = sum(m)/ns;             % Fraction of samples with one or more matches
d  = arrep(c,n);
p  = PX(d);
p  = reshape(p,size(d));    % This step not needed in version 5.1
ds = diff(sort(d))==0;
mm = sum(ds)>0;
m0 = find(1-mm);
pm0 = p(:,m0);              % Probabilities for arrangements with no matches
P0 = sum(prod(pm0));
disp('The sample is in column vector T')   % Displays of results
disp(['Sample average ex = ', num2str(ex),])
disp(['Population mean E(X) = ',num2str(EX),])
disp(['Sample variance vx = ',num2str(vx),])
disp(['Population variance V(X) = ',num2str(VX),])
disp(['Fraction of samples with one or more matches   pm = ', num2str(pm),])
disp(['Probability of one or more matches in a sample Pm = ', num2str(1-P0),])
```

comb.m

```
function y = comb(n,k)
% COMB y = comb(n,k) Binomial coefficients
% Version of 12/10/92
% Computes binomial coefficients C(n,k)
% k may be a matrix of integers between 0 and n
% result y is a matrix of the same dimensions
y = round(gamma(n+1)./(gamma(k + 1).*gamma(n + 1 - k)));
```

ibinom.m

```
function y = ibinom(n,p,k)
% IBINOM  y = ibinom(n,p,k) Individual binomial probabilities
% Version of 10/5/93
% n is a positive integer; p is a probability
% k a matrix of integers between 0 and n
% y = P(X>=k) (a matrix of probabilities)
if p > 0.5
a = [1 ((1-p)/p)*ones(1,n)];
b = [1 n:-1:1];
c = [1 1:n];
br = (p^n)*cumprod(a.*b./c);
bi = fliplr(br);
else
a = [1 (p/(1-p))*ones(1,n)];
b = [1 n:-1:1];
c = [1 1:n];
bi = ((1-p)^n)*cumprod(a.*b./c);
end
y = bi(k+1);
```

ipoisson.m

```
function y = ipoisson(mu,k)
% IPOISSON y = ipoisson(mu,k) Individual Poisson probabilities
% Version of 10/15/93
% mu = mean value
% k may be a row or column vector of integer values
% y = P(X = k) (a row vector of probabilities)
K = max(k);
p = exp(-mu)*cumprod([1 mu*ones(1,K)]./[1 1:K]);
y = p(k+1);
```

cpoisson.m

```
function y = cpoisson(mu,k)
% CPOISSON y = cpoisson(mu,k) Cumulative Poisson probabilities
% Version of 10/15/93
% mu = mean value mu
% k may be a row or column vector of integer values
% y = P(X >= k) (a row vector of probabilities)
K = max(k);
p = exp(-mu)*cumprod([1 mu*ones(1,K)]./[1 1:K]);
pc = [1 1 - cumsum(p)];
y = pc(k+1);
```

nbinom.m

```
function y = nbinom(m, p, k)
% NBINOM y = nbinom(m, p, k) Negative binomial probabilities
% Version of 12/10/92
% Probability the mth success occurs on the kth trial
% m a positive integer;  p a probability
% k a matrix of integers greater than or equal to m
% y = P(X=k) (a matrix of the same dimensions as k)
q = 1 - p;
y = ((p^m)/gamma(m)).*(q.^(k - m)).*gamma(k)./gamma(k - m + 1);
```

gaussian.m

```
function y = gaussian(m,v,t)
```

```
% GAUSSIAN y = gaussian(m,v,t) Gaussian distribution function
% Version of 11/18/92
% Distribution function for X ~ N(m, v)
% m = mean,  v = variance
% t is a matrix of evaluation points
% y = P(X<=t) (a matrix of the same dimensions as t)
u = (t - m)./sqrt(2*v);
if u >= 0
        y = 0.5*(erf(u) + 1);
else
        y = 0.5*erfc(-u);
end
```

gaussdensity.m

```
function y = gaussdensity(m,v,t)
% GAUSSDENSITY y = gaussdensity(m,v,t) Gaussian density
% Version of 2/8/96
% m = mean,  v = variance
% t is a matrix of evaluation points
y = exp(-((t-m).^2)/(2*v))/sqrt(v*2*pi);
```

norminv.m

```
function y = norminv(m,v,p)
% NORMINV y = norminv(m,v,p) Inverse gaussian distribution
% (quantile function for gaussian)
% Version of 8/17/94
% m = mean,  v = variance
% t is a matrix of evaluation points
if p >= 0
  u = sqrt(2)*erfinv(2*p - 1);
else
  u = -sqrt(2)*erfinv(1 - 2*p);
end
y = sqrt(v)*u + m;
```

gammadbn.m

```
function y = gammadbn(alpha, lambda, t)
% GAMMADBN y = gammadbn(alpha, lambda, t) Gamma distribution
% Version of 12/10/92
% Distribution function for X ~ gamma (alpha, lambda)
% alpha, lambda are positive parameters
% t may be a matrix of positive numbers
% y = P(X<= t) (a matrix of the same dimensions as t)
y = gammainc(lambda*t, alpha);
```

beta.m

```
function y = beta(r,s,t)
% BETA y = beta(r,s,t) Beta density function
% Version of 8/5/93
% Density function for Beta (r,s) distribution
% t is a matrix of evaluation points between 0 and 1
% y is a matrix of the same dimensions as t
y = (gamma(r+s)/(gamma(r)*gamma(s)))*(t.^(r-1).*(1-t).^(s-1));
```

betadbn.m

```
function y = betadbn(r,s,t)
```

```
% BETADBN y = betadbn(r,s,t) Beta distribution function
% Version of 7/27/93
% Distribution function for X  beta(r,s)
% y = P(X<=t) (a matrix of the same dimensions as t)
y = betainc(t,r,s);
```

weibull.m

```
function y = weibull(alpha,lambda,t)
% WEIBULL y = weibull(alpha,lambda,t) Weibull density
% Version of 1/24/91
% Density function for X ~ Weibull (alpha, lambda, 0)
% t is a matrix of positive evaluation points
% y is a matrix of the same dimensions as t
y = alpha*lambda*(t.^(alpha - 1)).*exp(-lambda*(t.^alpha));
```

weibulld.m

```
function y = weibulld(alpha, lambda, t)
% WEIBULLD y = weibulld(alpha, lambda, t) Weibull distribution function
% Version of 1/24/91
% Distribution function for X ~ Weibull (alpha, lambda, 0)
% t is a matrix of positive evaluation points
% y = P(X<=t) (a matrix of the same dimensions as t)
y = 1 - exp(-lambda*(t.^alpha));
```

bincomp.m

```
% BINCOMP file bincomp.m  Approx of binomial by Poisson and gaussian
% Version of 5/24/96
% Gaussian adjusted for "continuity correction"
% Plots distribution functions for specified parameters n, p
n = input('Enter the parameter n  ');
p = input('Enter the parameter p  ');
a = floor(n*p-2*sqrt(n*p));
a = max(a,1);                          % Prevents zero or negative indices
b = floor(n*p+2*sqrt(n*p));
k = a:b;
Fb = cumsum(ibinom(n,p,0:n));          % Binomial distribution function
Fp = cumsum(ipoisson(n*p,0:n));        % Poisson distribution function
Fg = gaussian(n*p,n*p*(1 - p),k+0.5); % Gaussian distribution function
stairs(k,Fb(k+1))                      % Plotting details
hold on
plot(k,Fp(k+1),'-.',k,Fg,'o')
hold off
xlabel('t values')                     % Graph labeling details
ylabel('Distribution function')
title('Approximation of Binomial by Poisson and Gaussian')
grid
legend('Binomial','Poisson','Adjusted Gaussian')
disp('See Figure for results')
```

poissapp.m

```
% POISSAPP file poissapp.m  Comparison of Poisson and gaussian
% Version of 5/24/96
% Plots distribution functions for specified parameter mu
mu = input('Enter the parameter mu  ');
n = floor(1.5*mu);
k = floor(mu-2*sqrt(mu)):floor(mu+2*sqrt(mu));
```

```
FP = cumsum(ipoisson(mu,0:n));
FG = gaussian(mu,mu,k);
FC = gaussian(mu,mu,k-0.5);
stairs(k,FP(k))
hold on
plot(k,FG,'-.',k,FC,'o')
hold off
grid
xlabel('t values')
ylabel('Distribution function')
title('Gaussian Approximation to Poisson Distribution')
legend('Poisson','Gaussian','Adjusted Gaussian')
disp('See Figure for results')
```

canonic.m

```
% CANONIC file canonic.m Distribution for simple rv in affine form
% Version of 6/12/95
% Determines the distribution for a simple random variable
% in affine form, when the minterm probabilities are available.
% Uses the m-functions mintable and csort.
% The coefficient vector must contain the constant term.
% If the constant term is zero, enter 0 in the last place.
c  = input(' Enter row vector of coefficients  ');
pm = input(' Enter row vector of minterm probabilities  ');
n  = length(c) - 1;
if 2^n ~= length(pm)
   error('Incorrect minterm probability vector length');
end
M  = mintable(n);            % Provides a table of minterm patterns
s  = c(1:n)*M + c(n+1);      % Evaluates X on each minterm
[X,PX] = csort(s,pm);        % s = values; pm = minterm probabilities
XDBN = [X;PX]';
disp('Use row matrices X and PX for calculations')
disp('Call for XDBN to view the distribution')
```

canonicf.m

```
function [x,px] = canonicf(c,pm)
% CANONICF  [x,px] = canonicf(c,pm)  Function version of canonic
% Version of 6/12/95
% Allows arbitrary naming of variables
n = length(c) - 1;
if 2^n ~= length(pm)
   error('Incorrect minterm probability vector length');
end
M  = mintable(n);              % Provides a table of minterm patterns
s  = c(1:n)*M + c(n+1);        % Evaluates X on each minterm
[x,px]  = csort(s,pm);         % s = values; pm = minterm probabilities
```

jcalc.m

```
% JCALC  file jcalc.m  Calculation setup for joint simple rv
% Version of 4/7/95 (Update of prompt and display 5/1/95)
% Setup for calculations for joint simple random variables
% The joint probabilities arranged as on the plane
% (top row corresponds to largest value of Y)
P = input('Enter JOINT PROBABILITIES (as on the plane)  ');
X = input('Enter row matrix of VALUES of X  ');
```

```
Y = input('Enter row matrix of VALUES of Y  ');
PX = sum(P);            % probabilities for X
PY = fliplr(sum(P'));   % probabilities for Y
[t,u] = meshgrid(X,fliplr(Y));
disp(' Use array operations on matrices X, Y, PX, PY, t, u, and P')
```

jcalcf.m

```
function [x,y,t,u,px,py,p] = jcalcf(X,Y,P)
% JCALCF [x,y,t,u,px,py,p] = jcalcf(X,Y,P) Function version of jcalc
% Version of 5/3/95
% Allows arbitrary naming of variables
if sum(size(P) ~= [length(Y) length(X)]) > 0
  error('     Incompatible vector sizes')
end
x = X;
y = Y;
p = P;
px = sum(P);
py = fliplr(sum(P'));
[t,u] = meshgrid(X,fliplr(Y));
```

jointzw.m

```
% JOINTZW file jointzw.m Joint dbn for two functions of (X,Y)
% Version of 4/29/97
% Obtains joint distribution for
% Z = g(X,Y) and W = h(X,Y)
% Inputs P, X, and Y as well as array
% expressions for g(t,u) and h(t,u)
P = input('Enter joint prob for (X,Y) ');
X = input('Enter values for X ');
Y = input('Enter values for Y ');
[t,u] = meshgrid(X,fliplr(Y));
G = input('Enter expression for g(t,u) ');
H = input('Enter expression for h(t,u) ');
[Z,PZ] = csort(G,P);
[W,PW] = csort(H,P);
r = length(W);
c = length(Z);
PZW = zeros(r,c);
for i = 1:r
  for j = 1:c
   a = find((G==Z(j))&(H==W(i)));
   if ~isempty(a)
    PZW(i,j) = total(P(a));
   end
  end
end
PZW = flipud(PZW);
[v,w] = meshgrid(Z,fliplr(W));
if (G==t)&(H==u)
  disp(' ')
  disp('  Note:  Z = X and W = Y')
  disp(' ')
elseif  G==t
  disp(' ')
  disp('  Note:  Z = X')
  disp(' ')
```

```
elseif H==u
  disp(' ')
  disp('  Note:  W = Y')
  disp(' ')
end
disp('Use array operations on Z, W, PZ, PW, v, w, PZW')
```

jdtest.m

```
function y = jdtest(P)
% JDTEST y = jdtest(P) Tests P for unit total and negative elements
% Version of 10/8/93
M = min(min(P));
S = sum(sum(P));
if M < 0
  y = 'Negative entries';
elseif abs(1 - S) > 1e-7
  y = 'Probabilities do not sum to one';
else
  y = 'P is a valid distribution';
end
```

tappr.m

```
% TAPPR file tappr.m  Discrete approximation to ac random variable
% Version of 4/16/94
% Sets up discrete approximation to distribution for
% absolutely continuous random variable  X
% Density is entered as a function of t
r = input('Enter matrix [a b] of x-range endpoints  ');
n = input('Enter number of x approximation points  ');
d = (r(2) - r(1))/n;
t = (r(1):d:r(2)-d) +d/2;
PX = input('Enter density as a function of t  ');
PX = PX*d;
PX = PX/sum(PX);
X  = t;
disp('Use row matrices X and PX as in the simple case')
```

tuappr.m

```
% TUAPPR file tuappr.m Discrete approximation to joint ac pair
% Version of 2/20/96
% Joint density entered as a function of t, u
% Sets up discrete approximations to X, Y, t, u, and density
rx = input('Enter matrix [a b] of X-range endpoints  ');
ry = input('Enter matrix [c d] of Y-range endpoints  ');
nx = input('Enter number of X approximation points  ');
ny = input('Enter number of Y approximation points  ');
dx = (rx(2) - rx(1))/nx;
dy = (ry(2) - ry(1))/ny;
X  = (rx(1):dx:rx(2)-dx) + dx/2;
Y  = (ry(1):dy:ry(2)-dy) + dy/2;
[t,u] = meshgrid(X,fliplr(Y));
P  = input('Enter expression for joint density  ');
P  = dx*dy*P;
P  = P/sum(sum(P));
PX = sum(P);
PY = fliplr(sum(P'));
```

