Introductory Probability Theory

A first Course in Probability Theory – Volume I

Nicholas N.N. Nsowah-Nuamah

bookboon
The eBook company

NICHOLAS N.N. NSOWAH-NUAMAH

INTRODUCTORY PROBABILITY THEORY

A FIRST COURSE IN PROBABILITY THEORY – VOLUME I

Introductory Probability Theory: A first Course in Probability Theory – Volume I
2nd edition

ISBN 978-87-403-1969-9
Peer review by Prof. Bashiru I.I. Saeed

CONTENTS

DEDICATION

To all students, both past and present, in the Department of Statistics and Institute of Statistical, Social and Economic Research (I.S.S.E.R) of the University of Ghana, without whose help and encouragement this book would never have been a reality.

PREFACE TO THE SECOND EDITION

The second edition of Volume I of the book has gone through a substantial revision. Most parts have been re-written and/or rearranged to bring clarity. In some cases, more relevant examples are used. Consequently, the old Volume I now comes in two volumes.

Volume I deals with the basic mathematical tools for the understanding of probability, basic probability concepts, probability calculus, laws and theorems in probability, random variables and probability distributions. Determination of central and non-central location of distributions as well as their spread are extensively discussed. The volume concludes with discussion of moments and moment-generating functions.

Volume II discusses special probability distributions, namely, Bernoulli Distribution, Binomial Distribution, Geometric Distribution, Negative Binomial Distribution, Poisson Distribution, Hypergeometric Distribution, Multinomial Distribution, Uniform Distribution, Exponential Distribution, Gamma Distribution, Beta Distribution and Normal Distribution.

The extension of univariate random variable to multivariate random variables with emphasis on Bivariate Distributions, and also Statistical Inequalities, Limit Laws and Sampling Distributions are in another e-book entitled "Advanced Topics in Basic Probability Theory".

While the typesetting was done by myself, the images were produced by Mr Joshua Appiah, a former staff of mine at Kumasi Technical University, where I was their Interim Vice Chancellor.

N.N.N. Nsowah-Nuamah
Accra, Ghana
August, 2017

PREFACE TO THE FIRST EDITION

This book comes in two volumes. Volume I deals with probability calculus and probability distributions and Volume II treats special topics in probability theory.

Much of the recent literature in probability appears to be inaccessible to many students, in large part, I believe, because rigour has taken precedence over the communication of fundamental ideas. What should be clear to authors of probability texts is that, most University students in Statistics go through a course in Probability with difficulty and disinterest and they do so because it is an unavoidable requirement at the Department(s) of Statistics and/or Mathematics. Some students had bitter experiences in schools with the teaching of probability and are already scared before starting the undergraduate course. The problem is compounded when the student realises that many available textbooks do not treat the topics in a friendly way.

This text has been written in a systematic way such that the student needs only to be introduced to the topic for him/her to understand the specified chapter. In view of this the rigour of other books has been avoided. I have also tried very hard to use notation that is at once consistent and intuitive. The only mathematical prerequisite in this Volume is a knowledge of elementary course in Calculus. An exposure to differentiation, integration, series and some ideas of convergence should be sufficient. Within the chapters, important concepts are set out as *definitions* (enclosed in double boxes) while the main theoretical results of concepts and meanings are set out as *theorems* (enclosed in single boxes). Results which follow almost immediately from theorems with very little additional argument are stated as *corollaries*. Even though the book does not provide a formal rigorous treatment, at times we adopt a "theorem-proof" format. Proofs have been provided to almost all theorems and corollaries stated. We implore the reader to clearly understand the definitions and implications of the theorems and corollaries although it is not always necessary to understand the proof of each theorem. Many instructors and students will prefer to skim through these proofs, especially in a first course, although we hope that most will choose not to omit them completely. The *notes*, a common feature of the book, are an important part in three respects. Firstly, they illuminate points made in the text. Secondly, they present material that does not quite fit anywhere else and thirdly, they discuss details whose exposition would disturb the flow of the argument.

In this Volume more than 200 worked examples have been provided and it is essential that the student goes through all of them. Probability theory requires a lot of practice from the student. I have therefore provided a set of questions at the end of each chapter. These exercises, which are more than 300, form an integral part of the text and it is essential that the student goes through many of them.

The Undergraduate programmes in Statistics may vary among countries and even among departments in the same country. Besides, the ability of students to grasp probability theory may be diverse from year to year. In view of this, there is purposely more material here than may be required. This will give the instructor the opportunity to choose topics to satisfy his/her own needs and taste.

