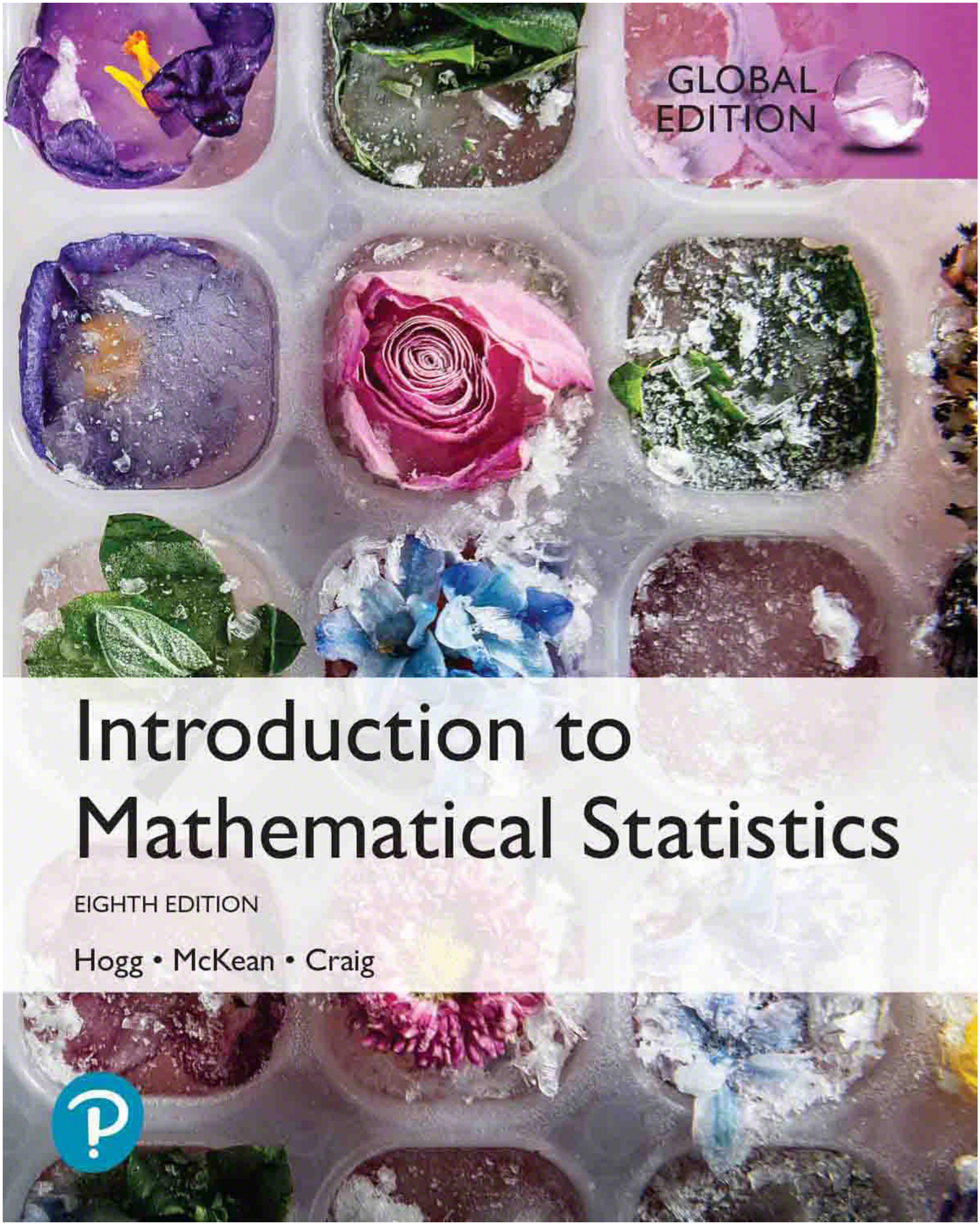
GLOBAL EDITION

Introduction to Mathematical Statistics

EIGHTH EDITION

Hogg • McKean • Craig



Introduction to Mathematical Statistics

Eighth Edition

Global Edition

Robert V. Hogg
University of Iowa

Joseph W. McKean
Western Michigan University

Allen T. Craig
Late Professor of Statistics
University of Iowa

Pearson

Director, Portfolio Management: Deirdre Lynch
Courseware Portfolio Manager: Patrick Barbera
Portfolio Management Assistant: Morgan Danna
Associate Editor, Global Edition: Shaoni Mukherjee
Content Producer: Lauren Morse
Managing Producer: Scott Disanno
Product Marketing Manager: Yvonne Vannatta
Field Marketing Manager: Evan St. Cyr
Marketing Assistant: Jon Bryant
Senior Author Support/Technology Specialist: Joe Vetere
Managing Editor, Digital Media, Global Edition: Gargi Banerjee
Manager, Rights and Permissions: Gina Cheselka
Manufacturing Buyer: Carol Melville, LSC Communications
Manufacturing Buyer, Global Edition: Kay Holman
Art Director: Barbara Atkinson
Production Coordination and Illustrations: Integra
Cover Design: Lumina Datamatics
Cover Image: Anna Mente/Shutterstock

Pearson Education Limited
KAO Two
KAO Park
Hockham Way
Harlow
Essex
CM17 9SR
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsonglobaleditions.com

ISBN 10: 1-292-26476-4

ISBN 13: 978-1-292-26476-9

eBook ISBN 13: 978-1-292-26478-3

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Typeset by Integra Software Services Pvt. Ltd.

Dedicated to my wife Marge
and to the memory of Bob Hogg

This page intentionally left blank

Contents

This page intentionally left blank

Preface

We have made substantial changes in this edition of *Introduction to Mathematical Statistics*. Some of these changes help students appreciate the connection between statistical theory and statistical practice while other changes enhance the development and discussion of the statistical theory presented in this book.

Many of the changes in this edition reflect comments made by our readers. One of these comments concerned the small number of real data sets in the previous editions. In this edition, we have included more real data sets, using them to illustrate statistical methods or to compare methods. Further, we have made these data sets accessible to students by including them in the free R package `hmcpkg`. They can also be individually downloaded in an R session at the url listed on page 12. In general, the R code for the analyses on these data sets is given in the text.

We have also expanded the use of the statistical software R. We selected R because it is a powerful statistical language that is free and runs on all three main platforms (Windows, Mac, and Linux). Instructors, though, can select another statistical package. We have also expanded our use of R functions to compute analyses and simulation studies, including several games. We have kept the level of coding for these functions straightforward. Our goal is to show students that with a few simple lines of code they can perform significant computations. Appendix B contains a brief R primer, which suffices for the understanding of the R used in the text. As with the data sets, these R functions can be sourced individually at the cited url; however, they are also included in the package `hmcpkg`.

We have supplemented the mathematical review material in Appendix A, placing it in the document *Mathematical Primer for Introduction to Mathematical Statistics*. It is freely available for students to download at the listed url. Besides sequences, this supplement reviews the topics of infinite series, differentiation, and integration (univariate and bivariate). We have also expanded the discussion of iterated integrals in the text. We have added figures to clarify discussion.

