SECOND EDITION

Probability and Random Processes

Venkatarama Krishnan
with contribution from
Kavitha Chandra

WILEY

PROBABILITY AND RANDOM PROCESSES

PROBABILITY AND RANDOM PROCESSES

SECOND EDITION

VENKATARAMA KRISHNAN
Professor Emeritus of Electrical and Computer Engineering
Center for Advanced Computation and Telecommunications
University of Massachusetts Lowell

With Contribution from

KAVITHA CHANDRA
Professor of Electrical and Computer Engineering
Center for Advanced Computation and Telecommunications
University of Massachusetts Lowell

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Krishnan, Venkatarama, 1929–
Probability and random processes / Venkatarama Krishnan. – Second edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-92313-9 (cloth : alk. paper)
1. Probabilities. 2. Stochastic processes. 3. Engineering–Statistical methods. 4. Science–Statistical methods. I. Title.
QA273.K74 2015
519.2–dc23

2015023986

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

2 2016

श्रीराम राम रामेति रमे रामे मनोरमे ।
सहस्रनाम तत्तुल्यं राम नाम वरानने ॥

Vishnu sahasranamam

Shree rāma rāma rāmēti
ramē rāmē manoramē
Sahasra nāmatathtulyam
rāmanāma varānanē

This book is respectfully dedicated to the
memory of my mother
Dharmambal Venkataraman
and the memory of my father
B. Venkataraman

மன்னனும் மாசறக்கற்றோனும் சீர்தூக்கின்
மன்னனிற்க்கற்றோன் சிறப்புடையோன்
மன்னர்க்குத் தன்தேசமல்லார்ச் சிறப்பில்லை
கற்றவனுக்குச் சென்றவிடெமல்லாம் சிறப்பு

Avvaiyar Tamil Poet

If a king and a well-learned person are balanced
A well-learned person weighs more than the king
Whereas the king is revered only in his country
A well-learned person is valued wherever he goes

On a well-learned person

CONTENTS

Preface for the Second Edition xii

Preface for the First Edition xiv

1 Sets, Fields, and Events 1

1.1 Set Definitions, 1
1.2 Set Operations, 2
1.3 Set Algebras, Fields, and Events, 5

2 Probability Space and Axioms 7

2.1 Probability Space, 7
2.2 Conditional Probability, 9
2.3 Independence, 11
2.4 Total Probability and Bayes' Theorem, 12

3 Basic Combinatorics 16

3.1 Basic Counting Principles, 16
3.2 Permutations, 16
3.3 Combinations, 18

4 Discrete Distributions 23

4.1 Bernoulli Trials, 23
4.2 Binomial Distribution, 23
4.3 Multinomial Distribution, 26
4.4 Geometric Distribution, 26
4.5 Negative Binomial Distribution, 27
4.6 Hypergeometric Distribution, 28
4.7 Poisson Distribution, 30
4.8 Newton–Pepys Problem and its Extensions, 33
4.9 Logarithmic Distribution, 40
4.9.1 Finite Law (Benford's Law), 40
4.9.2 Infinite Law, 43
4.10 Summary of Discrete Distributions, 44

5 Random Variables **45**

5.1 Definition of Random Variables, 45
5.2 Determination of Distribution and Density Functions, 46
5.3 Properties of Distribution and Density Functions, 50
5.4 Distribution Functions from Density Functions, 51

6 Continuous Random Variables and Basic Distributions **54**

6.1 Introduction, 54
6.2 Uniform Distribution, 54
6.3 Exponential Distribution, 55
6.4 Normal or Gaussian Distribution, 57

7 Other Continuous Distributions **63**

7.1 Introduction, 63
7.2 Triangular Distribution, 63
7.3 Laplace Distribution, 63
7.4 Erlang Distribution, 64
7.5 Gamma Distribution, 65
7.6 Weibull Distribution, 66
7.7 Chi-Square Distribution, 67
7.8 Chi and Other Allied Distributions, 68
7.9 Student-*t* Density, 71
7.10 Snedecor *F* Distribution, 72
7.11 Lognormal Distribution, 72
7.12 Beta Distribution, 73
7.13 Cauchy Distribution, 74
7.14 Pareto Distribution, 75
7.15 Gibbs Distribution, 75
7.16 Mixed Distributions, 75
7.17 Summary of Distributions of Continuous Random Variables, 76

8 Conditional Densities and Distributions **78**

8.1 Conditional Distribution and Density for $P\{A\} \neq 0$, 78
8.2 Conditional Distribution and Density for $P\{A\} = 0$, 80
8.3 Total Probability and Bayes' Theorem for Densities, 83

9 Joint Densities and Distributions **85**

9.1 Joint Discrete Distribution Functions, 85
9.2 Joint Continuous Distribution Functions, 86
9.3 Bivariate Gaussian Distributions, 90

10 Moments and Conditional Moments **91**

10.1 Expectations, 91
10.2 Variance, 92
10.3 Means and Variances of Some Distributions, 93
10.4 Higher-Order Moments, 94
10.5 Correlation and Partial Correlation Coefficients, 95
10.5.1 Correlation Coefficients, 95
10.5.2 Partial Correlation Coefficients, 106

11 Characteristic Functions and Generating Functions **108**

11.1 Characteristic Functions, 108
11.2 Examples of Characteristic Functions, 109
11.3 Generating Functions, 111
11.4 Examples of Generating Functions, 112
11.5 Moment Generating Functions, 113
11.6 Cumulant Generating Functions, 115
11.7 Table of Means and Variances, 116

12 Functions of a Single Random Variable **118**

12.1 Random Variable $g(X)$, 118
12.2 Distribution of $Y = g(X)$, 119
12.3 Direct Determination of Density $f_Y(y)$ from $f_X(x)$, 129
12.4 Inverse Problem: Finding $g(X)$ given $f_X(x)$ and $f_Y(y)$, 132
12.5 Moments of a Function of a Random Variable, 133

13 Functions of Multiple Random Variables **135**

13.1 Function of Two Random Variables, $Z = g(X,Y)$, 135
13.2 Two Functions of Two Random Variables, $Z = g(X,Y)$, $W = h(X,Y)$, 143
13.3 Direct Determination of Joint Density $f_{ZW}(z,w)$ from $f_{XY}(x,y)$, 146
13.4 Solving $Z = g(X,Y)$ Using an Auxiliary Random Variable, 150
13.5 Multiple Functions of Random Variables, 153

14 Inequalities, Convergences, and Limit Theorems **155**

14.1 Degenerate Random Variables, 155
14.2 Chebyshev and Allied Inequalities, 155
14.3 Markov Inequality, 158
14.4 Chernoff Bound, 159
14.5 Cauchy–Schwartz Inequality, 160
14.6 Jensen's Inequality, 162
14.7 Convergence Concepts, 163
14.8 Limit Theorems, 165

15 Computer Methods for Generating Random Variates **169**

15.1 Uniform-Distribution Random Variates, 169
15.2 Histograms, 170
15.3 Inverse Transformation Techniques, 172
15.4 Convolution Techniques, 178
15.5 Acceptance–Rejection Techniques, 178

16 Elements of Matrix Algebra **181**

16.1 Basic Theory of Matrices, 181
16.2 Eigenvalues and Eigenvectors of Matrices, 186
16.3 Vector and Matrix Differentiation, 190
16.4 Block Matrices, 194

17 Random Vectors and Mean-Square Estimation **196**

17.1 Distributions and Densities, 196

17.2 Moments of Random Vectors, 200
17.3 Vector Gaussian Random Variables, 204
17.4 Diagonalization of Covariance Matrices, 207
17.5 Simultaneous Diagonalization of Covariance Matrices, 209
17.6 Linear Estimation of Vector Variables, 210

18 Estimation Theory 212

18.1 Criteria of Estimators, 212
18.2 Estimation of Random Variables, 213
18.3 Estimation of Parameters (Point Estimation), 218
18.4 Interval Estimation (Confidence Intervals), 225
18.5 Hypothesis Testing (Binary), 231
18.6 Bayesian Estimation, 238

19 Random Processes 250

19.1 Basic Definitions, 250
19.2 Stationary Random Processes, 258
19.3 Ergodic Processes, 269
19.4 Estimation of Parameters of Random Processes, 273
19.4.1 Continuous-Time Processes, 273
19.4.2 Discrete-Time Processes, 280
19.5 Power Spectral Density, 287
19.5.1 Continuous Time, 287
19.5.2 Discrete Time, 294
19.6 Adaptive Estimation, 298

20 Classification of Random Processes 320

20.1 Specifications of Random Processes, 320
20.1.1 Discrete-State Discrete-Time (DSDT) Process, 320
20.1.2 Discrete-State Continuous-Time (DSCT) Process, 320
20.1.3 Continuous-State Discrete-Time (CSDT) Process, 320
20.1.4 Continuous-State Continuous-Time (CSCT) Process, 320
20.2 Poisson Process, 321
20.3 Binomial Process, 329
20.4 Independent Increment Process, 330
20.5 Random-Walk Process, 333
20.6 Gaussian Process, 338
20.7 Wiener Process (Brownian Motion), 340
20.8 Markov Process, 342
20.9 Markov Chains, 347
20.10 Birth and Death Processes, 357
20.11 Renewal Processes and Generalizations, 366
20.12 Martingale Process, 370
20.13 Periodic Random Process, 374
20.14 Aperiodic Random Process (Karhunen–Loeve Expansion), 377