```
disp('Use array operations on X, Y, PX, PY, t, u, and P')
```

dfappr.m

```
% DFAPPR file dfappr.m Discrete approximation from distribution function
% Version of 10/21/95
% Approximate discrete distribution from distribution
% function entered as a function of t
r = input('Enter matrix [a b] of X-range endpoints  ');
s = input('Enter number of X approximation points  ');
d = (r(2) - r(1))/s;
t = (r(1):d:r(2)-d) +d/2;
m  = length(t);
f  = input('Enter distribution function F as function of t  ');
f  = [0 f];
PX = f(2:m+1) - f(1:m);
PX = PX/sum(PX);
X  = t - d/2;
disp('Distribution is in row matrices X and PX')
```

acsetup.m

```
% ACSETUP file acsetup.m Discrete approx from density as string variable
% Version of 10/22/94
% Approximate distribution for absolutely continuous rv X
% Density is entered as a string variable function of t
disp('DENSITY f is entered as a STRING VARIABLE.')
disp('either defined previously or upon call.')
r  = input('Enter matrix [a b] of x-range endpoints  ');
s  = input('Enter number of x approximation points  ');
d  = (r(2) - r(1))/s;
t  = (r(1):d:r(2)-d) +d/2;
m  = length(t);
f  = input('Enter density as a function of t  ');
PX = eval(f);
PX = PX*d;
PX = PX/sum(PX);
X  = t;
disp('Distribution is in row matrices X and PX')
```

dfsetup.m

```
% DFSETUP file dfsetup.m  Discrete approx from string dbn function
% Version of 10/21/95
% Approximate discrete distribution from distribution
% function entered as string variable function of t
disp('DISTRIBUTION FUNCTION F is entered as a STRING')
disp('VARIABLE, either defined previously or upon call')
r = input('Enter matrix [a b] of X-range endpoints  ');
s = input('Enter number of X approximation points  ');
d = (r(2) - r(1))/s;
t = (r(1):d:r(2)-d) +d/2;
m  = length(t);
F  = input('Enter distribution function F as function of t  ');
f  = eval(F);
f  = [0 f];
PX = f(2:m+1) - f(1:m);
PX = PX/sum(PX);
X  = t - d/2;
```

```
disp('Distribution is in row matrices X and PX')
```

icalc.m

```
% ICALC file icalc.m  Calculation setup for independent pair
% Version of 5/3/95
% Joint calculation setup for independent pair
X  = input('Enter row matrix of X-values  ');
Y  = input('Enter row matrix of Y-values  ');
PX = input('Enter X probabilities  ');
PY = input('Enter Y probabilities  ');
[a,b] = meshgrid(PX,fliplr(PY));
P  = a.*b;                      % Matrix of joint independent probabilities
[t,u] = meshgrid(X,fliplr(Y));  % t, u matrices for joint calculations
disp(' Use array operations on matrices X, Y, PX, PY, t, u, and P')
```

icalcf.m

```
function [x,y,t,u,px,py,p] = icalcf(X,Y,PX,PY)
% ICALCF [x,y,t,u,px,py,p] = icalcf(X,Y,PX,PY) Function version of icalc
% Version of 5/3/95
% Allows arbitrary naming of variables
x = X;
y = Y;
px = PX;
py = PY;
if length(X) ~= length(PX)
  error('     X and PX of different lengths')
elseif length(Y) ~= length(PY)
  error('     Y and PY of different lengths')
end
[a,b] = meshgrid(PX,fliplr(PY));
p   = a.*b;                      % Matrix of joint independent probabilities
[t,u] = meshgrid(X,fliplr(Y));    % t, u matrices for joint calculations
```

icalc3.m

```
% ICALC3 file icalc3.m Setup for three independent rv
% Version of 5/15/96
% Sets up for calculations for three
% independent simple random variables
% Uses m-functions rep, elrep, kronf
X  = input('Enter row matrix of X-values  ');
Y  = input('Enter row matrix of Y-values  ');
Z  = input('Enter row matrix of Z-values  ');
PX = input('Enter X probabilities  ');
PY = input('Enter Y probabilities  ');
PZ = input('Enter Z probabilities  ');
n  = length(X);
m  = length(Y);
s  = length(Z);
[t,u] = meshgrid(X,Y);
t  = rep(t,1,s);
u  = rep(u,1,s);
v  = elrep(Z,m,n);  % t,u,v matrices for joint calculations
P  = kronf(PZ,kronf(PX,PY'));
disp('Use array operations on matrices X, Y, Z,')
disp('PX, PY, PZ, t, u, v, and P')
```

icalc4.m

```
% ICALC4 file icalc4.m Setup for four independent rv
% Version of 5/15/96
% Sets up for calculations for four
% independent simple random variables
% Uses m-functions rep, elrep, kronf
X  = input('Enter row matrix of X-values  ');
Y  = input('Enter row matrix of Y-values  ');
Z  = input('Enter row matrix of Z-values  ');
W  = input('Enter row matrix of W-values  ');
PX = input('Enter X probabilities  ');
PY = input('Enter Y probabilities  ');
PZ = input('Enter Z probabilities  ');
PW = input('Enter W probabilities  ');
n  = length(X);
m  = length(Y);
s  = length(Z);
r  = length(W);
[t,u] = meshgrid(X,Y);
t = rep(t,r,s);
u = rep(u,r,s);
[v,w] = meshgrid(Z,W);
v = elrep(v,m,n);  % t,u,v,w matrices for joint calculations
w = elrep(w,m,n);
P = kronf(kronf(PZ,PW'),kronf(PX,PY'));
disp('Use array operations on matrices X, Y, Z, W')
disp('PX, PY, PZ, PW, t, u, v, w, and P')
```

ddbn.m

```
% DDBN file ddbn.m Step graph of distribution function
% Version of 10/25/95
% Plots step graph of dbn function FX from
% distribution of simple rv (or simple approximation)
xc  = input('Enter row matrix of VALUES  ');
pc = input('Enter row matrix of PROBABILITIES  ');
m  = length(xc);
FX = cumsum(pc);
xt = [xc(1)-1-0.1*abs(xc(1)) xc xc(m)+1+0.1*abs(xc(m))];
FX = [0 FX 1];        % Artificial extension of range and domain
stairs(xt,FX)         % Plot of stairstep graph
hold on
plot(xt,FX,'o')       % Marks values at jump
hold off
grid
xlabel('t')
ylabel('u = F(t)')
title('Distribution Function')
```

cdbn.m

```
% CDBN file cdbn.m Continuous graph of distribution function
% Version of 1/29/97
% Plots continuous graph of dbn function FX from
% distribution of simple rv (or simple approximation)
xc  = input('Enter row matrix of VALUES  ');
pc = input('Enter row matrix of PROBABILITIES  ');
m  = length(xc);
FX = cumsum(pc);
```

```
xt = [xc(1)-0.01 xc xc(m)+0.01];
FX = [0 FX FX(m)];    % Artificial extension of range and domain
plot(xt,FX)           % Plot of continuous graph
grid
xlabel('t')
ylabel('u = F(t)')
title('Distribution Function')
```

simple.m

```
% SIMPLE file simple.m Calculates basic quantites for simple rv
% Version of 6/18/95
X  = input('Enter row matrix of X-values  ');
PX = input('Enter row matrix PX of X probabilities  ');
n  = length(X);          % dimension of X
EX = dot(X,PX)           % E[X]
EX2 = dot(X.^2,PX)       % E[X^2]
VX = EX2 - EX^2          % Var[X]
disp(' ')
disp('Use row matrices X and PX for further calculations')
```

jddbn.m

```
% JDDBN file jddbn.m  Joint distribution function
% Version of 10/7/96
% Joint discrete distribution function for
% joint  matrix P (arranged as on the plane).
% Values at lower left hand corners of grid cells
P = input('Enter joint probability matrix (as on the plane)  ');
FXY = flipud(cumsum(flipud(P)));
FXY = cumsum(FXY')';
disp('To view corner values for joint dbn function, call for FXY')
```

jsimple.m

```
% JSIMPLE file jsimple.m  Calculates basic quantities for joint simple rv
% Version of 5/25/95
% The joint probabilities are arranged as on the plane
% (the top row corresponds to the largest value of Y)
P = input('Enter JOINT PROBABILITIES (as on the plane)  ');
X = input('Enter row matrix of VALUES of X  ');
Y = input('Enter row matrix of VALUES of Y  ');
disp(' ')
PX = sum(P);              % marginal distribution for X
PY = fliplr(sum(P'));     % marginal distribution for Y
XDBN = [X; PX]';
YDBN = [Y; PY]';
PT  = idbn(PX,PY);
D  = total(abs(P - PT));  % test for difference
if D > 1e-8               % to prevent roundoff error masking zero
  disp('{X,Y} is NOT independent')
 else
  disp('{X,Y} is independent')
end
disp(' ')
[t,u] = meshgrid(X,fliplr(Y));
EX  = total(t.*P)         % E[X]
EY  = total(u.*P)         % E[Y]
EX2 = total((t.^2).*P)    % E[X^2]
```

```
EY2 = total((u.^2).*P)     % E[Y^2]
EXY = total(t.*u.*P)       % E[XY]
VX  = EX2 - EX^2           % Var[X]
VY  = EY2 - EY^2           % Var[Y]
cv = EXY - EX*EY;          % Cov[X,Y] = E[XY] - E[X]E[Y]
if abs(cv) > 1e-9          % to prevent roundoff error masking zero
   CV = cv
 else
   CV = 0
end
a = CV/VX                  % regression line of Y on X is
b = EY - a*EX              %       u = at + b
R = CV/sqrt(VX*VY);        % correlation coefficient rho
disp(['The regression line of Y on X is:  u = ',num2str(a),'t + ',num2str(b),])
disp(['The correlation coefficient is:  rho = ',num2str(R),])
disp(' ')
eYx = sum(u.*P)./PX;
EYX = [X;eYx]';
disp('Marginal dbns are in X, PX, Y, PY; to view, call XDBN, YDBN')
disp('E[Y|X = x] is in eYx; to view, call for EYX')
disp('Use array operations on matrices X, Y, PX, PY, t, u, and P')
```

japprox.m

```
% JAPPROX file japprox.m Basic quantities for ac pair {X,Y}
% Version of 5/7/96
% Assumes tuappr has set X, Y, PX, PY, t, u, P
EX  = total(t.*P)          % E[X]
EY  = total(u.*P)          % E[Y]
EX2 = total(t.^2.*P)       % E[X^2]
EY2 = total(u.^2.*P)       % E[Y^2]
EXY = total(t.*u.*P)       % E[XY]
VX  = EX2 - EX^2           % Var[X]
VY  = EY2 - EY^2           % Var[Y]
cv = EXY - EX*EY;          % Cov[X,Y] = E[XY] - E[X]E[Y]
if abs(cv) > 1e-9  % to prevent roundoff error masking zero
   CV = cv
 else
   CV = 0
end
a = CV/VX                  % regression line of Y on X is
b = EY - a*EX              % u = at + b
R = CV/sqrt(VX*VY);
disp(['The regression line of Y on X is:  u = ',num2str(a),'t + ',num2str(b),])
disp(['The correlation coefficient is:  rho = ',num2str(R),])
disp(' ')
eY = sum(u.*P)./sum(P);    % eY(t) = E[Y|X = t]
RL = a*X + b;
plot(X,RL,X,eY,'-.')
grid
title('Regression line and Regression curve')
xlabel('X values')
ylabel('Y values')
legend('Regression line','Regression curve')
clear eY                   % To conserve memory
clear RL
disp('Calculate with X, Y, t, u, P, as in joint simple case')
```

mgsum.m

```
function [z,pz] = mgsum(x,y,px,py)
% MGSUM [z,pz] = mgsum(x,y,px,py)  Sum of two independent simple rv
% Version of 5/6/96
% Distribution for the sum of two independent simple random variables
% x is a vector (row or column) of X values
% y is a vector (row or column) of Y values
% px is a vector (row or column) of X probabilities
% py is a vector (row or column) of Y probabilities
% z and pz are row vectors
[a,b] = meshgrid(x,y);
t  = a+b;
[c,d] = meshgrid(px,py);
p  = c.*d;
[z,pz]  = csort(t,p);
```

mgsum3.m

```
function [w,pw] = mgsum3(x,y,z,px,py,pz)
% MGSUM3 [w,pw] = mgsum3(x,y,z,px,py,y) Sum of three independent simple rv
% Version of 5/2/96
% Distribution for the sum of three independent simple random variables
% x is a vector (row or column) of X values
% y is a vector (row or column) of Y values
% z is a vector (row or column) of Z values
% px is a vector (row or column) of X probabilities
% py is a vector (row or column) of Y probabilities
% pz is a vector (row or column) of Z probabilities
% W and pW are row vectors
[a,pa] = mgsum(x,y,px,py);
[w,pw] = mgsum(a,z,pa,pz);
```

mgnsum.m

```
function [z,pz] = mgnsum(X,P)
% MGNSUM [z,pz] = mgnsum(X,P)  Sum of n independent simple rv
% Version of 5/16/96
% Distribution for the sum of n independent simple random variables
% X an n-row matrix of X-values
% P an n-row matrix of P-values
% padded with zeros, if necessary
% to make all rows the same length
[n,r] = size(P);
z  = 0;
pz = 1;
for i = 1:n
  x = X(i,:);
  p = P(i,:);
  x = x(find(p>0));
  p = p(find(p>0));
  [z,pz] = mgsum(z,x,pz,p);
end
```

mgsumn.m

```
function [z,pz] = mgsumn(varargin)
% MGSUMN [z,pz] = mgsumn([x1;p1],[x2;p2], ..., [xn;pn])
% Version of 6/2/97 Uses MATLAB version 5.1
% Sum of n independent simple random variables
% Utilizes distributions in the form [x;px] (two rows)
```