We have retained the order of elementary statistical inferences (Chapter 4) and asymptotic theory (Chapter 5). In Chapters 5 and 6, we have written brief reviews of the material in Chapter 4, so that Chapters 4 and 5 are essentially independent of one another and, hence, can be interchanged. In Chapter 3, we now begin the section on the multivariate normal distribution with a subsection on the bivariate normal distribution. Several important topics have been added. This includes Tukey's multiple comparison procedure in Chapter 9 and confidence intervals for the correlation coefficients found in Chapters 9 and 10. Chapter 7 now contains a discussion on standard errors for estimates obtained by bootstrapping the sample. Several topics that were discussed in the Exercises are now discussed in the text. Examples include quantiles, Section 1.7.1, and hazard functions, Section 3.3. In general, we have made more use of subsections to break up some of the discussion. Also, several more sections are now indicated by * as being optional.

Content and Course Planning

Chapters 1 and 2 develop probability models for univariate and multivariate variables while Chapter 3 discusses many of the most widely used probability models. Chapter 4 discusses statistical theory for much of the inference found in a standard statistical methods course. Chapter 5 presents asymptotic theory, concluding with the Central Limit Theorem. Chapter 6 provides a complete inference (estimation and testing) based on maximum likelihood theory. The EM algorithm is also discussed. Chapters 7–8 contain optimal estimation procedures and tests of statistical hypotheses. The final three chapters provide theory for three important topics in statistics. Chapter 9 contains inference for normal theory methods for basic analysis of variance, univariate regression, and correlation models. Chapter 10 presents nonparametric methods (estimation and testing) for location and univariate regression models. It also includes discussion on the robust concepts of efficiency, influence, and breakdown. Chapter 11 offers an introduction to Bayesian methods. This includes traditional Bayesian procedures as well as Markov Chain Monte Carlo techniques.

Several courses can be designed using our book. The basic two-semester course in mathematical statistics covers most of the material in Chapters 1–8 with topics selected from the remaining chapters. For such a course, the instructor would have the option of interchanging the order of Chapters 4 and 5, thus beginning the second semester with an introduction to statistical theory (Chapter 4). A one-semester course could consist of Chapters 1–4 with a selection of topics from Chapter 5. Under this option, the student sees much of the statistical theory for the methods discussed in a non-theoretical course in methods. On the other hand, as with the two-semester sequence, after covering Chapters 1–3, the instructor can elect to cover Chapter 5 and finish the course with a selection of topics from Chapter 4.

The data sets and R functions used in this book and the R package `hmcpkg` can be downloaded from this title's page at the site:
`www.pearsonglobaleditions.com`

Acknowledgments

Bob Hogg passed away in 2014, so he did not work on this edition of the book. Often, though, when I was trying to decide whether or not to make a change in the manuscript, I found myself thinking of what Bob would do. In his memory, I have retained the order of the authors for this edition.

As with earlier editions, comments from readers are always welcomed and appreciated. We would like to thank these reviewers of the previous edition: James Baldone, Virginia College; Steven Culpepper, University of Illinois at Urbana-Champaign; Yuichiro Kakihara, California State University; Jaechoul Lee, Boise State University; Michael Levine, Purdue University; Tingni Sun, University of Maryland, College Park; and Daniel Weiner, Boston University. We appreciated and took into consideration their comments for this revision. We appreciate the helpful comments of Thomas Hettmansperger of Penn State University, Ash Abebe of Auburn University, and Professor Ioannis Kalogridis of the University of Leuven. A special thanks to Patrick Barbera (Portfolio Manager, Statistics), Lauren Morse (Content Producer, Math/Stats), Yvonne Vannatta (Product Marketing Manager), and the rest of the staff at Pearson for their help in putting this edition together. Thanks also to Richard Ponticelli, North Shore Community College, who accuracy checked the page proofs. Also, a special thanks to my wife Marge for her unwavering support and encouragement of my efforts in writing this edition.

Joe McKean

Acknowledgments for the Global Edition

Pearson would like to thank and acknowledge the following people for their work on this Global Edition.

Contributors

Polina Dolmatova, American University of Central Asia
Tsung Fei Khang, University of Malaya
Eric A. L. LI, The University of Hong Kong
Choung Min Ng, University of Malaya
Neelesh S. Upadhye, Indian Institute of Technology Madras

Reviewers

Samprit Chakrabarti, ICFAI Business School Kolkata
Rubén Garcia Berasategui, Jakarta International College
Aneesh Kumar K., Mahatma Gandhi College, Iritty
Jesper Rydén, Swedish University of Agricultural Sciences
Pooja Sengupta, International Management Institute Kolkata
Neelesh S. Upadhye, Indian Institute of Technology Madras

This page intentionally left blank

Chapter 1

Probability and Distributions

1.1 Introduction

In this section, we intuitively discuss the concepts of a probability model which we formalize in Secton 1.3 Many kinds of investigations may be characterized in part by the fact that repeated experimentation, under essentially the same conditions, is more or less standard procedure. For instance, in medical research, interest may center on the effect of a drug that is to be administered; or an economist may be concerned with the prices of three specified commodities at various time intervals; or an agronomist may wish to study the effect that a chemical fertilizer has on the yield of a cereal grain. The only way in which an investigator can elicit information about any such phenomenon is to perform the experiment. Each experiment terminates with an *outcome*. But it is characteristic of these experiments that the outcome cannot be predicted with certainty prior to the experiment.

Suppose that we have such an experiment, but the experiment is of such a nature that a collection of every possible outcome can be described prior to its performance. If this kind of experiment can be repeated under the same conditions, it is called a **random experiment**, and the collection of every possible outcome is called the experimental space or the **sample space**. We denote the sample space by $\mathcal{C}$.

Example 1.1.1. In the toss of a coin, let the outcome tails be denoted by T and let the outcome heads be denoted by H. If we assume that the coin may be repeatedly tossed under the same conditions, then the toss of this coin is an example of a random experiment in which the outcome is one of the two symbols T or H; that is, the sample space is the collection of these two symbols. For this example, then, $\mathcal{C} = \{H, T\}$. ∎

Example 1.1.2. In the cast of one red die and one white die, let the outcome be the ordered pair (number of spots up on the red die, number of spots up on the white die). If we assume that these two dice may be repeatedly cast under the same conditions, then the cast of this pair of dice is a random experiment. The sample space consists of the 36 ordered pairs: $\mathcal{C} = \{(1,1), \ldots, (1,6), (2,1), \ldots, (2,6), \ldots, (6,6)\}$. ∎

We generally use small Roman letters for the elements of $\mathcal{C}$ such as a, b, or c. Often for an experiment, we are interested in the chances of certain subsets of elements of the sample space occurring. Subsets of $\mathcal{C}$ are often called **events** and are generally denoted by capitol Roman letters such as $A, B,$ or C. If the experiment results in an element in an event A, we say the event A has occurred. We are interested in the chances that an event occurs. For instance, in Example 1.1.1 we may be interested in the chances of getting heads; i.e., the chances of the event $A = \{H\}$ occurring. In the second example, we may be interested in the occurrence of the sum of the upfaces of the dice being "7" or "11;" that is, in the occurrence of the event $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(5,6),(6,5)\}$.