21 Random Processes and Linear Systems 383

21.1 Review of Linear Systems, 383
21.2 Random Processes through Linear Systems, 385
21.3 Linear Filters, 393
21.4 Bandpass Stationary Random Processes, 401

22 Wiener and Kalman Filters **413**

22.1 Review of Orthogonality Principle, 413
22.2 Wiener Filtering, 414
22.3 Discrete Kalman Filter, 425
22.4 Continuous Kalman Filter, 433

23 Probability Modeling in Traffic Engineering **437**

23.1 Introduction, 437
23.2 Teletraffic Models, 437
23.3 Blocking Systems, 438
23.4 State Probabilities for Systems with Delays, 440
23.5 Waiting-Time Distribution for M/M/c/∞ Systems, 441
23.6 State Probabilities for M/D/c Systems, 443
23.7 Waiting-Time Distribution for M/D/c/∞ System, 446
23.8 Comparison of M/M/c and M/D/c, 448
References, 451

24 Probabilistic Methods in Transmission Tomography **452**

24.1 Introduction, 452
24.2 Stochastic Model, 453
24.3 Stochastic Estimation Algorithm, 455
24.4 Prior Distribution $P\{\mathbf{M}\}$, 457
24.5 Computer Simulation, 458
24.6 Results and Conclusions, 460
24.7 Discussion of Results, 462
References, 462

APPENDICES

A A Fourier Transform Tables **463**

B Cumulative Gaussian Tables **467**

C Inverse Cumulative Gaussian Tables **472**

D Inverse Chi-Square Tables **474**

E Inverse Student-*t* Tables **481**

F Cumulative Poisson Distribution **484**

G Cumulative Binomial Distribution **488**

H Computation of Roots of $D(z) = 0$ **494**

References **495**

Index **498**

PREFACE FOR THE SECOND EDITION

In this second edition, additional sections have been incorporated to make it a more complete tome. In Chapter 4, a historical account of how Isaac Newton solved a binomial probability problem posed (1693) by the diarist Samuel Pepys when the field of probability was relatively unknown is described. In Chapter 10, Pearson correlation coefficient introduced earlier in the chapter has been expanded to include Spearman and Kendall correlation coefficients with their properties and usages discussed. In Chapter 14, convergences have been generalized. In Chapter 19 on estimation, two adaptive estimation techniques such as Recursive Least Squares and Least Mean Squares have been added. In Chapter 20 on random processes, additional topics such as Birth and Death Processes and Renewal Processes useful for analyzing queuing have been incorporated. Chapter 23 is a new chapter on Probability Modeling of Teletraffic Engineering written by my colleague Kavitha Chandra discussing various probability models in teletraffic engineering.

This new edition now contains 455 carefully detailed figures and 377 representative examples, some with multiple solutions with every step explained clearly enhancing the clarity of presentation to yield a more comprehensive reference tome. It explains in detail the essential topics of applied mathematical functions to problems that engineers and researchers solve daily in the course of their work. The usual topics like set theory, combinatorics, random variables, discrete and continuous probability distribution functions, moments, and convergence of random variables, autocovariance and cross covariance functions, stationarity concepts, Wiener and Kalman filtering, and tomographic imaging of the first edition have all been retained. Graphical Fourier transform tables and a number of probability tables with accuracy up to nine decimal places are given in the seven appendices. A new eighth appendix on finding the roots of the probability generating function has also been added. The index has been carefully prepared to enhance the utility of this reference.

With these added material on theory and applications of probability, the second edition of *Probability and Random Processes* is a more comprehensive reference tome for practicing scientists and engineers in economics, science, and technology.

As in the first edition, the additional graphs in this edition were also created with Mathcad software. All algebraic calculations were also verified with Mathcad software. Mathcad is a registered trademark of Parametric Technology Corporation Inc., http://www.ptc.com/product/mathcad/

Acknowledgments

Since the publication of the first edition of this book, many other books on Probability have appeared and they have influenced me to a considerable extent in the choice of additional topics. I once again extend my scientific debt of gratitude to all these authors.

I acknowledge with great pleasure the support given to me by my colleague Kavitha Chandra who not only contributed a chapter but also helped me to a considerable extent in the preparation of the manuscript. My grateful thanks are due to Kari Capone who encouraged me to bring out a second edition and did the entire spade work to make this project a reality. The courteous and thoughtful encouragement by Alex Castro during the final phases of the preparation of this manuscript is a pleasure for me to acknowledge. The efficiency with which Ramya Srinivasan, the production editor, went about the production job is indeed a pleasure to record. She also has the knack of smoothing ruffled feathers, which had kept me in an even keel. Finally, I also acknowledge with thanks the clean and efficient job performed by F. Pascal Raj, the assistant account manager and his team with their proactive suggestions.

The unequivocal support given to me by my wife Kamala has been a great source of comfort.

This is my third book with Wiley and I am always amazed at the way they can create a book with clockwork precision out of the rough manuscript.

VENKATARAMA KRISHNAN

Chelmsford, MA

March 2015

PREFACE FOR THE FIRST EDITION

Many good textbooks exist on probability and random processes written at the undergraduate level to the research level. However, there is no one handy and ready book that explains most of the essential topics, such as random variables and most of their frequently used discrete and continuous probability distribution functions; moments, transformation, and convergences of random variables; characteristic and generating functions; estimation theory and the associated orthogonality principle; vector random variables; random processes and their autocovariance and cross-covariance functions; stationarity concepts; and random processes through linear systems and the associated Wiener and Kalman filters. Engineering practitioners and students alike have to delve through several books to get the required formulas or tables either to complete a project or to finish a homework assignment. This book may alleviate this diffculty to some extent and provide access to a compendium of most distribution functions used by communication engineers, queuing theory specialists, signal processing engineers, biomedical engineers, and physicists. Probability tables with accuracy up to nine decimal places are given in the appendixes to enhance the utility of this book. A particular feature is the presentation of commonly occurring Fourier transforms where both the time and frequency functions are drawn to scale.

Most of the theory has been explained with figures drawn to scale. To understand the theory better, more than 300 examples are given with every step explained clearly. Following the adage that a figure is worth more than a thousand words, most of the examples are also illustrated with figures drawn to scale, resulting in more than 400 diagrams. This book will be of particular value to graduate and undergraduate students in electrical, computer, and civil engineering as well as students in physics and applied mathematics for solving homework assignments and projects. It will certainly be useful to communication and signal processing engineers, and computer scientists in an industrial setting. It will also serve as a good reference for research workers in biostatistics and financial market analysis.

The salient features of this book are

- Functional and statistical independence of random variables are explained.
- Ready reference to commonly occurring density and distribution functions and their means and variances is provided.
- A section on Benford's logarithmic law, which is used in detecting tax fraud, is included.
- More than 300 examples, many of them solved in different ways to illustrate the theory and various applications of probability, are presented.
- Most examples have been substantiated with graphs drawn to scale.
- More than 400 figures have been drawn to scale to improve the clarity of understanding the theory and examples.
- Bounds on the tails of Gaussian probability have been carefully developed.
- A chapter has been devoted to computer generation of random variates.
- Another chapter has been devoted to matrix algebra so that some estimation problems such as the Kalman filter can be cast in matrix framework.
- Estimation problems have been given considerable exposure.
- Random processes defined and classified.
- A section on martingale processes with examples has been included.

- Markov chains with interesting examples have been discussed.
- Wiener and Kalman filtering have been discussed with a number of examples.
- Important properties are summarized in tables.
- The final chapter is on applications of probability to tomographic imaging.
- The appendixes consist of probability tables with nine-place decimal accuracy for the most common probability distributions.
- An extremely useful feature are the Fourier transform tables with both time and frequency functions graphed carefully to scale.

After the introduction of functional independence in Section 1.2, the differences between functional and statistical independences are discussed in Section 2.3 and clarified in Example 2.3.2. The foundation for permutations and combinations are laid out in Chapter 3, followed by discrete distributions in Chapter 4 with an end-of-chapter summary of discrete distributions. Random variables are defined in Chapter 5, and most of the frequently occurring continuous distributions are explained in Chapters 6 and 7. Section 6.4 discusses the new bounds on Gaussian tails. A comprehensive table is given at the end of Chapter 7 for all the continuous distributions and densities. After discussion of conditional densities, joint densities, and moments in Chapters 8–10, characteristic and other allied functions are explained in Chapter 11 with a presentation of an end-of-chapter table of means and variances of all discrete and continuous distributions. Functions of random variables are discussed in Chapters 12 and 13 with numerous examples. Chapter 14 discusses the various bounds on probabilities along with some commonly occurring inequalities. Computer generation of random variates is presented in Chapter 15. Elements of matrix algebra along with vector and matrix differentiations are considered in Chapter 16. Vector random variables and diagonalization of covariance matrices and simultaneous diagonalization of two covariance matrices are taken up in Chapter 17. Estimation and allied hypothesis testing are discussed in Chapter 18. After random processes are explained in Chapter 19, they are carefully classified in Chapter 20, with Section 20.10 presenting martingale processes that find wide use in financial engineering. Chapter 21 discusses the effects of passing random processes through linear systems. A number of examples illustrate the basic ideas of Wiener and Kalman filters in Chapter 22. The final chapter, Chapter 23, presents a practical application of probability to tomographic imaging. The appendixes include probability tables up to nine decimal places for Gaussian, chi-square, Student-*t*, Poisson, and binomial distributions, and Fourier transform tables with graphs for time and frequency functions carefully drawn to scale.