Now a brief tour through this Volume. In Chapter 1 a brief account of Set Theory is given. Chapter 2 describes a number of counting tricks that are very useful in solving probability problems. These first two chapters do not really discuss probability at all. Instead, they present the essential mathematical tools that are needed in the rest of the book. Although you may have dealt with these concepts in high school, it is important that these two chapters be utilised (at least by the independent reader) to assure a sound base for the applications in probability. Together with this, a detailed treatment of basic concepts in probability theory, such as experiments, sample spaces, and events in Chapter 3 make the book self-contained both as a reference and a text. The materials in Chapters 4–12 form the heart of any introductory course in probability theory. Chapters 4, 5 and 6 give the basic theory of probability while Chapters 7 to 12 treat special probability distributions.

We shall go through the basic content of Volume II. Chapters 1–4 extend the concept of a univariate random variable to multivariate random variables, even though emphasis is on Bivariate Distributions. Chapter 1 discusses the Joint, Marginal and Conditional Probability Distributions, Joint Cumulative Distribution Functions and Independence of Bivariate Random Variables. Chapter 2 is on Sums, Differences, Products and Quotients of Bivariate Random Variables. Chapter 3 treats Expectation and Variance of Random Variables and their properties. Chapter 4 discusses various forms of Moments of bivariate distributions, and goes on to discuss Covariance, Correlation, Conditional Expectations and Regression Curves. Chapter 5 is on Statistical Inequalities (Markov's and Chebyshev's Inequalities in their various forms) and Limit Laws (the Law of Large Numbers and the Central Limit Theorem, also in their various forms). Sampling Distributions are discussed in Chapters 6 (Basic Concepts) and 7 (Sampling Distributions of a Statistic). We conclude Volume II with the Distributions derived from the Normal Distribution, namely, the χ^2, the t and the F Distributions.

We believe that the two Volumes of the book may also be used as a revision text for students starting their Master's programme.

N.N.N. Nsowah-Nuamah
Accra, Ghana
November, 1997

ACKNOWLEDGEMENT

First and foremost, I express my sincere thanks to the Almighty God both for the gift of writing He has endowed me with and for the stamina I had to complete this particular book.

It is possible that a book as detailed and technical as this one may not be free from errors. If there are few errors it was because l had the assistance of a great many people in reading the preliminary drafts.

I am indebted to my students in both the Institute of Statistical, Social and Economic Research (I.S.S.E.R) and the Department of Statistics all of the University of Ghana, in this respect. They pointed out a disconcertingly large number of mistakes, mostly typographical in nature. I am particularly grateful to Simon Harvey, David Ashong, Emmanuel Nunoofio, Albert Nyarko, Ezekiel Nortey and Matthew Kofie who critically cross-checked the answers to the questions which appear after each chapter. I thank Nathaniel Kwawukume for the many hours he devoted to going through most of the chapters and the extremely valuable suggestions he made. I am also indebted to Dr. F.K. Atsem of Department of Statistics, University of Ghana, who critically read through the final draft and made very useful suggestions.

I owe special thanks to two anonymous reviewers who spent substantial time reviewing the manuscript and who provided numerous helpful comments. On the quality of this text, I owe so much to the Swedish Agency for Research Cooperation with Developing Countries (SAREC) for financial support during my visit at the International Centre for Theoretical Physics. I am also very grateful to UNESCO, the International Atomic Energy Agency and the International Center for Theoretical Physics in two respects. Firstly, they provided me conducive conditions including an office space, and a good library and computer facilities which accorded me the opportunity to reassess the draft of the manuscript and prepare the camera-ready copy for the publishers. Secondly, they provided an opportunity for me to meet other scientists who commented on the manuscript. Prominent among them was Professor A.A.K. Majumdar of Mathematics Department, Jahangirnagar University in Bangladesh, who critically read the manuscript at its final stage and made constructive suggestions in both substance and style. Most of his and Dr. Atsem's comments have given this book its current shape. I am grateful to the University of Ghana for permission to use examination questions from the Department of Statistics and I.S.S.E.R. I also acknowledge that some of the questions in the exercises have been taken from some of the books listed in the bibliography.

The book was typeset using Latex whose authors I would like to thank for a superior product. I did the typesetting myself but I do not know how I could have completed it on time if David Mensah had not aided me. I say thank you to Prof. G.O.S Ekhaguere of Ibadan University who showed me a few more tricks in LaTeX when he visited Ghana. I also acknowledge Ato Kwamena Otu (who I call Leftie) for sketching some of the diagrams in the text (especially the Normal curves) which I could not immediately use the computer to draw.