Now conceive of our having made N repeated performances of the random experiment. Then we can count the number f of times (the **frequency**) that the event A actually occurred throughout the N performances. The ratio f/N is called the **relative frequency** of the event A in these N experiments. A relative frequency is usually quite erratic for small values of N, as you can discover by tossing a coin. But as N increases, experience indicates that we associate with the event A a number, say p, that is equal or approximately equal to that number about which the relative frequency seems to stabilize. If we do this, then the number p can be interpreted as that number which, in future performances of the experiment, the relative frequency of the event A will either equal or approximate. Thus, although we *cannot* predict the outcome of a random experiment, we *can*, for a large value of N, predict approximately the relative frequency with which the outcome will be in A. The number p associated with the event A is given various names. Sometimes it is called the *probability* that the outcome of the random experiment is in A; sometimes it is called the *probability* of the event A; and sometimes it is called the *probability measure* of A. The context usually suggests an appropriate choice of terminology.

Example 1.1.3. Let $\mathcal{C}$ denote the sample space of Example 1.1.2 and let B be the collection of every ordered pair of $\mathcal{C}$ for which the sum of the pair is equal to seven. Thus $B = \{(1,6),(2,5),(3,4),(4,3),(5,2)(6,1)\}$. Suppose that the dice are cast $N = 400$ times and let f denote the frequency of a sum of seven. Suppose that 400 casts result in $f = 60$. Then the relative frequency with which the outcome was in B is $f/N = \frac{60}{400} = 0.15$. Thus we might associate with B a number p that is close to 0.15, and p would be called the probability of the event B. ■

Remark 1.1.1. The preceding interpretation of probability is sometimes referred to as the *relative frequency approach*, and it obviously depends upon the fact that an experiment can be repeated under essentially identical conditions. However, many persons extend probability to other situations by treating it as a rational measure of belief. For example, the statement $p = \frac{2}{5}$ for an event A would mean to them that their *personal* or *subjective* probability of the event A is equal to $\frac{2}{5}$. Hence, if they are not opposed to gambling, this could be interpreted as a willingness on their part to bet on the outcome of A so that the two possible payoffs are in the ratio $p/(1-p) = \frac{2}{5}/\frac{3}{5} = \frac{2}{3}$. Moreover, if they truly believe that $p = \frac{2}{5}$ is correct, they would be willing to accept either side of the bet: (a) win 3 units if A occurs and lose 2 if it does not occur, or (b) win 2 units if A does not occur and lose 3 if it does. However, since the mathematical properties of probability given in Section 1.3 are consistent with either of these interpretations, the subsequent mathematical development does not depend upon which approach is used. ■

The primary purpose of having a mathematical theory of statistics is to provide mathematical models for random experiments. Once a model for such an experiment has been provided and the theory worked out in detail, the statistician may, within this framework, make inferences (that is, draw conclusions) about the random experiment. The construction of such a model requires a theory of probability. One of the more logically satisfying theories of probability is that based on the concepts of sets and functions of sets. These concepts are introduced in Section 1.2.

1.2 Sets

The concept of a *set* or a *collection* of objects is usually left undefined. However, a particular set can be described so that there is no misunderstanding as to what collection of objects is under consideration. For example, the set of the first 10 positive integers is sufficiently well described to make clear that the numbers $\frac{3}{4}$ and 14 are not in the set, while the number 3 is in the set. If an object belongs to a set, it is said to be an *element* of the set. For example, if C denotes the set of real numbers x for which $0 \le x \le 1$, then $\frac{3}{4}$ is an element of the set C. The fact that $\frac{3}{4}$ is an element of the set C is indicated by writing $\frac{3}{4} \in C$. More generally, $c \in C$ means that c is an element of the set C.

The sets that concern us are frequently *sets of numbers*. However, the language of sets of *points* proves somewhat more convenient than that of sets of numbers. Accordingly, we briefly indicate how we use this terminology. In analytic geometry considerable emphasis is placed on the fact that to each point on a line (on which an origin and a unit point have been selected) there corresponds one and only one number, say x; and that to each number x there corresponds one and only one point on the line. This one-to-one correspondence between the numbers and points on a line enables us to speak, without misunderstanding, of the "point x" instead of the "number x." Furthermore, with a plane rectangular coordinate system and with x and y numbers, to each symbol (x, y) there corresponds one and only one point in the plane; and to each point in the plane there corresponds but one such symbol. Here again, we may speak of the "point (x, y)," meaning the "ordered number pair x and y." This convenient language can be used when we have a rectangular coordinate system in a space of three or more dimensions. Thus the "point $(x_1, x_2, \ldots, x_n)$" means the numbers $x_1, x_2, \ldots, x_n$ in the order stated. Accordingly, in describing our sets, we frequently speak of a set of points (a set whose elements are points), being careful, of course, to describe the set so as to avoid any ambiguity. The notation $C = \{x : 0 \le x \le 1\}$ is read "C is the one-dimensional set of points x for which $0 \le x \le 1$." Similarly, $C = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$ can be read "C is the two-dimensional set of points (x, y) that are interior to, or on the boundary of, a square with opposite vertices at $(0, 0)$ and $(1, 1)$."

We say a set C is **countable** if C is finite or has as many elements as there are positive integers. For example, the sets $C_1 = \{1, 2, \ldots, 100\}$ and $C_2 = \{1, 3, 5, 7, \ldots\}$

are countable sets. The interval of real numbers $(0, 1]$, though, is not countable.

1.2.1 Review of Set Theory

As in Section 1.1, let $\mathcal{C}$ denote the sample space for the experiment. Recall that events are subsets of $\mathcal{C}$. We use the words event and subset interchangeably in this section. An elementary algebra of sets will prove quite useful for our purposes. We now review this algebra below along with illustrative examples. For illustration, we also make use of **Venn diagrams**. Consider the collection of Venn diagrams in Figure 1.2.1. The interior of the rectangle in each plot represents the sample space C. The shaded region in Panel (a) represents the event A.

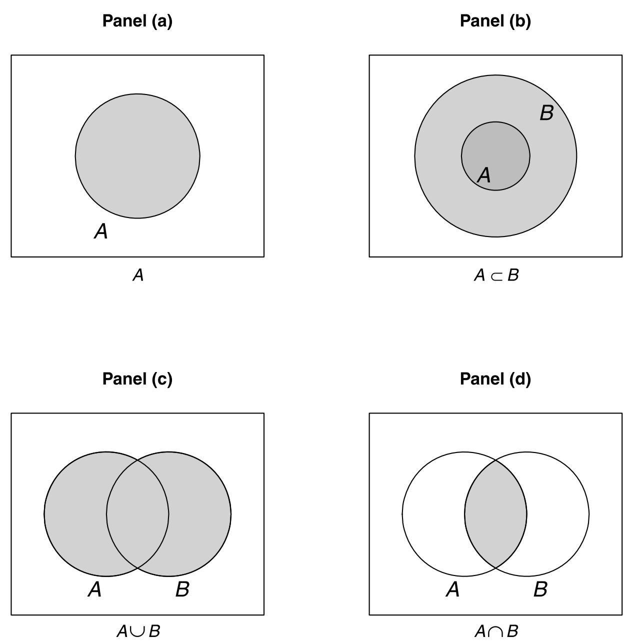


Figure 1.2.1: A series of Venn diagrams. The sample space C is represented by the interior of the rectangle in each plot. Panel (a) depicts the event A; Panel (b) depicts $A \subset B$; Panel (c) depicts $A \cup B$; and Panel (d) depicts $A \cap B$.