This book is suitable for students and practicing engineers who need a quick reference to any probability distribution or table. It is also useful for self-study where a number of carefully solved examples illustrate the applications of probability. Almost every topic has solved examples. It can be used as an adjunct to any textbook on probability and may also be prescribed as a textbook with homework problems drawn from other sources.

During the more than four decades that I have been continuously teaching probability and random processes, many authors who have written excellent textbooks have influenced me to a considerable extent. Many of the ideas and concepts that are expanded in this book are the direct result of this influence. I owe a scientific debt of gratitude to all the authors who have contributed to this ubiquitous and exciting field of probability. Some of these authors are listed in the reference, and the list is by no means exhaustive.

Most of the graphs and all the probability tables in this book were created with Mathcad software. All algebraic calculations were verified with Mathcad software. Mathcad and Mathsoft are registered trademarks of Mathsoft Engineering and Education, Inc., http://www.mathcad. com.

While facing extreme personal diffculties in writing this book, the unequivocal support of my entire family have been a source of inspiration in finishing it.

I acknowledge with great pleasure the support given to me by the Wiley staff. George Telecki welcomed the idea of this book; Rachel Witmer did more than her share of keeping me in good humor with her cheerful disposition in spite of the ever-expanding deadlines; and Kellsee Chu who did an excellent job in managing the production of this book. Finally, the enthusiastic support given to me by the Series Editor Emmanuel Desurvire was very refreshing. This is my second book with Wiley, and the skills of their copyeditors and staff who transform highly mathematical manuscript into a finished book continue to amaze me.

I have revised this book several times and corrected errors. Nevertheless, I cannot claim that it is error-free since correcting errors in any book is a convergent process. I sincerely hope that readers will bring to my attention any errors of commission or omission that they may come across.

Chelmsford, Massachusetts
March 2006

VENKATARAMA KRISHNAN

1

SETS, FIELDS, AND EVENTS

1.1 SET DEFINITIONS

The concept of sets play an important role in probability. We will define a set in the following paragraph.

Definition of Set

A *set* is a collection of objects called *elements*. The elements of a set can also be sets. Sets are usually represented by uppercase letters A, and elements are usually represented by lowercase letters a. Thus

$$A = \{a_1, a_2, \ldots, a_n\} \tag{1.1.1}$$

will mean that the set A contains the elements $a_1, a_2, \ldots, a_n$. Conversely, we can write that a_k is an element of A as

$$a_k \in A \tag{1.1.2}$$

and a_k is not an element of A as

$$a_k \notin A \tag{1.1.3}$$

A finite set contains a finite number of elements, for example, $S = \{2,4,6\}$. Infinite sets will have either countably infinite elements such as $A = \{x : x$ is all positive integers$\}$ or uncountably infinite elements such as $B = \{x : x$ is real number $\leq 20\}$.

Example 1.1.1 The set A of all positive integers less than 7 is written as

$$A = \{x : x \text{ is a positive integer} < 7\} : \text{ finite set.}$$

Example 1.1.2 The set N of all positive integers is written as

$$N = \{x : x \text{ is all positive integers}\} : \text{ countably infinite set.}$$

Example 1.1.3 The set R of all real numbers is written as

$$R = \{x : x \text{ is real}\} : \text{ uncountably infinite set.}$$

Example 1.1.4 The set R^2 of real numbers x, y is written as

$$R^2 = \{(x, y) : x \text{ is real, } y \text{ is real}\}$$

Example 1.1.5 The set C of all real numbers x,y such that $x + y \leq 10$ is written as

$$C = \{(x, y) : x + y \leq 10\} : \text{ uncountably infinite set.}$$

Venn Diagram

Sets can be represented graphically by means of a Venn diagram. In this case we assume tacitly that S is a universal set under consideration. In Example 1.1.5, the universal set $S = \{x : x$ is all positive integers$\}$. We shall represent the set A in Example 1.1.1 by means of a Venn diagram of Fig. 1.1.1.

Empty Set

An empty is a set that contains no element. It plays an important role in set theory and is denoted by $\varnothing$. The set $A = \{0\}$ is not an empty set since it contains the element 0.

Probability and Random Processes, Second Edition. Venkatarama Krishnan.

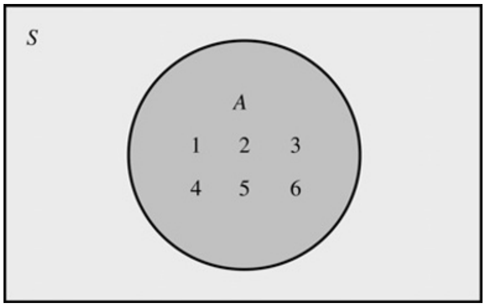


FIGURE 1.1.1

Cardinality

The number of elements in the set A is called the *cardinality* of set A, and is denoted by $|A|$. If it is an infinite set, then the cardinality is ∞.

Example 1.1.6 The cardinality of the set $A = \{2,4,6\}$ is 3, or $|A| = 3$. The cardinality of set $R = \{x : x \text{ is real}\}$ is ∞.

Example 1.1.7 The cardinality of the set $A = \{x : x \text{ is positive integer} <7\}$ is $|A| = 6$.

Example 1.1.8 The cardinality of the set $B = \{x : x \text{ is a real number} <10\}$ is infinity since there are infinite real numbers <10.

Subset

A set B is a subset of A if every element in B is an element of A and is written as $B \subseteq A$. B is a proper subset of A if *every* element of A is not in B and is written as $B \subset A$.

Equality of Sets

Two sets A and B are equal if $B \subseteq A$ and $A \subseteq B$, that is, if every element of A is contained in B and every element of B is contained in A. In other words, sets A and B contain exactly the same elements. Note that this is different from having the same cardinality, that is, containing the *same number* of elements.

Example 1.1.9 The set $B = \{1,3,5\}$ is a proper subset of $A = \{1,2,3,4,5,6\}$, whereas the set $C = \{x : x \text{ is a positive even integer} \leq 6\}$ and the set $D = \{2,4,6\}$ are the same since they contain the same elements. The cardinalities of B, C, and D are 3 and $C = D$.

We shall now represent the sets A and B and the sets C and D in Example 1.1.9 by means of the Venn diagram of Fig. 1.1.2 on a suitably defined universal set S.

Power Set

The power set of any set A is the set of all possible subsets of A and is denoted by $PS(A)$. Every power set of any set

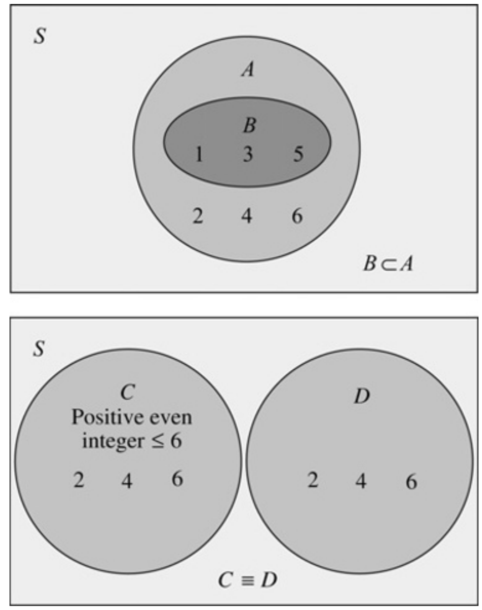


FIGURE 1.1.2

A must contain the set A itself and the empty set $\emptyset$. If n is the cardinality of the set A, then the cardinality of the power set $|PS(A)| = 2^n$.

Example 1.1.10 If the set $A = \{1,2,3\}$ then $PS(A) = \{\emptyset, (1,2,3), (1,2), (2,3), (3,1), (1), (2), (3)\}$. The cardinality $|PS(A)| = 8 = 2^3$.

1.2 SET OPERATIONS

Union

Let A and B be sets belonging to the universal set S. The *union* of sets A and B is another set C whose elements are those that are in either A or B, and is denoted by $A \cup B$. Where there is no confusion, it will also be represented as $A + B$:

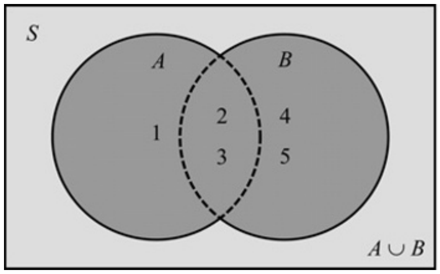
$$A \cup B = A + B = \{x : x \in A \text{ or } x \in B\} \tag{1.2.1}$$

Example 1.2.1 The union of sets $A = \{1,2,3\}$ and $B = \{2,3,4,5\}$ is the set $C = A \cup B = \{1,2,3,4,5\}$.