```
% Iterates mgsum
n  = length(varargin);   % The number of distributions
z  = 0;                  % Initialization
pz = 1;
for i = 1:n              % Repeated use of mgsum
  [z,pz] = mgsum(z,varargin{i}(1,:),pz,varargin{i}(2,:));
end
```

diidsum.m

```
function [x,px] = diidsum(X,PX,n)
% DIIDSUM [x,px] = diidsum(X,PX,n) Sum of n iid simple random variables
% Version of 10/14/95 Input rev 5/13/97
% Sum of n iid rv with common distribution X, PX
% Uses m-function mgsum
x  = X;                       % Initialization
px = PX;
for i = 1:n-1
  [x,px] = mgsum(x,X,px,PX);
end
```

itest.m

```
% ITEST file itest.m  Tests P for independence
% Version of 5/9/95
% Tests for independence the matrix of joint
% probabilities for a simple pair {X,Y}
pt = input('Enter matrix of joint probabilities  ');
disp(' ')
px = sum(pt);                  % Marginal probabilities for X
py = sum(pt');                 % Marginal probabilities for Y (reversed)
[a,b] = meshgrid(px,py);
PT = a.*b;                     % Joint independent probabilities
D  = abs(pt - PT) > 1e-9;      % Threshold set above roundoff
if total(D) > 0
  disp('The pair {X,Y} is NOT independent')
  disp('To see where the product rule fails, call for D')
else
  disp('The pair {X,Y} is independent')
end
```

idbn.m

```
function p = idbn(px,py)
% IDBN p = idbn(px,py)  Matrix of joint independent probabilities
% Version of 5/9/95
% Determines joint probability matrix for two independent
% simple random variables (arranged as on the plane)
[a,b] = meshgrid(px,fliplr(py));
p = a.*b
```

isimple.m

```
% ISIMPLE file isimple.m  Calculations for independent simple rv
% Version of 5/3/95
X  = input('Enter row matrix of X-values  ');
Y  = input('Enter row matrix of Y-values  ');
PX = input('Enter X probabilities  ');
PY = input('Enter Y probabilities  ');
[a,b] = meshgrid(PX,fliplr(PY));
```

```
P  = a.*b;                      % Matrix of joint independent probabilities
[t,u] = meshgrid(X,fliplr(Y));  % t, u matrices for joint calculations
EX  = dot(X,PX)                 % E[X]
EY  = dot(Y,PY)                 % E[Y]
VX  = dot(X.^2,PX) - EX^2       % Var[X]
VY  = dot(Y.^2,PY) - EY^2       % Var[Y]
disp(' Use array operations on matrices X, Y, PX, PY, t, u, and P')
```

dquant.m

```
function t = dquant(X,PX,U)
% DQUANT t = dquant(X,PX,U)  Quantile function for a simple random variable
% Version of 10/14/95
% U is a vector of probabilities
m  = length(X);
n  = length(U);
F  = [0 cumsum(PX)+1e-12];
F(m+1) = 1;                     % Makes maximum value exactly one
if U(n) >= 1                    % Prevents improper values of probability U
  U(n) = 1;
end
if U(1) <= 0
  U(1) = 1e-9;
end
f  = rowcopy(F,n);              % n rows of F
u  = colcopy(U,m);              % m columns of U
t  = X*((f(:,1:m) < u)&(u <= f(:,2:m+1)))';
```

dquanplot.m

```
% DQUANPLOT file dquanplot.m  Plot of quantile function for a simple rv
% Version of 7/6/95
% Uses stairs to plot the inverse of FX
X  = input('Enter VALUES for X  ');
PX = input('Enter PROBABILITIES for X  ');
m  = length(X);
F  = [0 cumsum(PX)];
XP = [X X(m)];
stairs(F,XP)
grid
title('Plot of Quantile Function')
xlabel('u')
ylabel('t = Q(u)')
hold on
plot(F(2:m+1),X,'o')          % Marks values at jumps
hold off
```

dsample.m

```
% DSAMPLE file dsample.m  Simulates sample from discrete population
% Version of 12/31/95 (Display revised 3/24/97)
% Relative frequencies vs probabilities for
% sample from discrete population distribution
X  = input('Enter row matrix of VALUES  ');
PX = input('Enter row matrix of PROBABILITIES  ');
n  = input('Sample size n  ');
U  = rand(1,n);
T  = dquant(X,PX,U);
[x,fr] = csort(T,ones(1,length(T)));
```

```
disp('    Value      Prob    Rel freq')
disp([x; PX; fr/n]')
ex = sum(T)/n;
EX = dot(X,PX);
vx = sum(T.^2)/n - ex^2;
VX = dot(X.^2,PX) - EX^2;
disp(['Sample average ex = ',num2str(ex),])
disp(['Population mean E[X] = ',num2str(EX),])
disp(['Sample variance vx = ',num2str(vx),])
disp(['Population variance Var[X] = ',num2str(VX),])
```

quanplot.m

```
% QUANPLOT file quanplot.m  Plots quantile function for dbn function
% Version of 2/2/96
% Assumes dfsetup or acsetup has been run
% Uses m-function dquant
X  = input('Enter row matrix of values  ');
PX = input('Enter row matrix of probabilities  ');
h  = input('Probability increment h  ');
U  = h:h:1;
T  = dquant(X,PX,U);
U  = [0 U 1];
Te = X(m) + abs(X(m))/20;
T  = [X(1) T Te];
plot(U,T)             % Plot rather than stairs for general case
grid
title('Plot of Quantile Function')
xlabel('u')
ylabel('t = Q(u)')
```

qsample.m

```
% QSAMPLE file qsample.m  Simulates sample for given population density
% Version of 1/31/96
% Determines sample parameters
% and approximate population parameters.
% Assumes dfsetup or acsetup has been run
X  = input('Enter row matrix of VALUES  ');
PX = input('Enter row matrix of PROBABILITIES  ');
n  = input('Sample size n =  ');
m  = length(X);
U  = rand(1,n);
T  = dquant(X,PX,U);
ex = sum(T)/n;
EX = dot(X,PX);
vx = sum(T.^2)/n - ex^2;
VX = dot(X.^2,PX) - EX^2;
disp('The sample is in column vector T')
disp(['Sample average ex = ', num2str(ex),])
disp(['Approximate population mean E(X) = ',num2str(EX),])
disp(['Sample variance vx = ',num2str(vx),])
disp(['Approximate population variance V(X) = ',num2str(VX),])
```

targetset.m

```
% TARGETSET file targetset.m  Setup for sample arrival at target set
% Version of 6/24/95
X  = input('Enter population VALUES  ');
```

```
PX = input('Enter population PROBABILITIES  ');
ms = length(X);
x = 1:ms;                    % Value indices
disp('The set of population values is')
disp(X);
E  = input('Enter the set of target values  ');
ne = length(E);
e  = zeros(1,ne);
for i = 1:ne
  e(i) = dot(E(i) == X,x);  % Target value indices
end
F  = [0 cumsum(PX)];
A  = F(1:ms);
B  = F(2:ms+1);
disp('Call for targetrun')
```

targetrun.m

```
% TARGETRUN file targetrun.m Number of trials to visit k target values
% Version of 6/24/95  Rev for Version 5.1 1/30/98
% Assumes the procedure targetset has been run.
r  = input('Enter the number of repetitions  ');
disp('The target set is')
disp(E)
ks = input('Enter the number of target values to visit  ');
if isempty(ks)
  ks = ne;
end
if ks > ne
  ks = ne;
end
clear T             % Trajectory in value indices (reset)
R0 = zeros(1,ms);   % Indicator for target value indices
R0(e) = ones(1,ne);
S  = zeros(1,r);    % Number of trials for each run (reset)
for k = 1:r
  R = R0;
  i = 1;
  while sum(R) > ne - ks
    u = rand(1,1);
    s = ((A < u)&(u <= B))*x';
    if R(s) == 1     % Deletes indices as values reached
      R(s) = 0;
    end
    T(i) = s;
    i = i+1;
  end
  S(k) = i-1;
end
if r == 1
  disp(['The number of trials to completion is ',int2str(i-1),])
  disp(['The initial value is ',num2str(X(T(1))),])
  disp(['The terminal value is ',num2str(X(T(i-1))),])
  N  = 1:i-1;
  TR = [N;X(T)]';
  disp('To view the trajectory, call for TR')
else
  [t,f] = csort(S,ones(1,r));
```

```
  D  = [t;f]';
  p  = f/r;
  AV = dot(t,p);
  SD = sqrt(dot(t.^2,p) - AV^2);
  MN = min(t);
  MX = max(t);
  disp(['The average completion time is ',num2str(AV),])
  disp(['The standard deviation is ',num2str(SD),])
  disp(['The minimum completion time is ',int2str(MN),])
  disp(['The maximum completion time is ',int2str(MX),])
  disp(' ')
  disp('To view a detailed count, call for D.')
  disp('The first column shows the various completion times;')
  disp('the second column shows the numbers of trials yielding those times')
  plot(t,cumsum(p))
  grid
  title('Fraction of Runs t Steps or Less')
  ylabel('Fraction of runs')
  xlabel('t = number of steps to complete run')
end
```

gend.m

```
% GEND file gend.m   Marginal and joint dbn for integer compound demand
% Version of 5/21/97
% Calculates marginal distribution for compound demand D
% and joint distribution for {N,D} in the integer case
% Do not forget zero coefficients for missing powers
% in the generating functions for N, Y
disp('Do not forget zero coefficients for missing powers')
gn = input('Enter gen fn COEFFICIENTS for gN  ');
gy = input('Enter gen fn COEFFICIENTS for gY  ');
n  = length(gn) - 1;           % Highest power in gN
m  = length(gy) - 1;           % Highest power in gY
P  = zeros(n + 1,n*m + 1);     % Base for generating P
y  = 1;                        % Initialization
P(1,1) = gn(1);                % First row of P (P(N=0) in the first position)
for i = 1:n                    % Row by row determination of P
   y  = conv(y,gy);            % Successive powers of gy
   P(i+1,1:i*m+1) = y*gn(i+1); % Successive rows of P
end
PD = sum(P);                   % Probability for each possible value of D
a  = find(gn);                 % Location of nonzero N probabilities
b  = find(PD);                 % Location of nonzero D probabilities
P  = P(a,b);                   % Removal of zero rows and columns
P  = rot90(P);                 % Orientation as on the plane
N  = 0:n;
N  = N(a);                     % N values with positive probabilites
PN = gn(a);                    % Positive N probabilities
Y  = 0:m;                      % All possible values of Y
Y  = Y(find(gy));              % Y values with positive probabilities
PY = gy(find(gy));             % Positive Y proabilities
D  = 0:n*m;                    % All possible values of D
PD = PD(b);                    % Positive D probabilities
D  = D(b);                     % D values with positive probabilities
gD = [D; PD]';                 % Display combination
disp('Results are in N, PN, Y, PY, D, PD, P')
disp('May use jcalc or jcalcf on N, D, P')
```

```
disp('To view distribution for D, call for gD')
```

gendf.m

```
function [d,pd] = gendf(gn,gy)
% GENDF [d,pd] = gendf(gN,gY) Function version of gend.m
% Calculates marginal for D in the integer case
% Version of 5/21/97
% Do not forget zero coefficients for missing powers
% in the generating functions for N, Y
n  = length(gn) - 1;           % Highest power in gN
m  = length(gy) - 1;           % Highest power in gY
P  = zeros(n + 1,n*m + 1);     % Base for generating P
y  = 1;                        % Initialization
P(1,1) = gn(1);                % First row of P (P(N=0) in the first position)
for i = 1:n                    % Row by row determination of P
   y  = conv(y,gy);            % Successive powers of gy
   P(i+1,1:i*m+1) = y*gn(i+1); % Successive rows of P
end
PD = sum(P);                   % Probability for each possible value of D
D  = 0:n*m;                    % All possible values of D
b  = find(PD);                 % Location of nonzero D probabilities
d  = D(b);                     % D values with positive probabilities
pd = PD(b);                    % Positive D probabilities
```

mgd.m

```
% MGD file mgd.m  Moment generating function for compound demand
% Version of 5/19/97
% Uses m-functions csort, mgsum
disp('Do not forget zeros coefficients for missing')
disp('powers in the generating function for N')
disp(' ')
g  = input('Enter COEFFICIENTS for gN  ');
y  = input('Enter VALUES for Y  ');
p  = input('Enter PROBABILITIES for Y  ');
n  = length(g);              % Initialization
a  = 0;
b  = 1;
D  = a;
PD = g(1);
for i = 2:n
  [a,b] = mgsum(y,a,p,b);
  D  = [D a];
  PD = [PD b*g(i)];
  [D,PD] = csort(D,PD);
end
r = find(PD>1e-13);
D = D(r);                    % Values with positive probability
PD = PD(r);                  % Corresponding probabilities
mD = [D; PD]';               % Display details
disp('Values are in row matrix D; probabilities are in PD.')
disp('To view the distribution, call for mD.')
```

Exercise 17.1.1

```
function [d,pd] = mgdf(pn,y,py)
% MGDF [d,pd] = mgdf(pn,y,py)  Function version of mgD
% Version of 5/19/97
```

```
% Uses m-functions mgsum and csort
% Do not forget zeros coefficients for missing
% powers in the generating function for N
n  = length(pn);               % Initialization
a  = 0;
b  = 1;
d  = a;
pd = pn(1);
for i = 2:n
  [a,b] = mgsum(y,a,py,b);
  d  = [d a];
  pd = [pd b*pn(i)];
  [d,pd] = csort(d,pd);
end
a  = find(pd>1e-13);          % Location of positive probabilities
pd = pd(a);                   % Positive probabilities
d  = d(a);                    % D values with positive probability
```

Exercise 17.1.1