We first define the complement of an event A.

Definition 1.2.1. *The* **complement** *of an event A is the set of all elements in C which are not in A. We denote the complement of A by A^c. That is, $A^c = \{x \in \mathcal{C} : x \notin A\}$.*

The complement of A is represented by the white space in the Venn diagram in Panel (a) of Figure 1.2.1.

The empty set is the event with no elements in it. It is denoted by ϕ. Note that $\mathcal{C}^c = \phi$ and $\phi^c = \mathcal{C}$. The next definition defines when one event is a subset of another.

Definition 1.2.2. *If each element of a set A is also an element of set B, the set A is called a* **subset** *of the set B. This is indicated by writing $A \subset B$. If $A \subset B$ and also $B \subset A$, the two sets have the same elements, and this is indicated by writing $A = B$.*

Panel (b) of Figure 1.2.1 depicts $A \subset B$.

The event A or B is defined as follows:

Definition 1.2.3. *Let A and B be events. Then the* **union** *of A and B is the set of all elements that are in A or in B or in both A and B. The union of A and B is denoted by $A \cup B$*

Panel (c) of Figure 1.2.1 shows $A \cup B$.

The event that both A and B occur is defined by,

Definition 1.2.4. *Let A and B be events. Then the* **intersection** *of A and B is the set of all elements that are in both A and B. The intersection of A and B is denoted by $A \cap B$*

Panel (d) of Figure 1.2.1 illustrates $A \cap B$.

Two events are **disjoint** if they have no elements in common. More formally we define

Definition 1.2.5. *Let A and B be events. Then A and B are* **disjoint** *if $A \cap B = \phi$*

If A and B are disjoint, then we say $A \cup B$ forms a **disjoint union.** The next two examples illustrate these concepts.

Example 1.2.1. Suppose we have a spinner with the numbers 1 through 10 on it. The experiment is to spin the spinner and record the number spun. Then $\mathcal{C} = \{1, 2, \ldots, 10\}$. Define the events A, B, and C by $A = \{1, 2\}$, $B = \{2, 3, 4\}$, and $C = \{3, 4, 5, 6\}$, respectively.

$$A^c = \{3, 4, \ldots, 10\}; \quad A \cup B = \{1, 2, 3, 4\}; \quad A \cap B = \{2\}$$
$$A \cap C = \phi; \quad B \cap C = \{3, 4\}; \quad B \cap C \subset B; \quad B \cap C \subset C$$
$$A \cup (B \cap C) = \{1, 2\} \cup \{3, 4\} = \{1, 2, 3, 4\} \tag{1.2.1}$$
$$(A \cup B) \cap (A \cup C) = \{1, 2, 3, 4\} \cap \{1, 2, 3, 4, 5, 6\} = \{1, 2, 3, 4\} \tag{1.2.2}$$

The reader should verify these results. ■

Example 1.2.2. For this example, suppose the experiment is to select a real number in the open interval $(0, 5)$; hence, the sample space is $\mathcal{C} = (0, 5)$. Let $A = (1, 3)$,

$B = (2, 4)$, and $C = [3, 4.5)$.

$$A \cup B = (1, 4); \quad A \cap B = (2, 3); \quad B \cap C = [3, 4)$$
$$A \cap (B \cup C) = (1, 3) \cap (2, 4.5) = (2, 3) \tag{1.2.3}$$
$$(A \cap B) \cup (A \cap C) = (2, 3) \cup \phi = (2, 3) \tag{1.2.4}$$

A sketch of the real number line between 0 and 5 helps to verify these results. ■

Expressions (1.2.1)–(1.2.2) and (1.2.3)–(1.2.4) are illustrations of general **distributive laws**. For any sets A, B, and C,

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned} \tag{1.2.5}$$

These follow directly from set theory. To verify each identity, sketch Venn diagrams of both sides.

The next two identities are collectively known as **DeMorgan's Laws**. For any sets A and B,

$$(A \cap B)^c = A^c \cup B^c \tag{1.2.6}$$
$$(A \cup B)^c = A^c \cap B^c. \tag{1.2.7}$$

For instance, in Example 1.2.1,

$$(A \cup B)^c = \{1, 2, 3, 4\}^c = \{5, 6, \ldots, 10\} = \{3, 4, \ldots, 10\} \cap \{\{1, 5, 6, \ldots, 10\} = A^c \cap B^c;$$

while, from Example 1.2.2,

$$(A \cap B)^c = (2, 3)^c = (0, 2] \cup [3, 5) = [(0, 1] \cup [3, 5)] \cup [(0, 2] \cup [4, 5)] = A^c \cup B^c.$$

As the last expression suggests, it is easy to extend unions and intersections to more than two sets. If $A_1, A_2, \ldots, A_n$ are any sets, we define

$$A_1 \cup A_2 \cup \cdots \cup A_n = \{x : x \in A_i, \text{ for some } i = 1, 2, \ldots, n\} \tag{1.2.8}$$
$$A_1 \cap A_2 \cap \cdots \cap A_n = \{x : x \in A_i, \text{ for all } i = 1, 2, \ldots, n\}. \tag{1.2.9}$$

We often abbreviative these by $\cup_{i=1}^n A_i$ and $\cap_{i=1}^n A_i$, respectively. Expressions for countable unions and intersections follow directly; that is, if $A_1, A_2, \ldots, A_n \ldots$ is a sequence of sets then

$$A_1 \cup A_2 \cup \cdots = \{x : x \in A_n, \text{ for some } n = 1, 2, \ldots\} = \cup_{n=1}^\infty A_n \tag{1.2.10}$$
$$A_1 \cap A_2 \cap \cdots = \{x : x \in A_n, \text{ for all } n = 1, 2, \ldots\} = \cap_{n=1}^\infty A_n. \tag{1.2.11}$$

The next two examples illustrate these ideas.