Intersection

The *intersection* of the sets A and B is another set C whose elements are the same as those in both A and B and is denoted by $A \cap B$. Where there is no confusion, it will also be represented by AB.

$$A \cap B = AB = \{x : x \in A \text{ and } x \in B\} \tag{1.2.2}$$



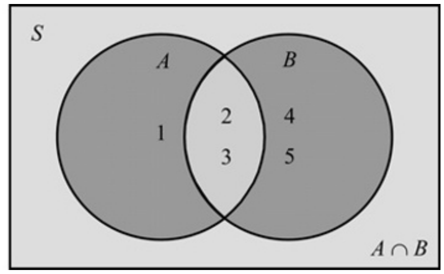


FIGURE 1.2.1

Example 1.2.2 The intersection of the sets A and B in Example 1.2.1 is the set $C = \{2,3\}$ Examples 1.2.1 and 1.2.2 are shown in the Venn diagram of Fig. 1.2.1.

Mutually Exclusive Sets

Two sets A and B are called *mutually exclusive* if their intersection is empty. Mutually exclusive sets are also called *disjoint*.

$$A \cap B = \emptyset \tag{1.2.3}$$

One way to determine whether two sets A and B are mutually exclusive is to check whether set B can occur when set A has already occurred and vice versa. If it cannot, then A and B are mutually exclusive. For example, if a single coin is tossed, the two sets, {heads} and {tails}, are mutually exclusive since {tails} cannot occur when {heads} has already occurred and vice versa.

Independence

We will consider two types of independence. The first is known as *functional independence* [58].[1] Two sets A and B can be called functionally independent if the occurrence of B does not in any way influence the occurrence of A and vice versa. The second one is *statistical independence*, which is a different concept that will be defined later. As an example, the tossing of a coin is functionally independent of the tossing of a die because they do not depend on each other. However, the tossing of a coin and a die are not mutually exclusive since any one can be tossed irrespective of the other. By the same token, pressure and temperature are not functionally independent because the physics of the problem, namely, Boyle's law, connects these quantities. They are certainly not mutually exclusive.

Cardinality of Unions and Intersections

We can now ascertain the cardinality of the union of two sets A and B that are not mutually exclusive. The cardinality of the union $C = A \cup B$ can be determined as follows. If we add the cardinality $|A|$ to the cardinality $|B|$, we have added the cardinality of the intersection $|A \cap B|$ twice. Hence we have to subtract once the cardinality $|A \cap B|$ as shown in Fig. 1.2.2. Or, in other words

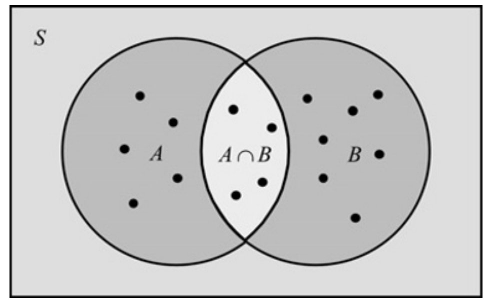


FIGURE 1.2.2

$$|A \cup B| = |A| + |B| - |A \cap B| \tag{1.2.4a}$$

In Fig. 1.2.2 the cardinality $|A| = 9$ and the cardinality $|B| = 11$ and the cardinality $|A \cup B|$ is $11 + 9 - 4 = 16$.

As a corollary, if sets A and B are mutually exclusive, then the cardinality of the union is the sum of the cardinalities; or

$$|A \cup B| = |A| + |B| \tag{1.2.4b}$$

The generalization of this result to an arbitrary union of n sets is called the *inclusion–exclusion* principle, given by

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{\substack{i,j=1\\ i\neq j}}^{n} |A_i \cap A_j| + \sum_{\substack{i,j,k=1\\ i\neq j\neq k}}^{n} |A_i \cap A_j \cap A_k| - (\pm 1)^n \sum_{\substack{i,j,k=1\\ i\neq j\neq k\ldots\neq n}}^{n} |A_i \cap A_j \cap A_k \cdots \cap A_n| \tag{1.2.5a}$$

If the sets $\{A_i\}$ are mutually exclusive, that is, $A_i \cap A_j = \emptyset$ for $i \neq j$, then we have

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| \tag{1.2.5b}$$

[1] Numbers in brackets refer to bibliographic entries in the References section at the end of the book.

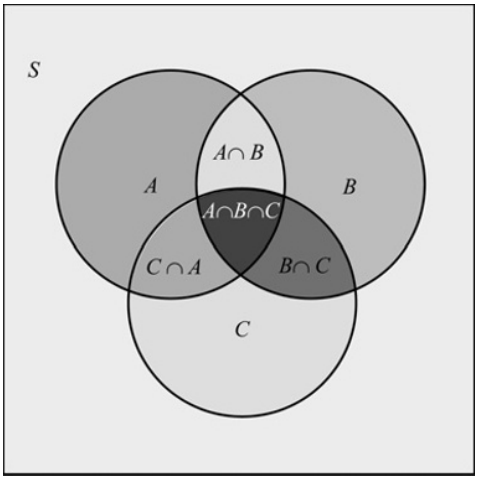


FIGURE 1.2.3

This equation is illustrated in the Venn diagram for $n = 3$ in Fig. 1.2.3, where if $|A \cup B \cup C|$ equals, $|A| + |B| + |C|$ then we have added twice the cardinalites of $|A \cap B|$, $|B \cap C|$ and $|C \cap A|$. However, if we subtract once $|A \cap B|$, $|B \cap C|$, and $|C \cap A|$ and write

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A|$$

then we have subtracted $|A \cap B \cap C|$ thrice instead of twice. Hence, adding $|A \cap B \cap C|$ we get the final result:

$$\begin{aligned} |A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| \\ - |C \cap A| + |A \cap B \cap C| \end{aligned} \tag{1.2.6}$$

Example 1.2.3 In a junior class, the number of students in electrical engineering (EE) is 100, in math (MA) 50, and in computer science (CS) 150. Among these 20 are taking both EE and MA, 25 are taking EE and CS, and 10 are taking MA and CS. Five of them are taking EE, CS and MA. Find the total number of students in the junior class.

From the problem we can identify the sets A = EE students, B = MA students, and C = CS students, and the corresponding intersections are $A \cap B$, $B \cap C$, $C \cap A$, and $\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}$. Here, $|A| = 100$, $|B| = 50$, and $|C| = 150$. We are also given $|A \cap B| = 20$, $|B \cap C| = 10$, and $|C \cap A| = 25$ and finally $|A \cap B \cap C| = 5$. Using Eq. (1.2.6) the total number of students are $100 + 50 + 150 - 20 - 10 - 25 + 5 = 250$.

Complement

If A is a subset of the universal set S, then the complement of A denoted by $\bar{A}$ is the elements in S not contained in A; or

$$\bar{A} = S - A = \{x : x \notin A \subset S \text{ and } x \in S\} \tag{1.2.7}$$

Example 1.2.4 If the universal set $S = \{1,2,3,4,5,6\}$ and $A = \{2,4,5\}$, then the complement of A is given by $\bar{A} = \{1,3,6\}$.

Difference

The complement of a subset $A \subset S$ as given by Eq. (1.2.7) is the difference between the universal set S and the subset A. The difference between any two sets A and B denoted by $A - B$ is the set C containing those elements that are in A but not in B and, $B - A$ is the set D containing those elements in B but not in A:

$$\begin{aligned} C = A - B = \{x : x \in A \text{ and } x \notin B\} \\ D = B - A = \{x : x \in B \text{ and } x \notin A\} \end{aligned} \tag{1.2.8}$$

$A - B$ is also called the *relative complement* of B with respect to A and $B - A$ is called the relative complement of A with respect to B. It can be verified that $A - B \neq B - A$.

Example 1.2.5 If the set A is given as $A = \{2,4,5,6\}$ and B is given as $B = \{5,6,7,8\}$, then the difference $A - B = \{2,4\}$ and the difference $B - A = \{7,8\}$. Clearly, $A - B \neq B - A$ if $A \neq B$.

Symmetric Difference

The symmetric difference between sets A and B is written as $A \,\Delta\, B$ and defined by

$$\begin{aligned} A \Delta B &= (A - B) \cup (B - A) \\ &= \{x : x \in A \text{ and } x \notin B) \text{ or } \{x : x \in B \text{ and } x \notin A) \end{aligned} \tag{1.2.9}$$

Example 1.2.6 The symmetric difference between the set A given by $A = \{2,4,5,6\}$ and B given by $B = \{5,6,7,8\}$ in Example 1.2.5 is $A \,\Delta\, B = \{2,4\} \cup \{7,8\} = \{2,4,7,8\}$.

The difference and symmetric difference for Examples 1.2.5 and 1.2.6 are shown in Fig. 1.2.4.