```
% RANDBERN file randbern.m   Random number of Bernoulli trials
% Version of 12/19/96; notation modified 5/20/97
% Joint and marginal distributions for a random number of Bernoulli trials
% N is the number of trials
% S is the number of successes
p  = input('Enter the probability of success  ');
N  = input('Enter VALUES of N  ');
PN = input('Enter PROBABILITIES for N  ');
n  = length(N);
m  = max(N);
S  = 0:m;
P  = zeros(n,m+1);
for i = 1:n
  P(i,1:N(i)+1) = PN(i)*ibinom(N(i),p,0:N(i));
end
PS = sum(P);
P  = rot90(P);
disp('Joint distribution N, S, P, and marginal PS')
```

Exercise 17.1.1

```
% INVENTORY1 file inventory1.m  Generates P for (m,M) inventory policy
% Version of 1/27/97
% Data for transition probability calculations
% for (m,M) inventory policy
M  = input('Enter value M of maximum stock  ');
m  = input('Enter value m of reorder point  ');
Y  = input('Enter row vector of demand values  ');
PY = input('Enter demand probabilities  ');
states = 0:M;
ms = length(states);
my = length(Y);
% Calculations for determining P
[y,s] = meshgrid(Y,states);
T  =  max(0,M-y).*(s < m) + max(0,s-y).*(s >= m);
P  = zeros(ms,ms);
for i = 1:ms
   [a,b] = meshgrid(T(i,:),states);
   P(i,:) = PY*(a==b)';
```

```
end
disp('Result is in matrix P')
```

branchp.m

```
% BRANCHP file branchp.m  Transition P for simple branching process
% Version of 7/25/95
% Calculates transition matrix for a simple branching
% process with specified maximum population.
disp('Do not forget zero probabilities for missing values of Z')
PZ = input('Enter PROBABILITIES for individuals  ');
M  = input('Enter maximum allowable population  ');
mz = length(PZ) - 1;
EZ = dot(0:mz,PZ);
disp(['The average individual propagation is ',num2str(EZ),])
P  = zeros(M+1,M+1);
Z  = zeros(M,M*mz+1);
k  = 0:M*mz;
a  = min(M,k);
z  = 1;
P(1,1) = 1;
for i = 1:M                  % Operation similar to gend
  z = conv(PZ,z);
  Z(i,1:i*mz+1) = z;
  [t,p] = csort(a,Z(i,:));
  P(i+1,:) = p;
end
disp('The transition matrix is P')
disp('To study the evolution of the process, call for branchdbn')
```

chainset.m

```
% CHAINSET file chainset.m Setup for simulating Markov chains
% Version of 1/2/96 Revise 7/31/97 for version 4.2 and 5.1
P  = input('Enter the transition matrix  ');
ms = length(P(1,:));
MS = 1:ms;
states = input('Enter the states if not 1:ms  ');
if isempty(states)
  states = MS;
end
disp('States are')
disp([MS;states]')
PI = input('Enter the long-run probabilities  ');
F  = [zeros(1,ms); cumsum(P')]';
A  = F(:,MS);
B  = F(:,MS+1);
e  = input('Enter the set of target states  ');
ne = length(e);
E  = zeros(1,ne);
for i = 1:ne
  E(i) = MS(e(i)==states);
end
disp(' ')
disp('Call for for appropriate chain generating procedure')
```

mchain.m

```
% MCHAIN file mchain.m  Simulation of Markov chains
```

```
% Version of 1/2/96  Revised 7/31/97 for version 4.2 and 5.1
% Assumes the procedure chainset has been run
n  = input('Enter the number n of stages   ');
st = input('Enter the initial state  ');
if ~isempty(st)
  s  = MS(st==states);
else
  s = 1;
end
T  = zeros(1,n);           % Trajectory in state numbers
U  = rand(1,n);
for i = 1:n
  T(i) = s;
  s = ((A(s,:) < U(i))&(U(i) <= B(s,:)))*MS';
end
N  = 0:n-1;
tr = [N;states(T)]';
n10 = min(n,11);
TR = tr(1:n10,:);
f  = ones(1,n)/n;
[sn,p] = csort(T,f);
if isempty(PI)
  disp('     State     Frac')
  disp([states; p]')
else
  disp('     State     Frac       PI')
  disp([states; p; PI]')
end
disp('To view the first part of the trajectory of states, call for TR')
```

arrival.m

```
% ARRIVAL file arrival.m  Arrival time to a set of states
% Version of 1/2/96  Revised 7/31/97 for version 4.2 and 5.1
% Calculates repeatedly the arrival
% time to a prescribed set of states.
% Assumes the procedure chainset has been run.
r  = input('Enter the number of repetitions  ');
disp('The target state set is:')
disp(e)
st = input('Enter the initial state  ');
if ~isempty(st)
  s1 = MS(st==states); % Initial state number
else
  s1 = 1;
end
clear T                % Trajectory in state numbers (reset)
S  = zeros(1,r);       % Arrival time for each rep  (reset)
TS = zeros(1,r);       % Terminal state number for each rep (reset)
for k = 1:r
  R  = zeros(1,ms);    % Indicator for target state numbers
  R(E) = ones(1,ne);   % reset for target state numbers
  s  = s1;
  T(1) = s;
  i  = 1;
  while R(s) ~= 1      % While s is not a target state number
    u = rand(1,1);
    s = ((A(s,:) < u)&(u <= B(s,:)))*MS';
```

```
    i = i+1;
    T(i) = s;
  end
  S(k) = i-1;          % i is the number of stages; i-1 is time
  TS(k) = T(i);
end
[ts,ft] = csort(TS,ones(1,r));  % ts = terminal state numbers  ft = frequencies
fts = ft/r;                     % Relative frequency of each ts
[a,at]  = csort(TS,S);          % at = arrival time for each ts
w  = at./ft;                    % Average arrival time for each ts
RES = [states(ts); fts; w]';
disp(' ')
if r == 1
  disp(['The arrival time is ',int2str(i-1),])
  disp(['The state reached is ',num2str(states(ts)),])
  N = 0:i-1;
  TR = [N;states(T)]';
  disp('To view the trajectory of states, call for TR')
else
  disp(['The result of ',int2str(r),' repetitions is:'])
  disp('Term state  Rel Freq   Av time')
  disp(RES)
  disp(' ')
  [t,f]  = csort(S,ones(1,r));  % t = arrival times   f = frequencies
  p  = f/r;                     % Relative frequency of each t
  dbn = [t; p]';
  AV = dot(t,p);
  SD = sqrt(dot(t.^2,p) - AV^2);
  MN = min(t);
  MX = max(t);
  disp(['The average arrival time is ',num2str(AV),])
  disp(['The standard deviation is ',num2str(SD),])
  disp(['The minimum arrival time is ',int2str(MN),])
  disp(['The maximum arrival time is ',int2str(MX),])
  disp('To view the distribution of arrival times, call for dbn')
  disp('To plot the arrival time distribution, call for plotdbn')
end
```

recurrence.m

```
% RECURRENCE file recurrence.m Recurrence time to a set of states
% Version of 1/2/96  Revised 7/31/97 for version 4.2 and 5.1
% Calculates repeatedly the recurrence time
% to a prescribed set of states, if initial
% state is in the set; otherwise arrival time.
% Assumes the procedure chainset has been run.
r  = input('Enter the number of repititions  ');
disp('The target state set is:')
disp(e)
st = input('Enter the initial state  ');
if ~isempty(st)
  s1 = MS(st==states); % Initial state number
else
  s1 = 1;
end
clear T               % Trajectory in state numbers (reset)
S  = zeros(1,r);      % Recurrence time for each rep  (reset)
TS = zeros(1,r);      % Terminal state number for each rep (reset)
```

```
for k = 1:r
  R  = zeros(1,ms);    % Indicator for target state numbers
  R(E) = ones(1,ne);   % reset for target state numbers
  s  = s1;
  T(1) = s;
  i  = 1;
  if R(s) == 1
    u = rand(1,1);
    s = ((A(s,:) < u)&(u <= B(s,:)))*MS';
    i = i+1;
    T(i) = s;
  end
  while R(s) ~= 1      % While s is not a target state number
    u = rand(1,1);
    s = ((A(s,:) < u)&(u <= B(s,:)))*MS';
    i = i+1;
    T(i) = s;
  end
  S(k) = i-1;          % i is the number of stages; i-1 is time
  TS(k) = T(i);
end
[ts,ft]  = csort(TS,ones(1,r)); % ts = terminal state numbers  ft = frequencies
fts = ft/r;                % Relative frequency of each ts
[a,tt]  = csort(TS,S);    % tt = total time for each ts
w  = tt./ft;               % Average time for each ts
RES = [states(ts); fts; w]';
disp(' ')
if r == 1
  disp(['The recurrence time is ',int2str(i-1),])
  disp(['The state reached is ',num2str(states(ts)),])
  N = 0:i-1;
  TR = [N;states(T)]';
  disp('To view the trajectory of state numbers, call for TR')
else
  disp(['The result of ',int2str(r),' repetitions is:'])
  disp('Term state  Rel Freq   Av time')
  disp(RES)
  disp(' ')
  [t,f]  = csort(S,ones(1,r));  % t = recurrence times   f = frequencies
  p  = f/r;                      % Relative frequency of each t
  dbn = [t; p]';
  AV = dot(t,p);
  SD = sqrt(dot(t.^2,p) - AV^2);
  MN = min(t);
  MX = max(t);
  disp(['The average recurrence time is ',num2str(AV),])
  disp(['The standard deviation is ',num2str(SD),])
  disp(['The minimum recurrence time is ',int2str(MN),])
  disp(['The maximum recurrence time is ',int2str(MX),])
  disp('To view the distribution of recurrence times, call for dbn')
  disp('To plot the recurrence time distribution, call for plotdbn')
end
```

kvis.m

```
% KVIS file kvis.m  Time to complete k visits to a set of states
% Version of 1/2/96 Revised 7/31/97 for version 4.2 and 5.1
% Calculates repeatedly the time to complete
```

```
% visits to k of the states in a prescribed set.
% Default is visit to all the target states.
% Assumes the procedure chainset has been run.
r  = input('Enter the number of repetitions  ');
disp('The target state set is:')
disp(e)
ks = input('Enter the number of target states to visit  ');
if isempty(ks)
  ks = ne;
end
if ks > ne
  ks = ne;
end
st = input('Enter the initial state  ');
if ~isempty(st)
  s1 = MS(st==states); % Initial state number
else
  s1 = 1;
end
disp(' ')
clear T                    % Trajectory in state numbers (reset)
R0 = zeros(1,ms);          % Indicator for target state numbers
R0(E) = ones(1,ne);        % reset
S = zeros(1,r);            % Terminal transitions for each rep (reset)
for k = 1:r
  R = R0;
  s  = s1;
  if R(s) == 1
    R(s) = 0;
  end
  i  = 1;
  T(1) = s;
  while sum(R) > ne - ks
    u = rand(1,1);
    s = ((A(s,:) < u)&(u <= B(s,:)))*MS';
    if R(s) == 1
      R(s) = 0;
    end
    i = i+1;
    T(i) = s;
  end
  S(k) = i-1;
end
if r == 1
  disp(['The time for completion is ',int2str(i-1),])
  N = 0:i-1;
  TR = [N;states(T)]';
  disp('To view the trajectory of states, call for TR')
else
  [t,f]  = csort(S,ones(1,r));
  p  = f/r;
  D  = [t;f]';
  AV = dot(t,p);
  SD = sqrt(dot(t.^2,p) - AV^2);
  MN = min(t);
  MX = max(t);
  disp(['The average completion time is ',num2str(AV),])
  disp(['The standard deviation is ',num2str(SD),])
```

```
  disp(['The minimum completion time is ',int2str(MN),])
  disp(['The maximum completion time is ',int2str(MX),])
  disp(' ')
  disp('To view a detailed count, call for D.')
  disp('The first column shows the various completion times;')
  disp('the second column shows the numbers of trials yielding those times')
end
```

plotdbn