Example 1.2.3. Suppose $\mathcal{C} = \{1, 2, 3, \ldots\}$. If $A_n = \{1, 3, \ldots, 2n - 1\}$ and $B_n = \{n, n + 1, \ldots\}$, for $n = 1, 2, 3, \ldots$, then

$$\cup_{n=1}^\infty A_n = \{1, 3, 5, \ldots\}; \quad \cap_{n=1}^\infty A_n = \{1\}; \tag{1.2.12}$$
$$\cup_{n=1}^\infty B_n = \mathcal{C}; \quad \cap_{n=1}^\infty B_n = \phi. \ ■ \tag{1.2.13}$$

Example 1.2.4. Suppose $\mathcal{C}$ is the interval of real numbers $(0, 5)$. Suppose $C_n = (1 - n^{-1}, 2 + n^{-1})$ and $D_n = (n^{-1}, 3 - n^{-1})$, for $n = 1, 2, 3, \ldots$. Then

$$\cup_{n=1}^{\infty} C_n = (0, 3); \quad \cap_{n=1}^{\infty} C_n = [1, 2] \tag{1.2.14}$$

$$\cup_{n=1}^{\infty} D_n = (0, 3); \quad \cap_{n=1}^{\infty} D_n = (1, 2). \quad \blacksquare \tag{1.2.15}$$

We occassionally have sequences of sets that are **monotone**. They are of two types. We say a sequence of sets $\{A_n\}$ is **nondecreasing, (nested upward)**, if $A_n \subset A_{n+1}$ for $n = 1, 2, 3, \ldots$. For such a sequence, we define

$$\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n. \tag{1.2.16}$$

The sequence of sets $A_n = \{1, 3, \ldots, 2n - 1\}$ of Example 1.2.3 is such a sequence. So in this case, we write $\lim_{n\to\infty} A_n = \{1, 3, 5, \ldots\}$. The sequence of sets $\{D_n\}$ of Example 1.2.4 is also a nondecreasing suquence of sets.

The second type of monotone sets consists of the **nonincreasing, (nested downward)** sequences. A sequence of sets $\{A_n\}$ is **nonincreasing**, if $A_n \supset A_{n+1}$ for $n = 1, 2, 3, \ldots$. In this case, we define

$$\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n. \tag{1.2.17}$$

The sequences of sets $\{B_n\}$ and $\{C_n\}$ of Examples 1.2.3 and 1.2.4, respectively, are examples of nonincreasing sequences of sets.

1.2.2 Set Functions

Many of the functions used in calculus and in this book are functions that map real numbers into real numbers. We are concerned also with functions that map sets into real numbers. Such functions are naturally called functions of a set or, more simply, **set functions**. Next we give some examples of set functions and evaluate them for certain simple sets.

Example 1.2.5. Let $\mathcal{C} = R$, the set of real numbers. For a subset A in $\mathcal{C}$, let $Q(A)$ be equal to the number of points in A that correspond to positive integers. Then $Q(A)$ is a set function of the set A. Thus, if $A = \{x : 0 < x < 5\}$, then $Q(A) = 4$; if $A = \{-2, -1\}$, then $Q(A) = 0$; and if $A = \{x : -\infty < x < 6\}$, then $Q(A) = 5$. $\blacksquare$

Example 1.2.6. Let $\mathcal{C} = R^2$. For a subset A of $\mathcal{C}$, let $Q(A)$ be the area of A if A has a finite area; otherwise, let $Q(A)$ be undefined. Thus, if $A = \{(x, y) : x^2 + y^2 \le 1\}$, then $Q(A) = \pi$; if $A = \{(0, 0), (1, 1), (0, 1)\}$, then $Q(A) = 0$; and if $A = \{(x, y) : 0 \le x, 0 \le y, x + y \le 1\}$, then $Q(A) = \frac{1}{2}$. $\blacksquare$

Often our set functions are defined in terms of sums or integrals.[1] With this in mind, we introduce the following notation. The symbol

$$\int_A f(x)\, dx$$

[1]Please see Chapters 2 and 3 of *Mathematical Comments*, at site noted in the Preface, for a review of sums and integrals

means the ordinary (Riemann) integral of $f(x)$ over a prescribed one-dimensional set A and the symbol

$$\int\!\!\int_A g(x,y)\,dxdy$$

means the Riemann integral of $g(x,y)$ over a prescribed two-dimensional set A. This notation can be extended to integrals over n dimensions. To be sure, unless these sets A and these functions $f(x)$ and $g(x,y)$ are chosen with care, the integrals frequently fail to exist. Similarly, the symbol

$$\sum_A f(x)$$

means the sum extended over all $x \in A$ and the symbol

$$\sum\sum_A g(x,y)$$

means the sum extended over all $(x,y) \in A$. As with integration, this notation extends to sums over n dimensions.

The first example is for a set function defined on sums involving a **geometric series**. As pointed out in Example 2.3.1 of *Mathematical Comments*,[2] if $|a| < 1$, then the following series converges to $1/(1-a)$:

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}, \quad \text{if } |a| < 1. \tag{1.2.18}$$

Example 1.2.7. Let $\mathcal{C}$ be the set of all nonnegative integers and let A be a subset of $\mathcal{C}$. Define the set function Q by

$$Q(A) = \sum_{n \in A} \left(\frac{2}{3}\right)^n. \tag{1.2.19}$$

It follows from (1.2.18) that $Q(\mathcal{C}) = 3$. If $A = \{1,2,3\}$ then $Q(A) = 38/27$. Suppose $B = \{1,3,5,\ldots\}$ is the set of all odd positive integers. The computation of $Q(B)$ is given next. This derivation consists of rewriting the series so that (1.2.18) can be applied. Frequently, we perform such derivations in this book.

$$\begin{aligned} Q(B) &= \sum_{n \in B} \left(\frac{2}{3}\right)^n = \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^{2n+1} \\ &= \frac{2}{3}\sum_{n=0}^{\infty}\left[\left(\frac{2}{3}\right)^2\right]^n = \frac{2}{3}\frac{1}{1-(4/9)} = \frac{6}{5} \quad \blacksquare \end{aligned}$$

In the next example, the set function is defined in terms of an integral involving the exponential function $f(x) = e^{-x}$.

[2]Downloadable at site noted in the Preface

Example 1.2.8. Let $\mathcal{C}$ be the interval of positive real numbers, i.e., $\mathcal{C} = (0, \infty)$. Let A be a subset of $\mathcal{C}$. Define the set function Q by

$$Q(A) = \int_A e^{-x}\,dx, \tag{1.2.20}$$

provided the integral exists. The reader should work through the following integrations:

$$\begin{aligned}
&Q[(1,3)] = \int_1^3 e^{-x}\,dx = -e^{-x}\Big|_1^3 = e^{-1} - e^{-3} \doteq 0.318\\
&Q[(5 \text{ and } \infty)] = \int_1^3 e^{-x}\,dx = -e^{-x}\Big|_5^\infty = e^{-5} \doteq 0.007\\
&Q[(1,3) \cup [3,5)] = \int_1^5 e^{-x}\,dx = \int_1^3 e^{-x}\,dx + \int_3^5 e^{-x}\,dx = Q[(1,3)] + Q([3,5)]\\
&Q(\mathcal{C}) = \int_0^\infty e^{-x}\,dx = 1. \quad \blacksquare
\end{aligned}$$

Our final example, involves an n dimensional integral.