Cartesian Product

This is a useful concept in set theory. If A and B are two sets, the Cartesian product $A \times B$ is the set of ordered pairs (x,y) such that $x \in A$ and $y \in B$ and is defined by

$$A \times B = \{(x,y) : x \in A,\ y \in B\} \tag{1.2.10}$$

When ordering is considered the cartesian product $A \times B \neq B \times A$. The cardinality of a Cartesian product is the product of the individual cardinalities, or $|A \times B| = |A| \cdot |B|$.

Example 1.2.7 If $A = \{2,3,4\}$ and $B = \{5,6\}$, then $C = A \times B = \{(2,5), (2,6), (3,5), (3,6), (4,5), (4,6)\}$ with cardinality $3 \times 2 = 6$. Similarly, $D = B \times A = \{(5,2), (5,3), (5,4), (6,2), (6,3),$

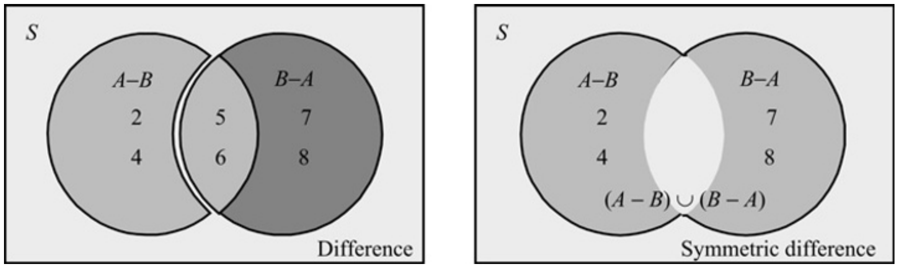


FIGURE 1.2.4

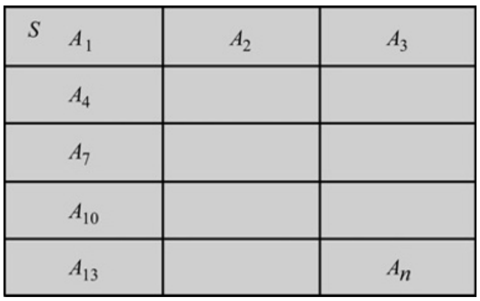


FIGURE 1.2.5

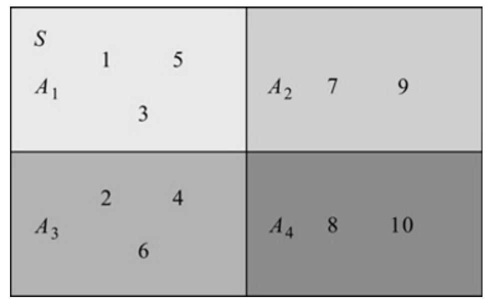


FIGURE 1.2.6

(6,4)} with the same cardinality of 6. However, C and D do not contain the same ordered pairs.

Partition

A *partition* is a collection of disjoint sets $\{A_i, i = 1, \ldots, n\}$ of a set S such that $\bigcup A_i$ over all i equals S and $A_i \cap A_j$, with $i \neq j$ empty. Partitions play a useful role in conditional probability and Bayes' theorem. Figure 1.2.5 shows the concept of a partition.

Example 1.2.8 If set $S = \{1,2,3,4,5,6,7,8,9,10\}$, then the collections of sets $\{A_1, A_2, A_3, A_4\}$ where $A_1 = \{1,3,5\}$, $A_2 = \{7,9\}$, $A_3 = \{2,4,6\}$, $A_4 = \{8,10\}$ is a partition of S as shown in Fig. 1.2.6.

A tabulation of set properties is shown on Table 1.2.1. Among the identities, De Morgan's laws are quite useful in simplifying set algebra.

TABLE 1.2.1 Table of set properties

Property	Equation
Identity	$A \cup \emptyset = A$ $A \cap S = A$
Domination	$A \cap \emptyset = \emptyset$ $A \cup S = S$
Idempotent	$A \cap A = A$ $A \cup A = A$
Complementation	$\bar{\bar{A}} = A$
Commutative	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associative	$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$
Distributive	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
Noncommutative	$A - B \neq B - A$
De Morgan's law	$\overline{A \cup B} = \bar{A} \cap \bar{B}$ $\overline{A \cap B} = \bar{A} \cup \bar{B}$

1.3 SET ALGEBRAS, FIELDS, AND EVENTS

Boolean Field

We shall now consider a universal set S and a collection $\mathscr{F}$ of subsets of S consisting of the sets $\{A_i\!: i = 1,2, \ldots, n, n + 1, \ldots\}$. The collection $\mathscr{F}$ is called a *Boolean field* if the following two conditions are satisfied:

1. If $A_i \in \mathscr{F}$, then $\bar{A}_i \in \mathscr{F}$.
2. If $A_1 \in \mathscr{F}$ and $A_2 \in \mathscr{F}$, then $A_1 \cup A_2 \in \mathscr{F}$, and hence if $\{A_i, i = 1, \ldots, n\} \in \mathscr{F}$, then $\cup_{i=1}^{n} A_i \in \mathscr{F}$.

Let us see the consequences of these conditions being satisfied. If $A_1 \in \mathscr{F}$ and $\bar{A}_1 \in \mathscr{F}$, then $A_1 \cup \bar{A}_1 = S \in \mathscr{F}$. If $S \in \mathscr{F}$, then $\bar{S} = \emptyset \in \mathscr{F}$. If $A_1 \cup A_2 \in \mathscr{F}$, then by condition 1 $\overline{A_1 \cup A_2} \in \mathscr{F}$, and by De Morgan's law, $\bar{A}_1 \cap \bar{A}_2 \in \mathscr{F}$. Hence $A_1 \cap A_2 \in \mathscr{F}$. Also $A_1 - A_2 \in \mathscr{F}$ and $(A_1 - A_2) \cup (A_2 - A_1) \in \mathscr{F}$. Thus the conditions listed above are suffcient for any field $\mathscr{F}$ to be *closed* under all set operations.

Sigma Field

The preceding definition of a field covers only finite set operations. Condition 3 for finite additivity may not hold for infinite additivity. For example, the sum of $1 + 2 + 2^2 + 2^3 + 2^4 = (2^5 - 1)/(2 - 1) = 31$, but the infinite sum $1 + 2 + 2^2 + 2^3 + 2^4 + \cdots$ diverges. Hence there is a need to extend finite additivity concept to infinite set operations. Thus, we have a σ field if the following extra condition is imposed for infinite additivity:

4. If $A_1, A_2, \ldots A_n, \ldots \in \mathscr{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$

Many σ fields may contain the subsets $\{A_i\}$ of S, but the smallest σ field containing the subsets $\{A_i\}$ is called the *Borel σ field*. The smallest σ field for S by itself is $\mathscr{F} = \{S, \emptyset\}$.

Example 1.3.1 We shall assume that $S = \{1,2,3,4,5,6\}$. Then the collection of the following subsets of S, $\mathscr{F} = \{S, \emptyset, (1,2,3), (4,5,6)\}$ is a field since it satisfies all the set operations such as $(1,2,3) \cup (4,5,6) = S$, $\overline{(1,2,3)} = (4,5,6)$. However, the collection of the following subsets of S, $\mathscr{F}_1 = \{S, \emptyset, (1,2,3), (4,5,6), (2)\}$ will not constitute a field because $(2) \cup (4,5,6) = (2,4,5,6)$ is not in the field. But we can adjoin the missing sets and make $\mathscr{F}_1$ into a field. This is known as *completion*. In the example above, if we adjoin the collection of sets $\mathscr{F}_2 = \{(2,4,5,6), (1,3)\}$ to $\mathscr{F}_1$, then the resulting collection $\mathscr{F} = \mathscr{F}_1 \cup \mathscr{F}_2 = \{S, \emptyset, (1,2,3), (4,5,6), \{2\}, (2,4,5,6), (1,3)\}$ is a field.

Event

Given the universal set S and the associated σ field $\mathscr{F}$, all subsets of S belonging to the field $\mathscr{F}$ are called *events*. All events are subsets of S and can be assigned a probability, but not all subsets of S are events unless they belong to an associated σ field.

2

PROBABILITY SPACE AND AXIOMS

2.1 PROBABILITY SPACE

A *probability space* is defined as the triplet $\{S, \mathscr{F}, P\}$, where S is the sample space, $\mathscr{F}$ is the field of subsets of S, and P is a probability measure. We shall now define these elements.

Sample Space S

Finest-grain, mutually exclusive, and collectively exhaustive listing of all possible outcomes of a mathematical experiment is called the *sample space*. The outcomes can be either finite, countably infinite, or uncountably infinite. If S is countable, we call it a *discrete sample space*; if it is uncountable, we call it *continuous sample space.*

Examples of Discrete Sample Space

Example 2.1.1 When a coin is tossed, the outcomes are heads $\{H\}$ or tails $\{T\}$. Thus the sample space $S = \{H,T\}$.

Example 2.1.2 When a die is tossed, the outcomes are 1,2,3,4,5,6. Thus the sample space is $S = \{1,2,3,4,5,6\}$.