```
% PLOTDBN file plotdbn.m
% Version of 1/23/98
% Plot arrival or recurrence time dbn
% Use after procedures arrival or recurrence
% to plot arrival or recurrence time distribution
plot(t,p,'-',t,p,'+')
grid
title('Time Distribution')
xlabel('Time in number of transitions')
ylabel('Relative frequency')
```

This page titled 17.1: Appendix A to Applied Probability- Directory of m-functions and m-procedures is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.2: Appendix B to Applied Probability- some mathematical aids

Series

1. **Geometric series** From the expression $(1-r)(1+r+r^2+\cdots+r^n)=1-r^{n+1}$, we obtain

$$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r} \text{ for } r \neq 1$$

For $|r|<1$, these sums converge to the geometric series $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$

Differentiation yields the following two useful series:

$$\sum_{k=1}^{\infty} kr^{k-1} = \frac{1}{(1-r)^2} \text{ for } |r|<1 \text{ and } \sum_{k=2}^{\infty} k(k-1)r^{k-2} = \frac{2}{(1-r)^3} \text{ for } |r|<1$$

For the finite sum, differentiation and algebraic manipulation yields

$\sum_{k=0}^{n} kr^{k-1} = \frac{1-r^n[1+n(1-r)]}{(1-r)^2}$ which converges to $\frac{1}{(1-r)^2}$ for $|r|<1$

2. **Exponential series**. $e^k = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ and $e^{-s} = \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!}$ for any x

Simple algebraic manipulation yields the following equalities useful for the Poisson distribution:

$$\sum_{k=n}^{\infty} k\frac{x^k}{k!} = x\sum_{k=n-1}^{\infty} \frac{x^k}{k!} \text{ and } \sum_{k=n}^{\infty} k(k-1)\frac{x^k}{k!} = x^2\sum_{k=n-2}^{\infty} x^k k!$$

3. **Sums of powers of integers** $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

Some useful integrals

1. **The gamma function** $\Gamma(r) = \int_0^{\infty} t^{r-1}e^{-t}\,dt$ for $r>0$

Integration by parts shows $Gamma(r) = (r-1)\Gamma(r-1)$ for $r>1$
By induction $\Gamma(r) = (r-1)(r-2)\cdots(r-k)\Gamma(r-k)$ for $r>k$
For a positive integer n, $\Gamma(n) = (n-1)!$ with $\Gamma(1) = 0! = 1$

2. By a change of variable in the gamma integral, we obtain

$$\int_0^{\infty} t^r e^{-\lambda t}\,dt = \frac{\Gamma(r+1)}{\lambda^{r+1}} \quad r>-1,\ \lambda>0$$

3. A well known indefinite integral gives

$$\int_a^{\infty} te^{-\lambda t}\,dt = \frac{m!}{\lambda^{m+1}}e^{-\lambda a}[1+\lambda a+\frac{(\lambda a)^2}{2!}+\cdots+\frac{(\lambda a)^m}{m!}]$$

4. The following integrals are important for the Beta distribution.

$$\int_0^1 u^r(1-u)^s\,du = \frac{\Gamma(r+1)\Gamma(s+1)}{\Gamma(r+s+2)} \quad r>-1,\ s>-1$$

For nonnegative integers m, n $\int_0^1 u^m(1-u)^n\ du = \dfrac{m!n!}{(m+n+1)!}$

Some basic counting problems

We consider three basic counting problems, which are used repeatedly as components of more complex problems. The first two, **arrangements** and **occupancy** are equivalent. The third is a basic **matching** problem.

Arrangements of r objects selected from among n distinguishable objects.
a. The order is significant.
b. The order is irrelevant.
For each of these, we consider two additional alternative conditions.
1. No element may be selected more than once.
2. Repitition is allowed.
Occupancy of n distinct cells by r objects. These objects are
a. Distinguishable.
b. Indistinguishable.
The occupancy may be
1. Exclusive.
2. Nonexclusive (i.e., more than one object per cell)

The results in the four cases may be summarized as follows:
a. 1. Ordered arrangements, without repetition (**permutations**). Distinguishable objects, exclusive occupancy.

$$P(n,r) = \frac{n!}{(n-r)!}$$

2. Ordered arrangements, with repitition allowed. Distinguishable objects, nonexclusive occupancy.

$$U(n,r) = n^r$$

b. 1. Arrangements without repetition, order irrelevant (**combinations**). Indistinguishable objects, exclusive occupancy.

$$C(n,r) = \frac{n!}{r!(n-r)!} = \frac{P(n,r)}{r!}$$

2. Unordered arrangements, with repetition. Indistinguishable objects, nonexclusive occupancy.

$$S(n,r) = C(n+r-1,r)$$

Matching n distinguishable elements to a fixed order. Let $M(n,k)$ be the number of permutations which give k matches.

$n = 5$

Natural order 1 2 3 4 5

Permutation 3 2 5 4 1 (Two matches– positions 2, 4)

We reduce the problem to determining $m(n,0)$, as follows:

Select k places for matches in $C(n,k)$ ways.
Order the $n-k$ remaining elements so that no matches in the other $n-k$ places.

$$M(n,k) = C(n,k)M(n-k,0)$$

Some algebraic trickery shows that $M(n,0)$ is the integer nearest $n!/e$. These are easily calculated by the MATLAB command `M = round(gamma(n+1)/exp(1))` For example `>> M = round(gamma([3:10]+1)/exp(1)); >> disp([3:6;M(1:4);7:10;M(5:8)]') 3 2 7 1854 4 9 8 14833 5 44 9 133496 6 265 10 1334961`

Extended binomial coefficients and the binomial series

The ordinary **binomial coefficient** is $C(n,k) = \frac{n!}{k!(n-k)!}$ for integers $n > 0$, $0 \leq k \leq n$

For any real x, any integer k, we extend the definition by

$$C(x,0) = 1,\ C(x,k) = 0 \text{ for } k < 0, \text{ and } C(n,k) = 0 \text{ for a positive integer } k > n$$

and

$$C(x,k) = \frac{x(x-1)(x-2)\cdots(x-k+1)}{k!} \text{ otherwise}$$

The **Pascal's relation** holds: $C(x,k) = C(x-1,k-1) + C(x-1,k)$
The power series expansion about $t = 0$ shows

$$(1+t)^x = 1 + C(x,1)t + C(x,2)t^2 + \cdots \forall x, -1 < t < 1$$

For $x = n$, a positive integer, the series becomes a polynomial of degree n

Cauchy's equation

Let f be a real-valued function defined on $(0,\infty)$, such that
a. $f(t+u) = f(t) + f(u)$ for $t, u > 0$, and
b. There is an open interval I on which f is bounded above (or is bounded below).
Then $f(t) = f(1)t\ \forall t > 0$
Let f be a real-valued function defined on $(0,\infty)$ such that
a. $f(t+u) = f(t)f(u)\ \forall t, u > 0$, and
b. There is an interval on which f is bounded above.
Then, either $f(t) = 0$ for $t > 0$, or there is a constant a such that \f(t) = e^{at}\) for $t > 0$

[For a proof, see Billingsley, *Probability and Measure*, second edition, appendix A20]

Countable and uncountable sets

A set (or class) is *countable* iff either it is finite or its members can be put into a one-to-one correspondence with the natural numbers.

Examples

- The set of odd integers is countable.
- Thee finite set $\{n : 1 \leq n \leq 1000\}$ is countable.
- The set of all rational numbers is countable. (This is established by an argument known as diagonalization).
- The set of pairs of elements from two countable sets is countable.
- The union of a countable class of countable sets is countable.

A set is *uncountable* iff it is neither finite nor can be put into a one-to-one correspondence with the natural numbers.

Examples

- The class of positive real numbers is uncountable. A well known operation shows that the assumption of countability leads to a contradiction.
- The set of real numbers in any finite interval is uncountable, since these can be put into a one-to-one correspondence of the class of all positive reals.

This page titled 17.2: Appendix B to Applied Probability- some mathematical aids is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.3: Appendix C- Data on some common distributions

Discrete distributions

Indicator function $X = I_E$ $P(X = 1) = P(E) = p$ $P(X = 0) = q = 1 - p$

$$E[X] = p \text{ Var}[X] = pq \; M_X(s) = q + pe^s \; g_X(s) = q + ps$$

Simple random variable $X = \sum_{i=1}^{n} t_i I_{A_i}$ (a primitive form) $P(A_i) = p_i$

$$E[X] = \sum_{i=1}^{n} t_i p_i \text{ Var}[X] = \sum_{i=1}^{n} t_i^2 p_i q_i - 2\sum_{i<j} t_i t_j p_i p_j \; M_X(s) = \sum_{i=1}^{n} p_i e^{st_i}$$

Binomial$(n, p)X = \sum_{i=1}^{n} I_{E_i}$ with $\{I_{E_i} : 1 \le i \le n\}$ iid $P(E_i) = p$

$$P(X = k) = C(n, k)p^k q^{n-k}$$

$$E[X] = np \text{ Var}[X] = npq \; M_X(s) = (q + pe^s)^n \; g_X(s) = (q + ps)^n$$

MATLAB: $P(X = k) = \text{ibinom}(n, p, k)$ $P(X \ge k) = \text{cbinom}(n, p, k)$

Geometric$(p) P(X = k) = pq^k \; \forall k \ge 0$

$$E[X] = q/p \text{ Var}[X] = q/p^2 \; M_X(s) = dfracp1 - qe^s \; g_X(s) = \frac{p}{1 - qs}$$

If $Y - 1$ ~ geometric (p), so that $P(Y = k) = pq^{k-1} \; \forall k \ge 1$, then

$$E[Y] = 1/p \text{ Var}[X] = q/p^2 \; M_Y(s) = \frac{pe^s}{1 - qe^s} \; g_Y(s) = \frac{ps}{1 - qs}$$

Negative binomial(m, p), X is the number of failures before the mth success.

$P(X = k) = C(m + k - 1, m - 1)p^m q^k \; \forall k \ge 0$

$$E[X] = mq/p \text{ Var}[X] = mq/p^2 \; M_X(s) = (\frac{p}{1 - qe^s})^m \; g_X(s) = (\frac{p}{1 - qs})^m$$

For $Y_m = X_m + m$, the number of the trial on which mth success occurs. $P(Y = k) = C(k - 1, m - 1)p^m q^{k-m} \; \forall k \ge m$.

$$E[Y] = m/p \text{ Var}[Y] = mq/p^2 \; M_Y(s) = (\frac{pe^s}{1 - qe^s})^m \; g_Y(s) = (\frac{ps}{1 - qs})^m$$

MATLAB: $P(Y = k) = \text{nbinom}(m, p, k)$

Poisson(μ). $P(X = k) = e^{-\mu}\frac{\mu^k}{k!} \; \forall k \ge 0$

$$E[X] = \mu \text{ Var}[X] = \mu \; M_X(s) = e^{\mu(e^s - 1)} \; g_X(s) = e^{\mu(s-1)}$$

MATLAB: $P(X = k) = \text{ipoisson}(m, k)$ $P(X \ge k) = \text{cpoisson}(m, k)$

Absolutely continuous distributions

Uniform(a, b) $f_x(t) = \frac{1}{b - a}$ $a < t < b$ (zero elsewhere)

$$E[X] = \frac{b + a}{2} \text{ Var}[X] = \frac{(b - a)^2}{12} \; M_X(s) = \frac{e^{sb} - e^{sa}}{s(b - a)}$$

Symmetric triangular $(-a, a)$ $f_X(t) = \begin{cases} (a+t)/a^2 & -a \le t < 0 \\ (a-t)/a^2 & 0 \le t \le a \end{cases}$

$$E[X] = 0 \ \mathrm{Var}[X] = \frac{a^2}{6} \ M_X(s) = \frac{e^{as} + e^{-as} - 2}{a^2 s^2} = \frac{e^{as} - 1}{as} \cdot \frac{1 - e^{-as}}{as}$$

Exponential$(\lambda) f_X(t) = \lambda e^{-\lambda t} \ t \ge 0$

$$E[X] = \frac{1}{\lambda} \ \mathrm{Var}[X] = \frac{1}{\lambda^2} \ M_X(s) = \frac{\lambda}{\lambda - s}$$

Gamma$(\alpha, \lambda) f_X(t) = \dfrac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} \ t \ge 0$

$$E[X] = \frac{\alpha}{\lambda} \ \mathrm{Var}[X] = \frac{\alpha}{\lambda^2} \ M_X(s) = (\frac{\lambda}{\lambda - s})^\alpha$$

MATLAB: $P(X \le t) = \mathrm{gammadbn}(\alpha, \lambda, t)$

Normal$N(\mu, \sigma^2) f_X(t) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp(-\dfrac{1}{2}(\dfrac{t-\mu}{\sigma})^2)$

$$E[X] = \mu \ \mathrm{Var}[X] \sigma^2 \ M_X(s) = \exp(\frac{\sigma^2 s^2}{2} + \mu s)$$

MATLAB: $P(X \le t) = \mathrm{gaussian}(\mu, \sigma^2, t)$

Beta(r, s)

$$f_X(t) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} t^{r-1}(1-t)^{s-1} \ 0 < t < 1, \, r > 0, \, s > 0$$

$$E[X] = \frac{r}{r+s} \ \mathrm{Var}[X] = \frac{rs}{(r+s)^2(r+s+1)}$$

MATLAB: $f_X(t) = \mathrm{beta}(r, s, t) \ P(X \le t) = \mathrm{betadbn}(r, s, t)$

Weibull(α, λ, ν)

$$F_X(t) = 1 - e^{-\lambda(t-\nu)^\alpha}, \ \alpha > 0, \lambda > 0, \nu \ge 0, t \ge \nu$$

$$E[X] = \frac{1}{\lambda^{1/\alpha}} \Gamma(1 + 1/\alpha) + \nu \ \mathrm{Var}[X] = \frac{1}{\lambda^{2/\alpha}} [\Gamma(1 + 2/\lambda) - \Gamma^2(1 + 1/\lambda)]$$

MATLAB: ($\nu = 0$ only)

$$f_X(t) = \mathrm{weibull}(a, l, t) \ P(X \le t) = \mathrm{weibull}(a, l, t)$$

Relationship between gamma and Poisson distributions

- If X ~ gamma (n, λ), then $P(X \le t) = P(Y \ge n)$ where Y ~ Poisson (λt).
- If Y ~ Poisson (λt), then $P(Y \ge n) = P(X \le t)$ where X ~ gamma (n, λ).

This page titled 17.3: Appendix C- Data on some common distributions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.4: Appendix D to Applied Probability- The standard normal distribution