Example 1.2.9. Let $\mathcal{C} = R^n$. For A in $\mathcal{C}$ define the set function

$$Q(A) = \int \cdots \underset{A}{\int} dx_1 dx_2 \cdots dx_n,$$

provided the integral exists. For example, if $A = \{(x_1, x_2, \ldots, x_n) : 0 \le x_1 \le x_2, 0 \le x_i \le 1,$ for $i = 2, 3, \ldots,$ n$\}$, then upon expressing the multiple integral as an iterated integral[3] we obtain

$$\begin{aligned}
Q(A) &= \int_0^1 \left[\int_0^{x_2} dx_1\right] dx_2 \bullet \prod_{i=3}^n \left[\int_0^1 dx_i\right]\\
&= \frac{x_2^2}{2}\Big|_0^1 \bullet 1 = \frac{1}{2}.
\end{aligned}$$

If $B = \{(x_1, x_2, \ldots, x_n) : 0 \le x_1 \le x_2 \le \cdots \le x_n \le 1\}$, then

$$\begin{aligned}
Q(B) &= \int_0^1 \left[\int_0^{x_n} \cdots \left[\int_0^{x_3} \left[\int_0^{x_2} dx_1\right] dx_2\right] \cdots dx_{n-1}\right] dx_n\\
&= \frac{1}{n!},
\end{aligned}$$

where $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1$. $\blacksquare$

[3]For a discussion of multiple integrals in terms of iterated integrals, see Chapter 3 of *Mathematical Comments*.

EXERCISES

1.2.1. Find the union $C_1 \cup C_2$ and the intersection $C_1 \cap C_2$ of the two sets C_1 and C_2, where

(a) $C_1 = \{2, 3, 5, 7\}$, $C_2 = \{1, 3, 5\}$.

(b) $C_1 = \{x : 0 \leq x \leq 3\}$, $C_2 = \{x : 2 < x < 4\}$.

(c) $C_1 = \{(x, y) : 0 < x < 1, 0 < y < 3\}$, $C_2 = \{(x, y) : 0 < x < 2, 2 \leq y < 3\}$.

1.2.2. Find the complement C^c of the set C with respect to the space $\mathcal{C}$ if

(a) $\mathcal{C} = \{x : 0 < x < 2\}$, $C = \left\{x : 0 < x < \frac{2}{3}\right\}$.

(b) $\mathcal{C} = \left\{(x, y, z) : x^2 + 2y^2 + 3z^2 \leq 4\right\}$, $C = \left\{(x, y, z) : x^2 + 2y^2 + 3z^2 < 4\right\}$.

(c) $\mathcal{C} = \left\{(x, y) : x^2 + y^2 \leq 1\right\}$, $C = \{(x, y) : |x| + |y| < 1\}$.

1.2.3. List all possible arrangements of the four letters l, a, m, and b. Let C_1 be the collection of the arrangements in which b is in the first position. Let C_2 be the collection of the arrangements in which a is in the third position. Find the union and the intersection of C_1 and C_2.

1.2.4. Concerning DeMorgan's Laws (1.2.6) and (1.2.7):

(a) Use Venn diagrams to verify the laws.

(b) Show that the laws are true.

(c) Generalize the laws to countable unions and intersections.

1.2.5. By the use of Venn diagrams, in which the space $\mathcal{C}$ is the set of points enclosed by a rectangle containing the circles C_1, C_2, and C_3, compare the following sets. These laws are called the **distributive laws**.

(a) $C_1 \cap (C_2 \cup C_3)$ and $(C_1 \cap C_2) \cup (C_1 \cap C_3)$.

(b) $C_1 \cup (C_2 \cap C_3)$ and $(C_1 \cup C_2) \cap (C_1 \cup C_3)$.

1.2.6. Show that the following sequences of sets, $\{C_k\}$, are nondecreasing, (1.2.16), then find $\lim_{k\to\infty} C_k$.

(a) $C_k = \{x : 1/k \leq x \leq 3 - 1/k\}$, $k = 1, 2, 3, \ldots$.

(b) $C_k = \{(x, y) : 1/k \leq x^2 + y^2 \leq 4 - 1/k\}$, $k = 1, 2, 3, \ldots$.

1.2.7. Show that the following sequences of sets, $\{C_k\}$, are nonincreasing, (1.2.17), then find $\lim_{k\to\infty} C_k$.

(a) $C_k = \{x : 2 - 1/k < x \leq 2\}$, $k = 1, 2, 3, \ldots$.

(b) $C_k = \{x : 2 < x \leq 2 + 1/k\}$, $k = 1, 2, 3, \ldots$.

(c) $C_k = \{(x, y) : 0 \le x^2 + y^2 \le 1/k\}$, $k = 1, 2, 3, \ldots$.

1.2.8. For every one-dimensional set C, define the function $Q(C) = \sum_C f(x)$, where $f(x) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^x$, $x = 0, 1, 2, \ldots$, zero elsewhere. If $C_1 = \{x : x = 0, 2, 4\}$ and $C_2 = \{x : x = 0, 1, 2, \ldots\}$, find $Q(C_1)$ and $Q(C_2)$.
Hint: Recall that $S_n = a + ar + \cdots + ar^{n-1} = a(1 - r^n)/(1 - r)$ and, hence, it follows that $\lim_{n\to\infty} S_n = a/(1 - r)$ provided that $|r| < 1$.

1.2.9. For every one-dimensional set C for which the integral exists, let $Q(C) = \int_C f(x)\,dx$, where $f(x) = \frac{3}{4}(1-x^2)$, $-1 < x < 1$, zero elsewhere; otherwise, let $Q(C)$ be undefined. If $C_1 = \{x : -\frac{1}{3} < x < \frac{1}{3}\}$, $C_2 = \{0\}$, and $C_3 = \{x : -1 < x < 5\}$, find $Q(C_1)$, $Q(C_2)$, and $Q(C_3)$.

1.2.10. For every two-dimensional set C contained in R^2 for which the integral exists, let $Q(C) = \int\int_C (x^2 + y^2)\,dxdy$. If $C_1 = \{(x, y) : -1 \le x \le 1, -1 \le y \le 1\}$, $C_2 = \{(x, y) : -1 \le x = y \le 1\}$, and $C_3 = \{(x, y) : x^2 + y^2 \le 1\}$, find $Q(C_1), Q(C_2)$, and $Q(C_3)$.

1.2.11. Let $\mathcal{C}$ denote the set of points that are interior to, or on the boundary of, a square with opposite vertices at the points (0, 0) and (1, 1). Let $Q(C) = \int\int_C dy\,dx$.

(a) If $C \subset \mathcal{C}$ is the set $\{(x, y) : 0 < y/2 < x < 1/2\}$, compute $Q(C)$.

(b) If $C \subset \mathcal{C}$ is the set $\{(x, y) : 0 < x < 1, x + y = 1\}$, compute $Q(C)$.

(c) If $C \subset \mathcal{C}$ is the set $\{(x, y) : 0 < x/2 < y \le x + 1/4 < 1\}$, compute $Q(C)$.

1.2.12. Let $\mathcal{C}$ be the set of points interior to or on the boundary of a cube with edge of length 1. Moreover, say that the cube is in the first octant with one vertex at the point $(0, 0, 0)$ and an opposite vertex at the point $(1, 1, 1)$. Let $Q(C) = \int\int\int_C dxdydz$.

(a) If $C \subset \mathcal{C}$ is the set $\{(x, y, z) : 0 < x < y < z < 1\}$, compute $Q(C)$.

(b) If C is the subset $\{(x, y, z) : 0 < x = y = z < 1\}$, compute $Q(C)$.