Example 2.1.3 The set of all positive integers is countable, and the sample space is $S = \{x : x \text{ is a positive integer}\}$

Examples of Continuous Sample Space

Example 2.1.4 The set of all real numbers between 0 and 2 is the sample space $S = \{x : 0 \le x \le 2\}$.

Example 2.1.5 The set of all real numbers x and y such that x is between 0 and 1 and y is between 1 and 2 is given by $S = \{x, y : 0 \le x \le 1, 1 \le y \le 2\}$.

Probability and Random Processes, Second Edition. Venkatarama Krishnan.

The continuous sample space given in Examples 2.1.4 and 2.1.5 is shown in Fig. 2.1.1.

Field $\mathscr{F}$

A *field* has already been defined in Section 1.3. We shall assume that a proper Borel σ field can be defined for all discrete and continuous sample spaces.

Probability Measure P

A *probability measure* is defined on the field $\mathscr{F}$ of events of the sample space S. It is a mapping of the sample space to the real line between 0 and 1 such that it satisfies the following three Kolmogorov axioms for all events belonging to the field $\mathscr{F}$:

Axiom 1:

$$\text{If A belongs to } \mathscr{F}, \text{ then } P(A) \ge 0 : \text{nonnegativity.} \tag{2.1.1}$$

Axiom 2: The probability of the sample space equals 1.

$$P(S) = 1 : \text{normalization.} \tag{2.1.2}$$

Axiom 3a: If A and B are disjoint sets belonging to $\mathscr{F}$, that is, $A \cap B = \emptyset$, then

$$P\{A \cup B\} = P(A) + P(B) : \text{Boolen additivity.} \tag{2.1.3}$$

Axiom 3b: If $\{A_1, A_2, \dots\}$ is a sequence of sets belonging to the field $\mathscr{F}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

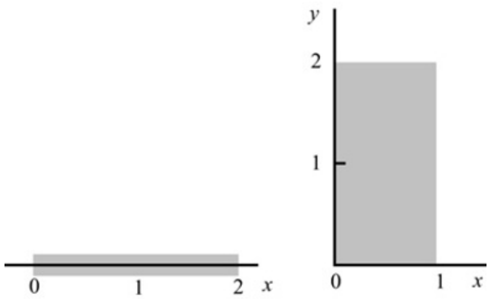


FIGURE 2.1.1

$$P\{A_1 \cup A_2 \cup \cdots \cup A_i \cup \cdots\} = \sum_{i=1}^{\infty} P\{A_j\} : \text{sigma additivity.} \tag{2.1.4}$$

From these axioms we can write the following corollaries:

Corollary 1:

$$P\{\varnothing\} = 0 \tag{2.1.5}$$

Note that the probability of the empty event is zero, but if the probability of an event A is zero, we cannot conclude that $A = \varnothing$. All sets with zero probability are defined as *null sets*. An empty set is also a null set, but null sets need not be empty.

Corollary 2: Since $A \cup \bar{A} = S$ and $A \cap \bar{A} = \varnothing$, we can write

$$P(\bar{A}) = 1 - P(A) \tag{2.1.6}$$

Corollary 3: For any A and B that are subsets of S, not mutually exclusive, we can write

$$P\{A \cup B\} = P\{A \cup [\bar{A} \cap B]\} = P\{A\} + P\{\bar{A} \cap B\}$$
$$P\{B\} = P\{(A \cap B) \cup (\bar{A} \cap B)\} = P\{A \cap B\} + P\{\bar{A} \cap B\}$$

Eliminating $P\{\bar{A} \cup B\}$ between the two equations, we obtain

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\} \tag{2.1.7}$$

Clearly, if $P\{A \cap B\} = 0$ implying A and B are mutually exclusive, then Eq. (2.1.7) reduces to Kolmogorov's axiom 3a. If the probabilities $P\{A\}$ and $P\{B\}$ are equal, then sets A and B are equal in probability. This does not mean that A is equal to B.

The probability of the union of three sets can be calculating from the Venn diagram for the three sets shown in Fig. 1.2.3 and redrawn in Fig. 2.1.2.

The results of the analysis for the cardinalities given in Eq. (1.2.6), can be extended to finding the $P(A \cup B \cup C)$. We

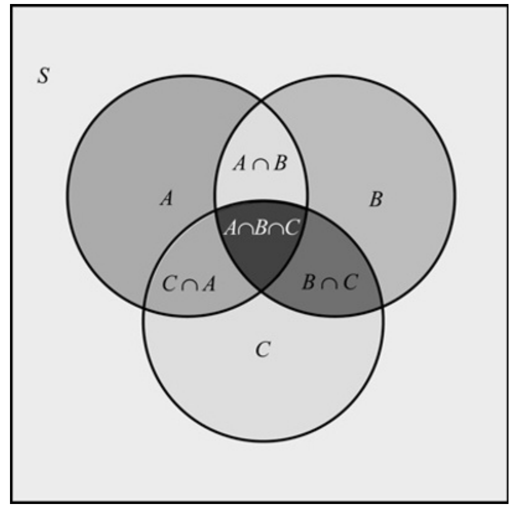


FIGURE 2.1.2

subtract $\{P(A \cap B) + P(B \cap C) + P(C \cap A)\}$ from $\{P(A) + P(B) + P(C)\}$ and add $P(A \cap B \cap C)$, resulting in

$$\begin{aligned} P\{A \cup B \cup C\} = {} & P\{A\} + P\{B\} + P\{C\} \\ & - P\{A \cap B\} - P\{B \cap C\} - P\{C \cap A\} \\ & + P\{A \cap B \cap C\} \end{aligned} \tag{2.1.8}$$

Using the inclusion–exclusion principle of Eq. (1.2.5a), we can also generalize Eq. (2.1.8) to the probabilities of the union of n sets $A_1, A_2, \ldots, A_n$ as follows:

$$\begin{aligned} P\{{}_{i=1}^{n} A_i\} = {} & \sum_{i=1}^{n} P\{A_i\} - \sum_{\substack{i,j=1 \\ i \neq j}}^{n} P\{A_i \cap A_j\} \\ & + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} P\{A_i \cap A_j \cap A_k\} \\ & -(\pm 1)^n \sum_{\substack{i,j,k=1 \\ i \neq j \neq k \ldots \neq n}}^{n} P\{A_i \cap A_j \cap A_k \cdots \cap A_n\} \end{aligned} \tag{2.1.9}$$

Example 2.1.6 A fair coin is tossed 3 times. Since each toss is a functionally independent trial and forms a Cartesian product, the cardinality of the three tosses is $2 \times 2 \times 2 = 8$ and the sample space consists of the following 8 events:

$$S = \{[HHH], [HHT], [HTT], [HTH], [THH], [TTH], [THT], [TTT]\}$$

In this sample space

$$P\{2 \text{ heads}\} = P\{[HHH], [HHT], [HTH], [THH]\} = \frac{4}{8}$$

$$P\{\text{both heads and tails}\} = \{[HHT], [HTT], [HTH], [THH], [TTH], [THT]\} = \frac{6}{8}$$

Example 2.1.7 A telephone call occurs at random between 9:00 a.m. and 11:00 a.m. Hence the probability that the telephone call occurs in the interval $0 < t \le t_0$ is given by $t_0/120$. Thus the probability that the call occurs between 9:30 a.m. and 10:45 a.m. is given by:

$$P\{30 < t \le 105\} = \frac{105-30}{120} = \frac{75}{120} = \frac{5}{8}$$

Example 2.1.8 Mortality tables give the probability density that a person who is born today assuming that the lifespan is 100 years as

$$f(t) = 3 \times 10^{-9} t^2 \left(100 - t^2\right) u(t)$$

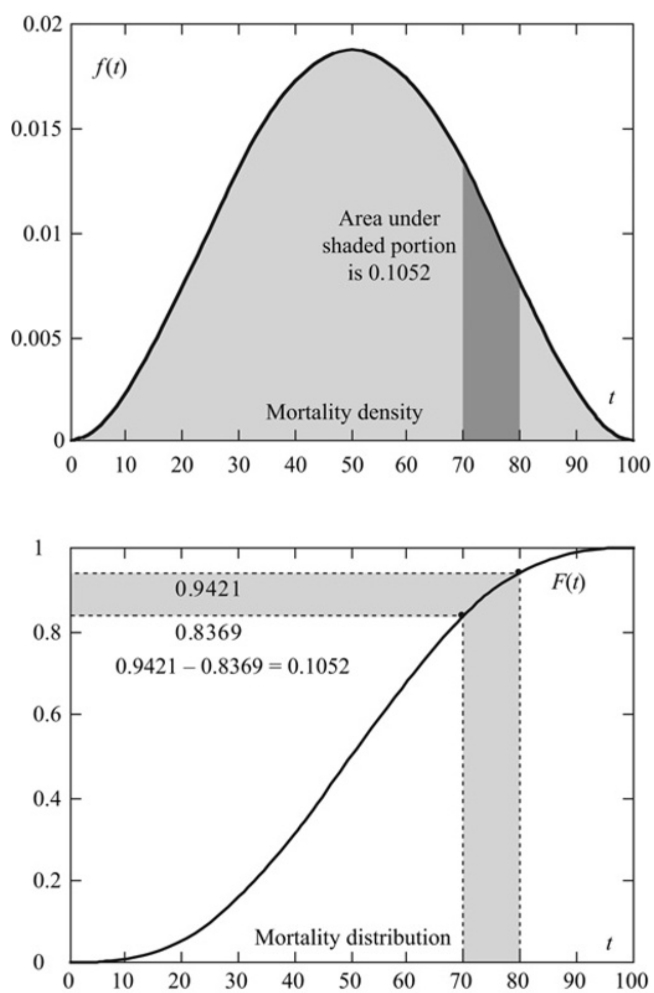


FIGURE 2.1.3

and the distribution function that the person will die at the age t_0 years as

$$F(t_0) = P\{t \le t_0\} = \int_0^{t_0} 3 \times 10^{-9} t^2 \left(100 - t^2\right) dt$$