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\mu^2/2}\, dt \quad \phi(-t) = 1 - \phi(t)$$

t	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7643	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989

This page titled 17.4: Appendix D to Applied Probability- The standard normal distribution is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.5: Appendix E to Applied Probability - Properties of Mathematical Expectation

$$E[g(X)] = \int g(X)\ dP$$

We suppose, without repeated assertion, that the random variables and Borel functions of random variables or random vectors are integrable. Use of an expression such as $I_M(X)$ involves the tacit assumption that M is a Borel set on the codomain of X.
(E1): $E[aI_A] = aP(A)$, any constant a, any event A
(E1a): $E[I_M(X)] = P(X \in M)$ and $E[I_M(X)I_N(Y)] - P(X \in M, Y \in N)$ for any Borel sets M, N (Extends to any finite product of such indicator functions of random vectors)
(E2): **Linearity**. For any constants a, b, $E[aX + bY) = aE[X] + bE[Y]$ (Extends to any finite linear combination)
(E3): **Positivity; monotonicity.**
a. $X \geq 0$ a.s. implies $E[X] \geq 0$, with equality iff $X = 0$ a.s.
b. $X \geq Y$ a.s. implies $E[X] \geq E[Y]$, with equality iff $X = Y$ a.s.
(E4): **Fundamental lemma**. If $X \geq 0$ is bounded, and $\{X_n : 1 \leq n\}$ is a.s. nonnegative, nondecreasing, with $\lim_n X_n(\omega) \geq X(\omega)$ for a.e. ω, then $\lim_n E[X_n] \geq E[X]$
(E4a): **Monotone convergence**. If for all n, $0 \leq X_n \leq X_{n+1}$ a.s. and $X_n \to X$ a.s.,then $E[X_n] \to E[X]$ (The theorem also holds if $E[X] = \infty$)

******
(E5): **Uniqueness**. * is to be read as one of the symbols $\leq, =$, or $\geq$
a. $E[I_M(X)g(X)]$ * $E[I_M(X)h(X)]$ for all M iff $g(X)$ * $h(X)$ a.s.
b. $E[I_M(X)I_N(Z)g(X,Z)] = E[I_M(X)I_N(Z)h(X,Z)]$ for all M, N iff $g(X,Z) = h(X,Z)$ a.s.
(E6): Fatou's lemma. If $X_n \geq 0$ a.s., for all n, then $E[\liminf X_n] \leq [\liminf E[X_n]$
(E7): Dominated convergence. If real or complex $X_n \to X$ a.s., $|X_n| \leq Y$ a.s. for all n, and Y is integrable, then $\lim_n E[X_n] = E[X]$
(E8): **Countable additivity and countable sums**.
a. If X is integrable over E, and $E = \bigvee_{i=1}^{\infty} E_i$ (disjoint union), then $E[I_E X] = \sum_{i=1}^{\infty} E[I_{E_i} X]$
b. If $\sum_{n=1}^{\infty} E[|X_n|] < \infty$, then $\sum_{n=1}^{\infty} |X_n| < \infty$, a.s. and $E[\sum_{n=1}^{\infty} X_n] = \sum_{n=1}^{\infty} E[X_n]$
(E9): **Some integrability conditions**
a. X is integrable iff both X^+ and X^- are integrable iff $|X|$ is integrable.
b. X is integrable iff $E[I_{\{|X|>a\}}|X|] \to 0$ as $a \to \infty$
c. If X is integrable, then X is a.s. finite
d. If $E[X]$ exists and $P(A) = 0$, then $E[I_A X] = 0$
(E10): **Triangle inequality**. For integrable X, real or complex, $|E[X]| \leq E[|X|]$
(E11): **Mean-value theorem**. If $a \leq X \leq b$ a.s. on A, then $aP(A) \leq E[I_A X] \leq bP(A)$
(E12): For nonnegative, Borel g, $E[g(X)] \geq aP(g(X) \geq a)$
(E13): **Markov's inequality**. If $g \geq 0$ and nondecreasing for $t \geq 0$ and $a \geq 0$, then

$$g(a)P(|X| \geq a) \leq E[g(|X|)]$$

(E14): **Jensen's inequality**. If g is convex on an interval which contains the range of random variable X, then $g(E[X]) \leq E[g(X)]$
(E15): **Schwarz' inequality**. For X, Y real or complex, $|E[XY]|^2 \leq E[|X|^2]E[|Y|^2]$, with equality iff there is a constant c such that $X = cY$ a.s.
(E16): **Hölder's inequality**. For $1 \leq p, q$, with $\frac{1}{p} + \frac{1}{q} = 1$, and X, Y real or complex.

$$E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q}$$

(E17): **Hölder's inequality**. For $1 < p$ and X, Y real or complex,

$$E[|X+Y|^p]^{1/p} \leq E[|X|^p]^{1/p} + E[|Y|^p]^{1/p}$$

(E18): **Independence and expectation**. The following conditions are equivalent.
a. The pair $\{X, Y\}$ is independent
b. $E[I_M(X)I_N(Y)] = E[I_M(X)]E[I_N(Y)]$ for all Borel M, N

c. $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for all Borel g, h such that $g(X)$, $h(Y)$ are integrable.
(E19): **Special case of the Radon-Nikodym theorem** If $g(Y)$ is integrable and X is a random vector, then there exists a real-valued Borel function $e(\cdot)$, defined on the range of X, unique a.s. $[P_X]$, such that $E[I_M(X)g(X)] = E[I_M(X)e(X)]$ for all Borel sets M on the codomain of X.
(E20): **Some special forms of expectation**
a. Suppose F is nondecreasing, right-continuous on $[0, \infty)$, with $F(0^-) = 0$. Let $F^*(t) = F(t-0)$. Consider $X \geq 0$ with $E[F(X)] < \infty$. Then,

$$\text{(1) } E[F(X)] = \int_0^\infty P(X \geq t) F\ (dt) \text{ and (2) } E[F^*(X)] = \int_0^\infty P(X > t) F\ (dt)$$

b. If X is integrable, then $E[X] = \int_{-\infty}^{\infty} [u(t) - F_X(t)]\ dt$
c. If X, Y are integrable, then $E[X - Y] = \int_{-\infty}^{\infty} [F_Y(t) - F_X(t)]\ dt$
d. if $X \geq 0$ is integrable, then

$$\textstyle\sum_{n=0}^{\infty} P(X \geq n+1) \leq E[X] \leq \sum_{n=0}^{\infty} P(X \geq n) \leq N \sum_{k=0}^{\infty} P(X \geq kN), \text{ for all } N \geq 1$$

e. If integrable $X \geq 0$ is integer-valued, then

$$E[X] = \textstyle\sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} P(X > n) E[X^2] = \sum_{n=1}^{\infty} (2n-1) P(X \geq n) = \sum_{n=0}^{\infty} (2n+1) P(X > n)$$

f. If Q is the quantile function for F_X, then $E[g(X)] = \int_0^1 g[Q(u)]\ du$

This page titled 17.5: Appendix E to Applied Probability - Properties of Mathematical Expectation is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.6: Appendix F to Applied Probability- Properties of conditional expectation, given a random vector

We suppose, without repeated assertion, that the random variables and functions of random vectors are integrable, as needed.
(CE1): **Defining condition**. $e(X) = E[g(Y)|X]$ a.s. iff $E[I_M(X)g(Y)] = E[I_M(X)e(X)]$ for each Borel set M on the codomain of X.
(CE1a): If $P(X \in M) > 0$, then $E[I_M(X)e(X)] = E[g(Y)|X \in M]P(X \in M)$
(CE1b): Law of total probability. $E[g(Y)] = E\{[g(Y)|X]\}$
(CE2): **Linearity**. For any constants a, b
$E[ag(Y) + bh(Z)|X] = aE[g(Y)|X] + bE[h(Z)|X]$ a.s.
(Extends to any finite linear combination)
(CE3): positivity; monotonicty.
a. $g(Y) \geq 0$ a.s. implies $E[g(Y)|X] \geq 0$ a.s.
b. $g(Y) \geq h(Z)$ a.s. implies $E[g(Y)|X] \geq E[h(Z)|X]$ a.s.
(CE4): **Monotone convergence**. $Y_n \to Y$ a.s. monotonically implies $E[Y_n|X] \to E[Y|X]$ a.s.
(CE5): **Independence**. $\{X, Y\}$ is an independent pair
a. iff $E[g(Y)|X] = E[g(Y)]$ a.s. for all Borel functions g
b. iff $E[I_N(Y)|X] = E[I_N(Y)]$ a.s. for all Borel sets N on the codomain of Y
(CE6): $e(X) = E[g(Y)|X]$ a.s. iff $E[h(X)g(Y)] = E[h(X)e(X)]$ a.s. for any Borel function h
(CE7): $E[h(X)|X] = h(X)$ a.s. for any Borel function h
(CE8): $E[h(X)g(Y)|X] = h(X)E[g(Y)|X]$ a.s. for any Borel function h
(CE9): If $X = h(W)$ and $W = k(X)$, with h, k Borel functions, then $E[g(Y)|X] = E[g(Y)|W]$ a.s.
(CE10): If $g$is a Borel function such that $E[g(t, Y)]$ is finite for all t on the range of X and $E[g(X, Y)]$ is finite, then
a. $E[g(X, Y)|X = t] = E[g(t, Y)|X = t]$ a.s. $[P_X]$
b. If $\{X, Y\}$ is independent, then $E[g(X, Y)|X = t] = E[g(t, Y)]$ a.s. $[P_X]$
(CE11): Suppose $\{X(t) : t \in T\}$ is a real-valued measurable random process whose parameter set T is a Borel subset of the real line and S is a random variable whose range is a subset of T, so that $X(S)$ is a random variable.
If $E[X(t)]$ is finite for all t in T and $E[X(S)]$ is finite, then
a. \9E[X(S)|S = t] = E[X(t)|S = t]\) a.s $[P_S]$
b. If, in addition, $\{S, X_T\}$ is independent, then $E[X(S)|S = t] = E[X(t)]$ a.s. $[P_S]$
(CE12): **Countable additivity and countable sums**.
a. If Y is integrable on A and $A = \bigvee_{n=1}^{\infty} A_n$.
then $E[I_A Y|X] = \sum_{n=1}^{\infty} E[I_A Y|X]$ a.s.
b. If $\sum_{n=1}^{\infty} E[|Y_n|] < \infty$, thne $E[\sum_{n=1}^{\infty} Y_n|X]$ a.s.
(CE13): **Triangle inequality**. $|E[g(Y)|X]| \leq E[|g(Y)||X]$ a.s.
(CE14): **Jensen's inequality**. If g is a convex function on an interval I which contains the range of a real random variable Y, then $g\{E[Y|X]\} \leq E[g(Y)|X]$ a.s.
(CE15): Suppose $E[|Y|^p] < \infty$ and $E[|Z|^p] < \infty$ for $1 \leq p < \infty$. Then $E\{|E[Y|X] - E[Z|X]|^p\} \leq E[|Y - Z|^p] < \infty$

This page titled 17.6: Appendix F to Applied Probability- Properties of conditional expectation, given a random vector is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.7: Appendix G to Applied Probability- Properties of conditional independence, given a random vector

Definition

The pair $\{X, Y\}$ is *conditionally independent, givenZ*, denoted $\{X, Y\}$ ci $|Z$ iff

$$E[I_M(X)I_N(Y)|Z] = E[I_M(X)|Z]E[I_N(Y)|Z] \text{ a.s. for all Borel sets } M, N$$

An arbitrary class $\{X_t : t \in T\}$ of random vectors is conditionally independent, give Z, iff such a product rule holds for each finite subclass or two or more members of the class.

Remark. The expression "for all Borel sets M, N," here and elsewhere, implies the sets are on the appropriate codomains. Also, the expressions below "for all Borel functions g," etc., imply that the functions are real-valued, such that the indicated expectations are finite.

The following are equivalent. Each is necessary and sufficient that $\{X, Y\}$ ci $|Z$.

(CI1): $E[I_M(X)I_N(Y)|Z] = E[I_M(X)|Z]E[I_N(Y)|Z]$ a.s. for all Borel sets M, N
(CI2): $E[I_M(X)|Z, Y] = E[I_M(X)|Z]$ a.s. for all Borel sets M
(CI3): $E[I_M(X)I_Q(Z)|Z, Y] = E[I_M(X)I_Q(Z)|Z]$ a.s. for all Borel sets M, Q
(CI4): $E[I_M(X)I_Q(Z)|Y] = E\{E[I_M(X)I_Q(Z)|Z]|Y\}$ a.s. for all Borel sets M, Q

****
(CI5): $E[g(X, Z)h(Y, Z)|Z] = E[g(X, Z)|Z]E[h(Y, Z)|Z]$ a.s. for all Borel functions g, h
(CI6): $E[g(X, Z)|Z, Y] = E[g(X, Z)|Z]$ a.s. for all Borel function g
(CI7): For any Borel function g, there exists a Borel function e_g such that

$$E[g(X, Z)|Z, Y] = e_g(Z) \text{ a.s.}$$

(CI8): $E[g(X, Z)|Y] = E\{E[g(X, Z)|Z]|Y\}$ a.s. for all Borel functions g

****
(CI9): $\{U, V\}$ ci $|Z$, where $U = g(X, Z)$ and $V = h(Y, Z)$, for any Borel functions g, h.

Additional properties of conditional independence

(CI10): $\{X, Y\}$ ci $|Z$ implies $\{X, Y\}$ ci $|(Z, U)$, $\{X, Y\}$ ci $|(Z, V)$, and $\{X, Y\}$ ci $|(Z, U, V)$, where $U = h(X)$ and $V = k(Y)$, with h, k Borel.
(CI11): $\{X, Z\}$ ci $|Y$ and $\{X, W\}$ ci $|(Y, Z)$ iff $\{X, (Z, W)\}$ ci $|Y$.
(CI12): $\{X, Z\}$ ci $|Y$ and $\{(X, Y), W\}$ ci $|Z$ implies $\{X, (Z, W)\}$ is independent.
(CI13): $\{X, Y\}$ is independent and $\{X, Y\}$ ci $|Y$ iff $\{X, (Y, Z)\}$ is independent.
(CI14): $\{X, Y\}$ ci $|Z$ implies $E[g(X, Y)|Y = u, Z = v] = E[g(X, u)|Z = v]$ a.s. $[P_{YZ}]$
(CI15): $\{X, Y\}$ ci $|Z$ implies
a. $E[g(X, Z)h(Y, Z)] = E\{E[g(X, Z)|Z]E[h(Y, Z)|Z]\} = E[e_1(Z)e_2(Z)]$
b. $E[g(Y)|X \in M]P(X \in M) = E\{E[I_M(X)|Z]E[g(Y)|Z]\}$
(CI16): $\{(X, Y), Z\}$ ci $|W$ iff $E[I_M(X)I_N(Y)I_Q(Z)|W] = E[I_M(X)I_N(Y)|W]E[I_Q(Z)|W]$ a.s. for all Borel sets M, N, Q

This page titled 17.7: Appendix G to Applied Probability- Properties of conditional independence, given a random vector is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

17.8: Matlab files for "Problems" in "Applied Probability"

npr02_04