1.2.13. Let C denote the set $\{(x, y, z) : x^2 + y^2 + z^2 \le 1\}$. Using spherical coordinates, evaluate

$$Q(C) = \int\int\int_C \sqrt{x^2 + y^2 + z^2}\,dxdydz.$$

1.2.14. To join a certain club, a person must be either a statistician or a mathematician or both. Of the 35 members in this club, 25 are statisticians and 17 are mathematicians. How many persons in the club are both a statistician and a mathematician?

1.2.15. After a hard-fought football game, it was reported that, of the 11 starting players, 7 hurt a hip, 5 hurt an arm, 7 hurt a knee, 3 hurt both a hip and an arm, 3 hurt both a hip and a knee, 2 hurt both an arm and a knee, and 1 hurt all three. Comment on the accuracy of the report.

1.3 The Probability Set Function

Given an experiment, let $\mathcal{C}$ denote the sample space of all possible outcomes. As discussed in Section 1.1, we are interested in assigning probabilities to events, i.e., subsets of $\mathcal{C}$. What should be our collection of events? If $\mathcal{C}$ is a finite set, then we could take the set of all subsets as this collection. For infinite sample spaces, though, with assignment of probabilities in mind, this poses mathematical technicalities that are better left to a course in probability theory. We assume that in all cases, the collection of events is sufficiently rich to include all possible events of interest and is closed under complements and countable unions of these events. Using DeMorgan's Laws, (1.2.6)–(1.2.7), the collection is then also closed under countable intersections. We denote this collection of events by $\mathcal{B}$. Technically, such a collection of events is called a **σ-field** of subsets.

Now that we have a sample space, $\mathcal{C}$, and our collection of events, $\mathcal{B}$, we can define the third component in our probability space, namely a probability set function. In order to motivate its definition, we consider the relative frequency approach to probability.

Remark 1.3.1. The definition of probability consists of three axioms which we motivate by the following three intuitive properties of relative frequency. Let $\mathcal{C}$ be a sample space and let $A \subset \mathcal{C}$. Suppose we repeat the experiment N times. Then the relative frequency of A is $f_A = \#\{A\}/N$, where $\#\{A\}$ denotes the number of times A occurred in the N repetitions. Note that $f_A \geq 0$ and $f_{\mathcal{C}} = 1$. These are the first two properties. For the third, suppose that A_1 and A_2 are disjoint events. Then $f_{A_1 \cup A_2} = f_{A_1} + f_{A_2}$. These three properties of relative frequencies form the axioms of a probability, except that the third axiom is in terms of countable unions. As with the axioms of probability, the readers should check that the theorems we prove below about probabilities agree with their intuition of relative frequency. ∎

Definition 1.3.1 (Probability)**.** *Let $\mathcal{C}$ be a sample space and let $\mathcal{B}$ be the set of events. Let P be a real-valued function defined on $\mathcal{B}$. Then P is a* **probability set function** *if P satisfies the following three conditions:*

1. *$P(A) \geq 0$, for all $A \in \mathcal{B}$.*
2. *$P(\mathcal{C}) = 1$.*
3. *If $\{A_n\}$ is a sequence of events in $\mathcal{B}$ and $A_m \cap A_n = \phi$ for all $m \neq n$, then*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

A collection of events whose members are pairwise disjoint, as in (3), is said to be a **mutually exclusive** collection and its union is often referred to as a **disjoint union**. The collection is further said to be **exhaustive** if the union of its events is the sample space, in which case $\sum_{n=1}^{\infty} P(A_n) = 1$. We often say that a mutually exclusive and exhaustive collection of events forms a **partition** of $\mathcal{C}$.

A probability set function tells us how the probability is distributed over the set of events, $\mathcal{B}$. In this sense we speak of a distribution of probability. We often drop the word "set" and refer to P as a probability function.

The following theorems give us some other properties of a probability set function. In the statement of each of these theorems, $P(A)$ is taken, tacitly, to be a probability set function defined on the collection of events $\mathcal{B}$ of a sample space $\mathcal{C}$.

Theorem 1.3.1. *For each event $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$.*

Proof: We have $\mathcal{C} = A \cup A^c$ and $A \cap A^c = \phi$. Thus, from (2) and (3) of Definition 1.3.1, it follows that

$$1 = P(A) + P(A^c),$$

which is the desired result. ■

Theorem 1.3.2. *The probability of the null set is zero; that is, $P(\phi) = 0$.*

Proof: In Theorem 1.3.1, take $A = \phi$ so that $A^c = \mathcal{C}$. Accordingly, we have

$$P(\phi) = 1 - P(\mathcal{C}) = 1 - 1 = 0$$

and the theorem is proved. ■

Theorem 1.3.3. *If A and B are events such that $A \subset B$, then $P(A) \leq P(B)$.*

Proof: Now $B = A \cup (A^c \cap B)$ and $A \cap (A^c \cap B) = \phi$. Hence, from (3) of Definition 1.3.1,

$$P(B) = P(A) + P(A^c \cap B).$$

From (1) of Definition 1.3.1, $P(A^c \cap B) \geq 0$. Hence, $P(B) \geq P(A)$. ■

Theorem 1.3.4. *For each $A \in \mathcal{B}$, $0 \leq P(A) \leq 1$.*

Proof: Since $\phi \subset A \subset \mathcal{C}$, we have by Theorem 1.3.3 that

$$P(\phi) \leq P(A) \leq P(\mathcal{C}) \quad \text{or} \quad 0 \leq P(A) \leq 1,$$

the desired result. ■

Part (3) of the definition of probability says that $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint, i.e., $A \cap B = \phi$. The next theorem gives the rule for any two events regardless if they are disjoint or not.

Theorem 1.3.5. *If A and B are events in $\mathcal{C}$, then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof: Each of the sets $A \cup B$ and B can be represented, respectively, as a union of nonintersecting sets as follows:

$$A \cup B = A \cup (A^c \cap B) \quad \text{and} \quad B = (A \cap B) \cup (A^c \cap B). \tag{1.3.1}$$

That these identities hold for all sets A and B follows from set theory. Also, the Venn diagrams of Figure 1.3.1 offer a verification of them.

Thus, from (3) of Definition 1.3.1,

$$P(A \cup B) = P(A) + P(A^c \cap B)$$

and

$$P(B) = P(A \cap B) + P(A^c \cap B).$$

If the second of these equations is solved for $P(A^c \cap B)$ and this result is substituted in the first equation, we obtain

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

This completes the proof. ∎

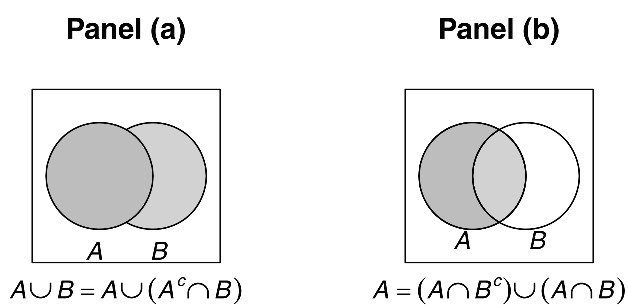


Figure 1.3.1: Venn diagrams depicting the two disjoint unions given in expression (1.3.1). Panel (a) depicts the first disjoint union while Panel (b) shows the second disjoint union.