These probabilities are shown in Fig. 2.1.3. The probability that this person will die between the ages 70 to 80 is given by $P\{70 < t \le 80\} = 0.1052$, obtained from the following integral:

$$P\{70 < t \le 80\} = F(80) - F(70) = \int_{70}^{80} 3 \times 10^{-9} t^2 (100 - t^2)\, dt = 0.1052$$

The density and distribution functions are shown in Fig. 2.1.3.

2.2 CONDITIONAL PROBABILITY

The conditional probability of an event B conditioned on the occurrence of another event B is defined by

$$\left.\begin{aligned} P\{B \mid A\} &= \frac{P\{B \cap A\}}{P\{A\}} \\ P\{B \cap A\} &= P\{B \mid A\} P\{A\} \end{aligned}\right\} \text{ if } P\{A\} \neq 0 \qquad (2.2.1)$$

If $P\{A\} = 0$, then the conditional probability $P\{B \mid A\}$ is not defined. Since $P\{B \mid A\}$ is a probability measure, it also satisfies Kolmogorov axioms.

Conditional probability $P\{B \mid A\}$ can also be interpreted as the probability of the event B in the reduced sample space $S_1 = A$. Sometimes it is simpler to use the reduced sample space to solve conditional probability problems. We shall use both these interpretations in obtaining conditional probabilities as shown in the following examples.

Example 2.2.1 In the tossing of a fair die, we shall calculate the probability that 3 has occurred conditioned on the toss being odd. We shall define the following events:

$$A = \{\text{odd number}\} : \; B = 3$$

$P\{A\} = \frac{1}{2}$ and $P\{B\} = \frac{1}{6}$ and the event $A \cap B = \{3\}$ and hence

$$P\{3 \mid \text{odd}\} = \frac{P\{3\}}{P\{\text{odd}\}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Note that $P\{2 \mid \text{odd}\} = 0$ since 2 is an impossible event given that an odd number has occurred. Also note that $P\{3 \mid 3\} = 1$.

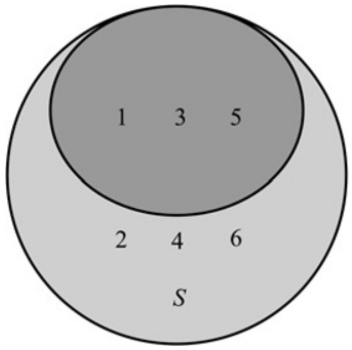


FIGURE 2.2.1

The conditional probability of 3 occurring, given that 3 has occurred is obviously equal to 1.

We can also find the solution using the concept of reduced sample space. Since we are given that an odd number has occurred, we can reduce the sample space from $S = \{1,2,3,4,5,6\}$ to $S_1 = \{1,3,5\}$ as shown in Fig. 2.2.1.

The probability of 3 in this reduced sample space is $\frac{1}{3}$, agreeing with the previous result.

Example 2.2.2 Consider a family of exactly two children. We will find the probabilities (1) that both are girls, (2) both are girls given that one of them is a girl, (3) both are girls given that the elder child is a girl, and (4) both are girls given that both are girls.

The sample space S has 4 points $\{gg, gb, bg, bb\}$, where g is girl and b is boy. The event $B = \{\text{both children are girls}\}$. The conditioning event A will depend on the following four cases:

1. The event $A = S$. Hence $P\{B\} = \frac{1}{4}$.
2. The event $A = \{\text{one child is a girl}\}$. The reduced sample space is $S_1 = \{gg, gb, bg\}$ and the conditional probability $P\{B \mid A\} = \frac{1}{3}$.
3. The event $A = \{\text{elder child is a girl}\}$. In this case, the reduced sample space is $S_2 = \{gg,\ gb\}$ and $P\{B \mid A\} = \frac{1}{2}$.
4. The event $= \{\text{both are girls}\}$. The reduced sample space has only one point, namely, $S_3 = \{gg\}$ and $P\{B \mid A\} = 1$.

As our knowledge of the experiment increases by the conditioning event, the uncertainty decreases and the probability increases from 0.25, through 0333, 0.5, and finally to 1, meaning that there is no uncertainty.

Example 2.2.3 The game of craps as played in Las Vegas has the following rules. A player rolls two dice. He wins on the first roll if he throws a 7 or a 11. He loses if the first throw is a 2, 3, or 12. If the first throw is a 4, 5, 6, 8, 9, or 10, it is called a *point* and the game continues. He goes on rolling until he throws the point for a win or a 7 for a loss. We have to find the probability of the player winning.

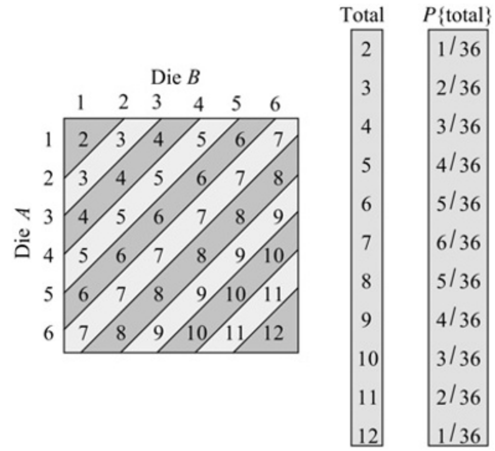


FIGURE 2.2.2

We will solve this problem both from the definition of conditional probability and the reduced sample space. Figure 2.2.2 shows the number of ways the sums of the pips on the two dice can equal 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and their probabilities.

Solution Using Definition of Conditional Probability: The probability of winning in the first throw is

$$P\{7\} + P\{11\} = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = 0.22222$$

The probability of losing in the first throw is

$$P\{2\} + P\{3\} + P\{12\} = \frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{4}{36} = 0.11111$$

To calculate the probability of winning in the second throw, we shall assume that i is the point with probability p. The probability of not winning in any given throw after the first is given by $r = P\{\text{not } i \text{ and not } 7\} = 1 - p - \frac{1}{6}$. We compute the conditional probability

$$P\{\text{win} \mid i \text{ in first throw}\} = p + rp + r^2p + \cdots$$
$$= \frac{p}{1-r} = \frac{p}{p+\frac{1}{6}}$$

as an infinite geometric series:

$$P\{\text{win in second or subsequent throws}\} = \frac{p^2}{p+\frac{1}{6}}$$

The probability p of making the point $i = 4,5,6,8,9,10$ is obtained from Fig. 2.2.2. Thus, for $i = 4,5,6$, we obtain

$$P\{\text{win after point } i=4\} = \left(\frac{3}{36}\right)^2 \bigg/ \left(\frac{3}{36}+\frac{1}{6}\right) = \frac{1}{36}$$

$$P\{\text{win after point } i=5\} = \left(\frac{4}{36}\right)^2 \bigg/ \left(\frac{4}{36}+\frac{1}{6}\right) = \frac{2}{45}$$

$$P\{\text{win after point } i=6\} = \left(\frac{5}{36}\right)^2 \bigg/ \left(\frac{5}{36}+\frac{1}{6}\right) = \frac{25}{396}$$

with similar probabilities for 8,9,10. Thus the probability of winning in craps is

$$P\{\text{win}\} = P\{\text{win in roll 1}\} + P\{\text{win in roll } 2,3,\ldots\}$$
$$= \frac{8}{36} + 2\left[\frac{1}{36}+\frac{2}{45}+\frac{25}{396}\right] = \frac{244}{495} = 0.4929$$

Solution Using Reduced Sample Space: We can see from the figure that there are 6 ways of rolling 7 for a loss. $(i-1, i=4,5,6)$ ways of rolling the point for a win, and $(13-i, i=8,9,10)$ ways of rolling the point for a win. After the point i has been rolled in the first throw, the reduced sample space consists of $(i-1+6)$ points for $i=4,5,6$ and $(13-i+6)$ points for $i=8,9,10$ in which the game will terminate. Thus

$$P\{\text{win in roll } 2,3,\ldots \mid i \text{ in roll } 1\} = P\{W \mid i\}$$
$$= \begin{cases} \dfrac{i-1}{i-1+6}, & i=4,5,6 \\ \dfrac{13-i}{13-i+6}, & i=8,9,10 \end{cases}$$

Thus, $P\{W \mid 4\} = \frac{3}{9}$, $P\{W \mid 5\} = \frac{4}{10}$, $P\{W \mid 6\} = \frac{5}{11}$, $P\{W \mid 8\} = \frac{5}{11}$, $P\{W \mid 9\} = \frac{4}{10}$, $P\{W \mid 10\} = \frac{3}{9}$. Hence probability of a win after the first roll is $P\{W \mid i\} \cdot P\{i\}$. Hence

$$P\{\text{win after first roll}\}$$
$$= 2\left[\frac{3}{9}\cdot\frac{3}{36}+\frac{4}{10}\cdot\frac{4}{36}+\frac{5}{11}\cdot\frac{5}{36}\right] = \frac{134}{495}$$
$$P\{\text{win}\} = \frac{8}{36} + 2\left[\frac{3}{9}\cdot\frac{3}{36}+\frac{4}{10}\cdot\frac{4}{36}+\frac{5}{11}\cdot\frac{5}{36}\right]$$
$$= \frac{8}{36} + \frac{134}{495} + \frac{244}{495} = 0.4929$$

a result obtained with comparative ease.