```
% file npr02_04.m
% Data for problem P2-4
pm = [0.0168  0.0392  0.0672  0.1568  0.0072  0.0168  0.0288  0.0672 ...
      0.0252  0.0588  0.1008  0.2352  0.0108  0.0252  0.0432  0.1008];
disp('Minterm probabilities are in pm.  Use mintable(4)')
```

npr02_05

```
% file npr02_05.m
% Data for problem P2-5
pm = [0.0216  0.0144  0.0504  0.0336  0.0324  0.0216  0.0756  0.0504  0.0216 ...
      0.0144  0.0504  0.0336  0.0324  0.0216  0.0756  0.0504  0.0144  0.0096 ...
      0.0336  0.0224  0.0216  0.0144  0.0504  0.0336  0.0144  0.0096  0.0336 ...
      0.0224  0.0216  0.0144  0.0504  0.0336];
disp('Minterm probabilities are in pm.  Use mintable(5)')
```

npr02_06

```
% file npr02_06.m
% Data for problem P2-6
minvec3
DV = [A|Ac; A|(Bc&C); A&C; Ac&B; Ac&Cc; B&Cc];
DP = [1      0.65     0.20 0.25  0.25   0.30];
TV = [((A&Cc)|(Ac&C))&Bc; ((A&Bc)|Ac)&Cc; Ac&(B|Cc)];
disp('Call for mincalc')
```

npr02_07

```
% file npr02_07.m
% Data for problem P2-7
minvec3
DV = [A|Ac; ((A&Bc)|(Ac&B))&C; A&B; Ac&Cc;  A;  C; A&Bc&Cc];
DP = [ 1        0.4            0.2   0.3   0.6 0.5   0.1];
TV = [(Ac&Cc)|(A&C); ((A&Bc)|Ac)&Cc; Ac&(B|Cc)];
disp('Call for mincalc')
```

npr02_08

```
% file npr02_08.m
% Data for problem P2-8
minvec3
DV = [A|Ac; A;  C;  A&C; Ac&B; Ac&Bc&Cc];
DP = [ 1   0.6 0.4  0.3  0.2     0.1];
TV = [(A|B)&Cc; (A&Cc)|(Ac&C); (A&Cc)|(Ac&B)];
```

```
disp('Call for mincalc')
```

npr02_09

```
% file npr02_09.m
% Data for problem P2-9
minvec3
DV = [A|Ac;  A; A&B; A&C; A&B&Cc];
DP = [ 1    0.5 0.3  0.3   0.1];
TV = [A&(~(B&Cc)); (A&B)|(A&C)|(B&C)];
disp('Call for mincalc')

% Modification for part 2
% DV = [DV; Ac&Bc&Cc; Ac&B&C];
% DP = [DP 0.1 0.05];
```

npr02_10

```
% file npr02_10.m
% Data for problem P2-10
minvec4
DV = [A|Ac;  A;  Ac&Bc; A&Cc; A&C&Dc];
DP = [1     0.6  0.2    0.4    0.1];
TV = [(Ac&B)|(A&(Cc|D))];
disp('Call for mincalc')
```

npr02_11

```
% file npr02_11.m
% Data for problem P2-11
% A = male;  B = on campus;  C = active in sports
minvec3
DV = [A|Ac;  A;   B;  A|C; B&Cc; A&B&C; A&Bc; A&Cc];
DP = [ 1    0.52 0.85 0.78 0.30  0.32   0.08 0.17];
TV = [A&B; A&B&Cc; Ac&C];
disp('Call for mincalc')
```

npr02_12

```
% file npr02_12.m
% Data for problem P2-12
% A = male;  B = party member; C = voted last election
minvec3
DV = [A|Ac;  A;  A&Bc;  B;  Bc&C; Ac&Bc&C];
DP = [  1   0.60 0.30  0.50 0.20  0.10];
TV = [Bc&Cc];
disp('Call for mincalc')
```

npr02_13

```
% file npr02_13.m
% Data for problem P2-13
% A = rain in Austin;  B = rain in Houston;
% C = rain in San Antonio
minvec3
DV = [A|Ac; A&B; A&Bc; A&C; (A&Bc)|(Ac&B); B&C; Bc&C; Ac&Bc&Cc];
DP = [  1   0.35 0.15  0.20    0.45        0.30 0.05   0.15];
TV = [A&B&C; (A&B&Cc)|(A&Bc&C)|(Ac&B&C); (A&Bc&Cc)|(Ac&B&Cc)|(Ac&Bc&C)];
disp('Call for mincalc')
```

npr02_14

```
% file npr02_14.m
% Data for problem P2-14
% A = male;  B = engineering;
% C = foreign language; D = graduate study
minvec4
DV = [A|Ac; A; B; Ac&B; C; Ac&C; A&D; Ac&D; A&B&D; ...
      Ac&B&D; B&C&D; Bc&Cc&D; Ac&Bc&C&D];
DP = [1 0.55 0.23 0.10 0.75 0.45 0.26 0.19 0.13 0.08 0.20 0.05 0.11];
TV = [C&D; Ac&Dc; A&((C&Dc)|(Cc&D))];
disp('Call for mincalc')
```

npr02_15

```
% file npr02_15.m
% Data for problem P2-15
% A = men; B = on campus; C = readers; D = active
minvec4
DV = [A|Ac; A;  B;  Ac&B;  C;  Ac&C;  D;  B&D; C&D; ...
     Ac&B&D; Ac&Bc&D; Ac&B&C&D; Ac&Bc&C&D; A&Bc&Cc&D];
DP = [1  0.6 0.55 0.25 0.40 0.25 0.70 0.50 0.35 0.25 0.05 0.10 0.05 0.05];
TV = [A&D&(Cc|Bc); A&Dc&Cc];
disp('Call for mincalc')
```

npr02_16

```
% file npr02_16.m
% Data for problem P2-16
minvec3
DV = [A|Ac; A;    B;    C; (A&B)|(A&C)|(B&C); A&B&C; A&C; (A&B)-2*(B&C)];
DP = [ 1  0.221 0.209 0.112   0.197           0.045  0.062      0];
TV = [A|B|C; (A&Bc&Cc)|(Ac&B&Cc)|(Ac&Bc&C)];
disp('Call for mincalc')
```

npr02_17

```
% file npr02_17.m
% Data for problem P2-17
```

```
% A = alignment;  B = brake work;  C = headlight
minvec3
DV = [A|Ac; A&B&C; (A&B)|(A&C)|(B&C); B&C;    A  ];
DP = [ 1    0.100      0.325          0.125 0.550];
TV = [A&Bc&Cc; Ac&(~(B&C))];
disp('Call for mincalc')
```

npr02_18

```
% file npr02_18.m
% Date for problem P2-18
minvec3
DV = [A|Ac; A&(B|C); Ac; Ac&Bc&Cc];
DP = [ 1     0.3     0.6    0.1];
TV = [B|C; (((A&B)|(Ac&Bc))&Cc)|(A&C); Ac&(B|Cc)];
disp('Call for mincalc')

% Modification
% DV = [DV; Ac&B&C; Ac&B];
% DP = [DP   0.2     0.3];
```

npr02_19

```
% file npr02_19.m
% Data for problem P2-19
% A = computer;  B = monitor;  C = printer
minvec3
DV = [A|Ac; A&B; A&B&Cc; A&C; B&C; (A&Cc)|(Ac&C); ...
          (A&Bc)|(Ac&B); (B&Cc)|(Bc&C)];
DP = [1 0.49 0.17 0.45 0.39 0.50 0.43 0.43];
TV = [A; B; C; (A&B&Cc)|(A&Bc&C)|(Ac&B&C); (A&B)|(A&C)|(B&C); A&B&C];
disp('Call for mincalc')
```

npr02_20

```
% file npr02_20.m
% Data for problem P2-20
minvec3
DV = [A|Ac; A;     B;  A&B&C; A&C; (A&B)|(A&C)|(B&C); B&C - 2*(A&C)];
DP = [  1  0.232 0.228 0.045 0.062      0.197            0];
TV = [A|B|C; Ac&Bc&C];
disp('Call for mincalc')
% Modification
% DV = [DV; C];
% DP = [DP  0.230 ];
```

npr02_21

```
% file npr02_21.m
% Data for problem P2-21
minvec3
DV = [A|Ac; A;  A&B; A&B&C;  C;  Ac&Cc];
```

```
DP = [ 1   0.4  0.3  0.25   0.65  0.3 ];
TV = [(A&Cc)|(Ac&C); Ac&Bc; A|B; A&Bc];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&Cc; Ac&Bc];
% DP = [DP   0.1      0.3 ];
```

npr02_22

```
% file npr02_22.m
% Data for problem P2-22
minvec3
DV = [A|Ac; A;  A&B; A&B&C;  C;  Ac&Cc];
DP = [ 1   0.4  0.5  0.25   0.65  0.3 ];
TV = [(A&Cc)|(Ac&C); Ac&Bc; A|B; A&Bc];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&Cc; Ac&Bc];
% DP = [DP   0.1      0.3 ];
```

npr02_23

```
% file npr02_23.m
% Data for problem P2-23
minvec3
DV = [A|Ac; A;  A&C; A&B&C;  C;  Ac&Cc];
DP = [ 1   0.4  0.3  0.25   0.65  0.3 ];
TV = [(A&Cc)|(Ac&C); Ac&Bc; A|B; A&Bc];
disp('Call for mincalc')
% Modification
% DV = [DV; Ac&B&Cc; Ac&Bc];
% DP = [DP   0.1      0.3 ];
```

npr03_01

```
% file npr03_01.m
% Data for problem P3-1
minvec3
DV = [A|Ac; A;  A&B; B&C; Ac|(B&C); Ac&B&Cc];
DP = [ 1   0.55 0.30 0.20   0.55     0.15  ];
TV = [Ac&B; B];
disp('Call for mincalc')
```

npr04_04

```
% file npr04_04.m
% Data for problem P4-4
pm = [0.032 0.016 0.376 0.011 0.364 0.073 0.077 0.051];
disp('Minterm probabilities for P4-4 are in pm')
```

npr04_05

```
% file npr04_05.m
% Data for problem P4-5
pm = [0.084 0.196 0.036 0.084 0.085 0.196 0.035 0.084 ...
          0.021 0.049 0.009 0.021 0.020 0.049 0.010 0.021];
disp('Minterm probabilities for P4-5 are in pm')
```

npr04_06

```
% file npr04_06.m
% Data for problem P4-6
pm =  [0.085 0.195 0.035 0.085 0.080 0.200 0.035 0.085 ...
        0.020 0.050 0.010 0.020 0.020 0.050 0.015 0.015];
disp('Minterm probabilities for P4-6 are in pm')
```

mpr05_16

```
% file mpr05_16.m
% Data for Problem P5-16
A = [51 26  7; 42 32 10; 19 54 11; 24 53  7; 27 52  5;
     49 19 16; 16 59  9; 47 32  5; 55 17 12; 24 53  7];
B = [27 34  5; 19 43  4; 39 22  5; 38 19  9; 28 33  5;
     19 41  6; 37 21  8; 19 42  5; 27 33  6; 39 21  6];
disp('Call for oddsdf')
```

npr05_17

```
% file npr05_17.m
% Data for problem P5-17
PG1 = 84/150;
PG2 = 66/125;
A = [0.61 0.31 0.08
     0.50 0.38 0.12
     0.23 0.64 0.13
     0.29 0.63 0.08
     0.32 0.62 0.06
     0.58 0.23 0.19
     0.19 0.70 0.11
     0.56 0.38 0.06
     0.65 0.20 0.15
     0.29 0.63 0.08];
B = [0.41 0.51 0.08
     0.29 0.65 0.06
     0.59 0.33 0.08
     0.57 0.29 0.14
     0.42 0.50 0.08
     0.29 0.62 0.09
     0.56 0.32 0.12
     0.29 0.64 0.08
     0.41 0.50 0.09
     0.59 0.32 0.09];
disp('Call for oddsdp')
```

npr06_10

```
% file npr06_10.m
% Data for problem P6-10
pm = [ 0.072 0.048 0.018 0.012 0.168 0.112 0.042 0.028 ...
       0.062 0.048 0.028 0.010 0.170 0.110 0.040 0.032];
c  = [-5.3 -2.5 2.3 4.2 -3.7];
disp('Minterm probabilities are in pm, coefficients in c')
```

npr06_12

```
% file npr06_12.m
% Data for problem P6-12
pm = 0.001*[5 7 6 8 9 14 22 33 21 32 50 75 86 129 201 302];
c = [1 1 1 1 0];
disp('Minterm probabilities in pm, coefficients in c')
```

npr06_18.m

```
% file npr06_18.m
% Data for problem P6-18
cx = [5 17 21 8 15 0];
cy = [8 15 12 18 15 12 0];
pmx = minprob(0.01*[37 22 38 81 63]);
pmy = minprob(0.01*[77 52 23 41 83 58]);
disp('Data in cx, cy, pmx, pmy')
```

npr07_01

```
\begin{verbatim}
% file npr07_01.m
% Data for problem P7-1
T  = [1 3 2 3 4 2 1 3 5 2];
pc = 0.01*[ 8 13  6  9 14 11 12  7 11  9];
disp('Data are in T and pc')
\end{verbatim}
```

npr07_02

```
% file npr07_02.m
% Data for problem P7-2
T = [3.5 5.0 3.5 7.5 5.0 5.0 3.5 7.5];
pc = 0.01*[10 15 15 20 10  5 10 15];
disp('Data are in T, pc')
```

npr08_01

```
% file npr08_01.m
% Solution for problem P8-1
X = 0:2;
```

```
Y = 0:2;
Pn = [132  24   0; 864 144  6; 1260 216 6];
P = Pn/(52*51);
disp('Data in Pn, P, X, Y')
```

npr08_02

```
% file npr08_02.m
% Solution for problem P8-2
X = 0:2;
Y = 0:2;
Pn = [6 0 0; 18 12 0; 6 12 2];
P = Pn/56;
disp('Data are in X, Y,Pn, P')
```

npr08_03

```
% file npr08_03.m
% Solution for problem  P8-3
X = 1:6;
Y = 0:6;
P0 = zeros(6,7);       % Initialize
for i = 1:6            % Calculate rows of Y probabilities
    P0(i,1:i+1) = (1/6)*ibinom(i,1/2,0:i);
end
P = rot90(P0);         % Rotate to orient as on the plane
PY = fliplr(sum(P'));  % Reverse to put in normal order
disp('Answers are in X, Y, P, PY')
```

npr08_04

```
% file npr08_04.m
% Solution for problem P8-4
X = 2:12;
Y = 0:12;
PX = (1/36)*[1 2 3 4 5 6 5 4 3 2 1];
P0 = zeros(11,13);
for i = 1:11
    P0(i,1:i+2) = PX(i)*ibinom(i+1,1/2,0:i+1);
end
P = rot90(P0);
PY = fliplr(sum(P'));
disp('Answers are in X, Y, PY, P')
```

npr08_05