Example 1.3.1. Let $\mathcal{C}$ denote the sample space of Example 1.1.2. Let the probability set function assign a probability of $\frac{1}{36}$ to each of the 36 points in $\mathcal{C}$; that is, the dice are fair. If $C_1 = \{(1,1), (2,1), (3,1), (4,1), (5,1)\}$ and $C_2 = \{(1,2), (2,2), (3,2)\}$, then $P(C_1) = \frac{5}{36}$, $P(C_2) = \frac{3}{36}$, $P(C_1 \cup C_2) = \frac{8}{36}$, and $P(C_1 \cap C_2) = 0$. ∎

Example 1.3.2. Two coins are to be tossed and the outcome is the ordered pair (face on the first coin, face on the second coin). Thus the sample space may be represented as $\mathcal{C} = \{(H,H), (H,T), (T,H), (T,T)\}$. Let the probability set function assign a probability of $\frac{1}{4}$ to each element of $\mathcal{C}$. Let $C_1 = \{(H,H), (H,T)\}$ and $C_2 = \{(H,H), (T,H)\}$. Then $P(C_1) = P(C_2) = \frac{1}{2}$, $P(C_1 \cap C_2) = \frac{1}{4}$, and, in accordance with Theorem 1.3.5, $P(C_1 \cup C_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$. ∎

For a finite sample space, we can generate probabilities as follows. Let $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$ be a finite set of m elements. Let $p_1, p_2, \ldots, p_m$ be fractions such that

$$0 \le p_i \le 1 \text{ for } i = 1, 2, \ldots, m \text{ and } \textstyle\sum_{i=1}^m p_i = 1. \tag{1.3.2}$$

Suppose we define P by

$$P(A) = \sum_{x_i \in A} p_i, \quad \text{for all subsets } A \text{ of } \mathcal{C}. \tag{1.3.3}$$

Then $P(A) \ge 0$ and $P(\mathcal{C}) = 1$. Further, it follows that $P(A \cup B) = P(A) + P(B)$ when $A \cap B = \phi$. Therefore, P is a probability on $\mathcal{C}$. For illustration, each of the following four assignments forms a probability on $\mathcal{C} = \{1, 2, \ldots, 6\}$. For each, we also compute $P(A)$ for the event $A = \{1, 6\}$.

$$\begin{aligned}
&p_1 = p_2 = \cdots = p_6 = \frac{1}{6}; \quad P(A) = \frac{1}{3}. && (1.3.4)\\
&p_1 = p_2 = 0.1, p_3 = p_4 = p_5 = p_6 = 0.2; \quad P(A) = 0.3.\\
&p_i = \frac{i}{21}, \quad i = 1, 2, \ldots, 6; \quad P(A) = \frac{7}{21}.\\
&p_1 = \frac{3}{\pi}, p_2 = 1 - \frac{3}{\pi}, p_3 = p_4 = p_5 = p_6 = 0.0; \quad P(A) = \frac{3}{\pi}.
\end{aligned}$$

Note that the individual probabilities for the first probability set function, (1.3.4), are the same. This is an example of the equilikely case which we now formally define.

Definition 1.3.2 (Equilikely Case)**.** *Let $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$ be a finite sample space. Let $p_i = 1/m$ for all $i = 1, 2, \ldots, m$ and for all subsets A of $\mathcal{C}$ define*

$$P(A) = \sum_{x_i \in A} \frac{1}{m} = \frac{\#(A)}{m},$$

where $\#(A)$ denotes the number of elements in A. Then P is a probability on $\mathcal{C}$ and it is refereed to as the **equilikely case***.* ■

Equilikely cases are frequently probability models of interest. Examples include: the flip of a fair coin; five cards drawn from a well shuffled deck of 52 cards; a spin of a fair spinner with the numbers 1 through 36 on it; and the upfaces of the roll of a pair of balanced dice. For each of these experiments, as stated in the definition, we only need to know the number of elements in an event to compute the probability of that event. For example, a card player may be interested in the probability of getting a pair (two of a kind) in a hand of five cards dealt from a well shuffled deck of 52 cards. To compute this probability, we need to know the number of five card hands and the number of such hands which contain a pair. Because the equilikely case is often of interest, we next develop some counting rules which can be used to compute the probabilities of events of interest.

1.3.1 Counting Rules

We discuss three counting rules that are usually discussed in an elementary algebra course.

The first rule is called the mn-**rule** (m times n-rule), which is also called the **multiplication rule**. Let $A = \{x_1, x_2, \ldots, x_m\}$ be a set of m elements and let $B = \{y_1, y_2, \ldots, y_n\}$ be a set of n elements. Then there are mn ordered pairs, (x_i, y_j), $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, of elements, the first from A and the second from B. Informally, we often speak of ways, here. For example there are five roads (ways) between cities I and II and there are ten roads (ways) between cities II and III. Hence, there are $5 * 10 = 50$ ways to get from city I to city III by going from city I to city II and then from city II to city III. This rule extends immediately to more than two sets. For instance, suppose in a certain state that driver license plates have the pattern of three letters followed by three numbers. Then there are $26^3 * 10^3$ possible license plates in this state.

Next, let A be a set with n elements. Suppose we are interested in k-tuples whose components are elements of A. Then by the extended mn rule, there are $n \cdot n \cdots n = n^k$ such k-tuples whose components are elements of A. Next, suppose $k \leq n$ and we are interested in k-tuples whose components are distinct (no repeats) elements of A. There are n elements from which to choose for the first component, $n-1$ for the second component, $\ldots$, $n-(k-1)$ for the kth. Hence, by the mn rule, there are $n(n-1)\cdots(n-(k-1))$ such k-tuples with distinct elements. We call each such k-tuple a **permutation** and use the symbol P_k^n to denote the number of k permutations taken from a set of n elements. This number of permutations, P_k^n is our second counting rule. We can rewrite it as

$$P_k^n = n(n-1)\cdots(n-(k-1)) = \frac{n!}{(n-k)!}. \tag{1.3.5}$$

Example 1.3.3 (Birthday Problem)**.** Suppose there are n people in a room. Assume that $n < 365$ and that the people are unrelated in any way. Find the probability of the event A that at least 2 people have the same birthday. For convenience, assign the numbers 1 though n to the people in the room. Then use n-tuples to denote the birthdays of the first person through the nth person in the room. Using the mn-rule, there are 365^n possible birthday n-tuples for these n people. This is the number of elements in the sample space. Now assume that birthdays are equilikely to occur on any of the 365 days. Hence, each of these n-tuples has probability 365^{-n}. Notice that the complement of A is the event that all the birthdays in the room are distinct; that is, the number of n-tuples in A^c is P_n^{365}. Thus, the probability of A is

$$P(A) = 1 - \frac{P_n^{365}}{365^n}.$$

For instance, if $n = 2$ then $P(A) = 1 - (365 * 364)/(365^2) = 0.0027$. This formula is not easy to compute by hand. The following R function[4] computes the $P(A)$ for the input n and it can be downloaded at the sites mentioned in the Preface.

[4]An R primer for the course is found in Appendix B.