Example 2.2.4 We shall continue Example 2.1.8 and find the probability that a person will die between the ages 70 and 80 given that this individual has lived upto the age 70. We will define the events A and B as follows:

$$A = \{\text{person has lived upto} 70\}$$
$$B = \{\text{person dies between ages 70 and 80}\}$$

The required conditional probability equals the number of persons who die between the ages 70 and 80 divided by the number of persons who have lived upto the age 70:

$$P\{B|A\} = \frac{P\{70 < t \le 80 \text{ and } t \ge 70\}}{P\{t \ge 70\}} = \frac{P\{70 < t \le 80\}}{P\{t \ge 70\}}$$
$$= \frac{\displaystyle\int_{70}^{80} 3\times 10^{-9} t^2(100-t^2)\,dt}{\displaystyle\int_{70}^{100} 3\times 10^{-9} t^2(100-t^2)\,dt} = \frac{0.105}{0.163} = 0.644$$

This type of analysis is useful in setting up actuarial tables for life insurance purposes.

2.3 INDEPENDENCE

We have already seen the concept of *functional independence* of sets in Chapter 1. We will now define *statistical independence* of events. If A and B are two events in a field of events $\mathscr{F}$, then A and B are statistically independent if and only if

$$P\{A \cap B\} = P\{A\}P\{B\} \tag{2.3.1}$$

Statistical independence neither implies nor is implied by functional independence. Unless otherwise indicated, we will call statistically independent events "independent" without any qualifiers.

Three events A_1, A_2, A_3 are independent if they satisfy, in addition to

$$P\{A_1 \cap A_2 \cap A_3\} = P\{A_1\}P\{A_2\}P\{A_3\} \tag{2.3.2}$$

the pairwise independence

$$P\{A_i \cap A_j\} = P\{A_i\}P\{A_j\}, \quad i, j = 1,2,3, \quad i \ne j \tag{2.3.3}$$

Note that if A and B are mutually exclusive, then, from Kolmogorov's axiom 3a, we obtain

$$P\{A \cup B\} = P\{A\} + P\{B\} \tag{2.3.4}$$

The concepts of independence and mutual exclusivity are fundamental to probability theory. For any two sets A and B, we recall from Eq. (2.1.7) that

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\} \tag{2.1.7}$$

TABLE 2.3.1

1, *H*	2, *H*	3, *H*	4, *H*	5, *H*	6, *H*
1, *T*	2, *T*	3, *T*	4, *T*	5, *T*	6, *T*

If A and B are mutually exclusive, then $P\{A\cap B\}=0$, but if they are independent, then $P\{A\cap B\}=P\{A\}\cdot P\{B\}$. If this has to be zero for mutual exclusivity, then one of the probabilities must be zero. Therefore, we conclude that mutual exclusivity and independence are incompatible unless we are dealing with null events.

We shall now give examples to clarify the concepts of functional independence and statistical independence.

Example 2.3.1 The probability of an ace on the throw of a die is $\frac{1}{6}$. The probability of a head on the throw of a coin is $\frac{1}{2}$. Since the throw of a die is not dependent on the throw of a coin, we have functional independence and the combined sample space is the Cartesian product with 12 elements as shown in Table 2.3.1.

We assume as usual that all the 12 elements are equiprobable events and the probability of ace and head is $\frac{1}{12}$ since there is only one occurrence of $\{\mathbf{1},\ \mathbf{H}\}$ among the 12 points as seen in the first row and first column in Table 2.3.1. Using Eq. (2.3.1), we have $P\{1,\ H\}=P\{1\}\cdot P\{H\}=\frac{1}{6}\times\frac{1}{2}=\frac{1}{12}$, illustrating that in this example, functional independence coincides with statistical independence.

Example 2.3.2 Two dice, one red and the other blue, are tossed. These tosses are functionally independent, and we have the Cartesian product of $6\times 6=36$ elementary events in the combined sample space, where each event is equiprobable. We seek the probability that an event B defined by the sum of the numbers showing on the dice equals 9. There are four points $\{(6,3),\ (5,4),\ (4,5),\ (3,6)\}$, and hence $P\{B\}=\frac{4}{36}=\frac{1}{9}$. We now condition the event B with an event A defined as the red die shows odd numbers. Clearly, the events A and B are functionally dependent. The probability of the event A is $P\{A\}=\frac{18}{36}=\frac{1}{2}$. We want to determine whether the events A and B are statistically independent. These events are shown in Fig. 2.3.1, where the first number is for the red die and the second number is for the blue die.

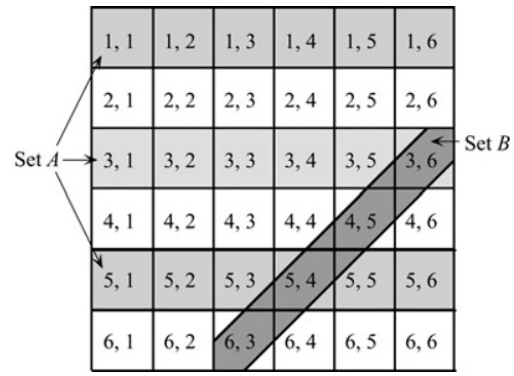


FIGURE 2.3.1

From Fig. 2.3.1 $P\{A\cap B\}=P\{(3,6),\ (5,4)\}=\frac{2}{36}=\frac{1}{18}$ and $=P\{A\}\times P\{B\}=\frac{1}{2}\times\frac{1}{9}=\frac{1}{18}$ showing statistical independence. We compute $P\{B\mid A\}$ from the point of view of reduced sample space. The conditioning event reduces the sample space from 36 points to 18 equiprobable points and the event $\{B\mid A\}=\{(3,6),\ (5,4)\}$. Hence $P\{B\mid A\}=\frac{2}{18}=\frac{1}{9}=P\{B\}$, or, the conditioning event has no influence on B. Here, even though the events A and B are functionally dependent, they are statistically independent.

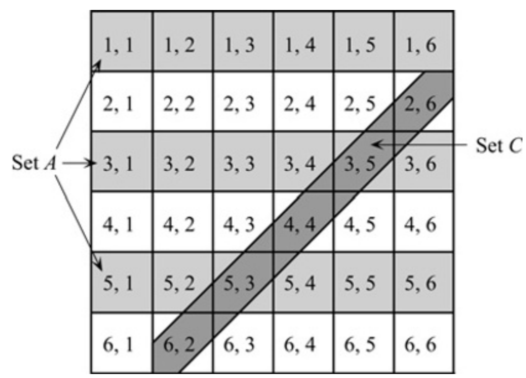


FIGURE 2.3.2

However, if another set C is defined by the sum being equal to 8 as shown in Fig. 2.3.2, then $P\{C\}=\frac{5}{36}$.

Here the events C and A are not statistically independent because $P\{C\}\cdot P\{A\}=\frac{5}{36}\times\frac{1}{2}=\frac{5}{72}\neq P\{C\cap A\}=P\{(3,5),\ (5,3)\}=\frac{2}{18}=\frac{4}{72}$.

In this example, we have the case where the events A and C are neither statistically independent nor functionally independent.

2.4 TOTAL PROBABILITY AND BAYES' THEOREM

Let $\{A_i,\ i=1,\ldots,n\}$ be a partition of the sample space and let B an arbitrary event in the sample space S as shown in Fig. 2.4.1.

We will determine the *total probability* $P\{B\}$, given the conditional probabilities $P\{B\mid A_1\},\ P\{B\mid A_2\},\ldots,\ P\{B\mid A_n\}$. The event B can be given in terms of the partition as

$$\begin{aligned} B=\{B\cap S\}&=B\cap\{A_1\cup A_1\cup\cdots\cup A_n\}\\ &=\{B\cap A_1\}\cup\{B\cap A_2\}\cup\cdots\cup\{B\cap A_n\} \end{aligned} \tag{2.4.1}$$