```
% file npr08_05.m
% Data and basic calculations for P8-5
PX = (1/36)*[1 2 3 4 5 6 5 4 3 2 1];
X = 2:12;
Y = 0:12;
P0 = zeros(11,13);
```

```
for i = 1:11
  P0(i,1:i+2) = PX(i)*ibinom(i+1,1/6,0:i+1);
end
P = rot90(P0);
PY = fliplr(sum(P'));
disp('Answers are in X, Y, P, PY')
```

npr08_06

```
% file  Newprobs/pr08_06.m
% Data for problem P8-6 (from Exam 2, 95f)
P = [0.0483    0.0357    0.0420    0.0399    0.0441
     0.0437    0.0323    0.0380    0.0361    0.0399
     0.0713    0.0527    0.0620    0.0609    0.0551
     0.0667    0.0493    0.0580    0.0651    0.0589];
X = [-2.3 -0.7 1.1 3.9 5.1];
Y = [ 1.3  2.5 4.1 5.3];
disp('Data are in X, Y, P')
```

npr08_07

```
% file pr08_07.m  (from Exam3, 96s)
% Data for problem P8-7
X = [-3.1 -0.5  1.2  2.4  3.7 4.9];
Y = [-3.8 -2.0  4.1  7.5];
P = [ 0.0090    0.0396    0.0594    0.0216    0.0440    0.0203;
      0.0495         0    0.1089    0.0528    0.0363    0.0231;
      0.0405    0.1320    0.0891    0.0324    0.0297    0.0189;
      0.0510    0.0484    0.0726    0.0132         0    0.0077];
disp('Data are in X, Y, P')
```

npr08_08

```
% file Newprobs/pr08_08.m (from Exam 4 96s)
% Data for problem P8-8
P = [0.0156  0.0191  0.0081  0.0035  0.0091  0.0070  0.0098  0.0056  0.0091  0.0049;
     0.0064  0.0204  0.0108  0.0040  0.0054  0.0080  0.0112  0.0064  0.0104  0.0056;
     0.0196  0.0256  0.0126  0.0060  0.0156  0.0120  0.0168  0.0096  0.0056  0.0084;
     0.0112  0.0182  0.0108  0.0070  0.0182  0.0140  0.0196  0.0012  0.0182  0.0038;
     0.0060  0.0260  0.0162  0.0050  0.0160  0.0200  0.0280  0.0060  0.0160  0.0040;
     0.0096  0.0056  0.0072  0.0060  0.0256  0.0120  0.0268  0.0096  0.0256  0.0084;
     0.0044  0.0134  0.0180  0.0140  0.0234  0.0180  0.0252  0.0244  0.0234  0.0126;
     0.0072  0.0017  0.0063  0.0045  0.0167  0.0090  0.0026  0.0172  0.0217  0.0223];

X = 1:2:19;
Y = [-5  -3  -1  3  5 9 10 12];
disp('Data are in X, Y, P')
```

npr08_09

```
% file pr08_09.m   (from Exam3 95f)
% Data for problem P8-9
```

```
P = [0.0390    0.0110    0.0050    0.0010    0.0010;
     0.0650    0.0700    0.0500    0.0150    0.0100;
     0.0310    0.0610    0.1370    0.0510    0.0330;
     0.0120    0.0490    0.1630    0.0580    0.0390;
     0.0030    0.0090    0.0450    0.0250    0.0170];
X = [1 1.5 2 2.5 3];
Y = [1 2 3 4 5];
disp('Data are in X, Y, P')
```

npr09_02

```
\begin{verbatim}
% file Newprobs/npr09_02.m
% Data for problem P9-2
P = [0.0589    0.0342    0.0304    0.0456    0.0209;
     0.0961    0.0556    0.0498    0.0744    0.0341;
     0.0682    0.0398    0.0350    0.0528    0.0242;
     0.0868    0.0504    0.0448    0.0672    0.0308];
X = [-3.9 -1.7 1.5 2.8 4.1];
Y = [-2 1 2.6 5.1];
disp('Data are in X, Y, P')
\end{verbatim}

```

npr10_16

```
\begin{verbatim}
% file npr10_16.m
% Data for problem P10-16
cx = [-2 1 3 0];
pmx = 0.001*[255  25 375  45 108  12 162  18];
cy = [1 3 1 -3];
pmy = minprob(0.01*[32 56 40]);
Z = [-1.3 1.2 2.7 3.4 5.8];
PZ = 0.01*[12 24 43 13  8];
disp('Data are in cx, pmx, cy, pmy, Z, PZ')
\end{verbatim}
```

npr12_10

```
% file npr12_10.m
% Data for problems P12-10, P12_11
cx = [-3.3 -1.7 2.3 7.6 -3.4];
pmx = 0.0001*[475 725 120 180 1125 1675  280 420 480 720 130 170 1120 1680 270 430];
cy = [10 17 20 -10];
pmy = 0.01*[6 14 9 21 6 14 9 21];
disp('Data are in cx, cy, pmx and pmy')
```

npr16_07

```
\begin{verbatim}
% file npr16_07.m
```

```
% Transition matrix for problem P16-7

P = [0.23    0.32    0.02    0.22    0.21;
     0.29    0.41    0.10    0.08    0.12;
     0.22    0.07    0.31    0.14    0.26;
     0.32    0.15    0.05    0.33    0.15;
     0.08    0.23    0.31    0.09    0.29];

disp('Transition matrix is P')
\end{verbatim}
```

npr16_09

```
% file npr16_09.m
% Transition matrix for problem P16-9
P = [0.2 0.5 0.3  0   0   0   0;
     0.6 0.1 0.3  0   0   0   0;
     0.2 0.7 0.1  0   0   0   0;
      0   0   0  0.6 0.4  0   0;
      0   0   0  0.5 0.5  0   0;
     0.1 0.3  0  0.2 0.1 0.1 0.2;
     0.1 0.2 0.1 0.2 0.2 0.2  0 ];
disp('Transition matrix is P')
```

This page titled 17.8: Matlab files for "Problems" in "Applied Probability" is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.

INDEX

B

Bernoulli sequence
4.3: Composite Trials

C

central limit theorem
13.2: Convergence and the Central Limit Theorem

composite trials
4.3: Composite Trials

conditional probability
3.1: Conditional Probability
3.2: Problems on Conditional Probability

conditioning event
3.1: Conditional Probability

convergence
13.2: Convergence and the Central Limit Theorem

correlation coefficient
12.2: Covariance and the Correlation Coefficient

Covariance
12.1: Variance
12.2: Covariance and the Correlation Coefficient

D

distribution function
7.1: Distribution and Density Functions

G

gamma distribution
7.2: Distribution Approximations

geometric distribution
7.1: Distribution and Density Functions

I

independent pairs
9.1: Independent Classes of Random Variables

J

joint distributions
8.1: Random Vectors and Joint Distributions

joint probability mass distribution
9.1: Independent Classes of Random Variables

L

Lebesque integral
11.2: Mathematical Expectation and General Random Variables

likelihood
1.1: Likelihood

M

marginal distributions
8.1: Random Vectors and Joint Distributions

minterm expansion
2.1: Minterms

minterms
2.1: Minterms

P

partitions
2.1: Minterms

Poisson approximation
7.2: Distribution Approximations

Poisson distribution
7.2: Distribution Approximations

Q

quantile function
10.3: The Quantile Function

R

rectangle test
9.1: Independent Classes of Random Variables

S

standard deviation
12.1: Variance

V

variance
12.1: Variance

W

Weibull density
7.1: Distribution and Density Functions

Weibull distribution
10.3: The Quantile Function

This page titled Index is shared under a license and was authored, remixed, and/or curated by LibreTexts.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a license and was authored, remixed, and/or curated by LibreTexts.

DETAILED LICENSING

Overview

Title: Applied Probability (Pfeiffer)

Webpages: 85

All licenses found:

- CC BY 3.0: 92.9% (79 pages)
- Undeclared: 7.1% (6 pages)

By Page

- Applied Probability (Pfeiffer) - *CC BY 3.0*
  - Front Matter - *CC BY 3.0*
    - TitlePage - *Undeclared*
    - InfoPage - *CC BY 3.0*
    - Preface - *CC BY 3.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Probability Systems - *CC BY 3.0*
    - 1.1: Likelihood - *CC BY 3.0*
    - 1.2: Probability Systems - *CC BY 3.0*
    - 1.3: Interpretations - *CC BY 3.0*
    - 1.4: Problems on Probability Systems - *CC BY 3.0*
  - 2: Minterm Analysis - *CC BY 3.0*
    - 2.1: Minterms - *CC BY 3.0*
    - 2.2: Minterms and MATLAB Calculations - *CC BY 3.0*
    - 2.3: Problems on Minterm Analysis - *CC BY 3.0*
  - 3: Conditional Probability - *CC BY 3.0*
    - 3.1: Conditional Probability - *CC BY 3.0*
    - 3.2: Problems on Conditional Probability - *CC BY 3.0*
  - 4: Independence of Events - *CC BY 3.0*
    - 4.1: Independence of Events - *CC BY 3.0*
    - 4.2: MATLAB and Independent Classes - *CC BY 3.0*
    - 4.3: Composite Trials - *CC BY 3.0*
    - 4.4: Problems on Independence of Events - *CC BY 3.0*
  - 5: Conditional Independence - *CC BY 3.0*
    - 5.1: Conditional Independence - *CC BY 3.0*
    - 5.2: Patterns of Probable Inference - *CC BY 3.0*
    - 5.3: Problems on Conditional Independence - *CC BY 3.0*
  - 6: Random Variables and Probabilities - *CC BY 3.0*
    - 6.1: Random Variables and Probabilities - *CC BY 3.0*
    - 6.2: Problems on Random Variables and Probabilities - *CC BY 3.0*
  - 7: Distribution and Density Functions - *CC BY 3.0*
    - 7.1: Distribution and Density Functions - *CC BY 3.0*
    - 7.2: Distribution Approximations - *CC BY 3.0*
    - 7.3: Problems on Distribution and Density Functions - *CC BY 3.0*
  - 8: Random Vectors and Joint Distributions - *CC BY 3.0*
    - 8.1: Random Vectors and Joint Distributions - *CC BY 3.0*
    - 8.2: Random Vectors and MATLAB - *CC BY 3.0*
    - 8.3: Problems on Random Vectors and Joint Distributions - *CC BY 3.0*
  - 9: Independent Classes of Random Variables - *CC BY 3.0*
    - 9.1: Independent Classes of Random Variables - *CC BY 3.0*
    - 9.2: Problems on Independent Classes of Random Variables - *CC BY 3.0*
  - 10: Functions of Random Variables - *CC BY 3.0*
    - 10.1: Functions of a Random Variable - *CC BY 3.0*
    - 10.2: Function of Random Vectors - *CC BY 3.0*
    - 10.3: The Quantile Function - *CC BY 3.0*
    - 10.4: Problems on Functions of Random Variables - *CC BY 3.0*
  - 11: Mathematical Expectation - *CC BY 3.0*
    - 11.1: Mathematical Expectation- Simple Random Variables - *CC BY 3.0*
    - 11.2: Mathematical Expectation and General Random Variables - *CC BY 3.0*
    - 11.3: Problems on Mathematical Expectation - *CC BY 3.0*
  - 12: Variance, Covariance, and Linear Regression - *CC BY 3.0*
    - 12.1: Variance - *CC BY 3.0*
    - 12.2: Covariance and the Correlation Coefficient - *CC BY 3.0*
    - 12.3: Linear Regression - *CC BY 3.0*
    - 12.4: Problems on Variance, Covariance, Linear Regression - *CC BY 3.0*

- 13: Transform Methods - *CC BY 3.0*
  - 13.1: Transform Methods - *CC BY 3.0*
  - 13.2: Convergence and the Central Limit Theorem - *CC BY 3.0*
  - 13.3: Simple Random Samples and Statistics - *CC BY 3.0*
  - 13.4: Problems on Transform Methods - *CC BY 3.0*
- 14: Conditional Expectation, Regression - *CC BY 3.0*
  - 14.1: Conditional Expectation, Regression - *CC BY 3.0*
  - 14.2: Problems on Conditional Expectation, Regression - *CC BY 3.0*
- 15: Random Selection - *CC BY 3.0*
  - 15.1: Random Selection - *CC BY 3.0*
  - 15.2: Some Random Selection Problems - *CC BY 3.0*
  - 15.3: Problems on Random Selection - *CC BY 3.0*
- 16: Conditional Independence, Given a Random Vector - *CC BY 3.0*
  - 16.1: Conditional Independence, Given a Random Vector - *CC BY 3.0*
  - 16.2: Elements of Markov Sequences - *CC BY 3.0*
  - 16.3: Problems on Conditional Independence, Given a Random Vector - *CC BY 3.0*
- 17: Appendices - *CC BY 3.0*
  - 17.1: Appendix A to Applied Probability- Directory of m-functions and m-procedures - *CC BY 3.0*
  - 17.2: Appendix B to Applied Probability- some mathematical aids - *CC BY 3.0*
  - 17.3: Appendix C- Data on some common distributions - *CC BY 3.0*
  - 17.4: Appendix D to Applied Probability- The standard normal distribution - *CC BY 3.0*
  - 17.5: Appendix E to Applied Probability - Properties of Mathematical Expectation - *CC BY 3.0*
  - 17.6: Appendix F to Applied Probability- Properties of conditional expectation, given a random vector - *CC BY 3.0*
  - 17.7: Appendix G to Applied Probability- Properties of conditional independence, given a random vector - *CC BY 3.0*
  - 17.8: Matlab files for "Problems" in "Applied Probability" - *CC BY 3.0*
- Back Matter - *CC BY 3.0*
  - Index - *Undeclared*
  - Glossary - *Undeclared*
  - Detailed Licensing - *Undeclared*

This page titled Applied Probability (Pfeiffer) is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Paul Pfeiffer via source content that was edited to the style and standards of the LibreTexts platform.