And finally, I appreciate very much the care, cooperation and attention devoted to the preparation of the book by the staff of Ghana Universities Press, especially the Director, Mr. K.M. Ganu, and Mr. V.K. Boadu, who personally went through the scripts to check for consistencies and typographical errors.

In spite of careful checking by me and others some of whom I have acknowledged, it is possible that the occasional errors might have crept through. All such errors are of course my responsibility, including any flaws and inadequacies. I would appreciate receiving information about such mistakes or comments which might be discovered to be sent directly to me, (my e-mail address is, sadaocgh@ghana.com) or via Ghana Universities Press.

GREEK ALPHABETS

A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	xi
Γ	γ	gamma	O	o	omicron
Δ	δ	delta	Π	π	pi
E	ϵ	epsilon	P	ρ	rho
Z	ζ	zeta	Σ	σ	sigma
H	η	eta	T	τ	tau
Θ	θ	theta	Υ	υ	upsilon
I	ι	iota	Φ	ϕ, φ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ψ	psi
M	μ	mu	Ω	ω	omega

LIST OF MATHEMATICAL NOTATIONS AND CONVENTIONS

$=$	equal to
$\neq$	not equal to
$<$	is less than
$\leq$	less than or equal to
$>$	greater than
$\geq$	greater than or equal to
$\approx$	approximately equal to
$\sim$	asymptotically equal to (or distributed as)
$\pm$	plus or minus
$\emptyset$	the empty (null) set
$\in$	is a member of (belongs to)
$\notin$	is not a member of (does not belongs to)
$\subseteq$	is a subset of (is contained in)
$\nsubseteq$	is not a subset of (is not contain in)
$\supseteq$	contains
$\nsupseteq$	does not contain
$\subset$	is a proper subset of
$\not\subset$	is not a proper subset of
$\Rightarrow$	implies
$\nRightarrow$	does not imply
$\Leftrightarrow$	equivalent to
$\cup$	the union of
$\cap$	the intersection of
A^c, A', $\overline{A}$	A complement
$A \backslash B$, $A - B$	difference of A and B
$A \Delta B$	symmetric difference of A and B
$\mathcal{P}(\mathcal{A})$	power set of A
$n!$	$n -$ factorial
$\binom{n}{r}$, ${}_nC_r$,	binomial coefficient (n combination r)
${}_nP_r$,	n permutation r
Σ	sum of
Π	product of
$n(\cdot)$, $\#(\cdot)$	number of elements in
∞	infinity

$P(A)$	probability of A
$P(A\|B)$	conditional probability of A given B
$p(x_i)$	probability mass function
$f(x)$	probability density function
$F(x)$	cumulative distribution function
$E(X)$	expectation of X
$\mathrm{Var}(X)$, σ^2	variance of X
σ	standard deviation
$\mathrm{Cov}(X, Y)$	covariance of X and Y
$\mathrm{Corr}(X, Y)$	correlation of X and Y
$M_X(t)$	moment-generating function of X
a_3	skewness
a_4	kurtosis
$b(x; n, p)$	Binomial probability distribution
$B(x; n, p)$	Binomial distribution function
$G(x; p)$	Geometric probability distribution
$b^-(x; n, p)$	Negative binomial probability distribution
$p(x; \lambda)$	Poisson probability distribution
$P(x; \lambda)$	Poisson distribution function
$h(x; N, n, M)$	Hypergeometric probability distribution
$H(x; N, n, M)$	Hypergeometric distribution function
$M(x_i; n, p_i)$	Multinomial probability distribution
$U(a, b)$	Uniform probability distribution
$\mathrm{Exp}(x; \lambda)$	Exponential probability distribution
$\Gamma(\alpha)$	Gamma function
$\Gamma_x(\alpha)$	Incomplete gamma function
$\mathrm{Gamma}(x; \alpha, \beta)$	General Gamma probability distribution
$B(\alpha, \beta)$	Beta function
$B_x(\alpha, \beta)$	Incomplete beta function
$\mathrm{Beta}(x; \alpha; \beta)$	Beta probability distribution
$N(\mu, \sigma)$	General Normal probability distribution
$N(0, 1)$	Standard Normal probability distribution
$\xrightarrow{p}$	converges in probability to
$\xrightarrow{a.s}$	converges almost surely to
$\xrightarrow{d}$	converges in distribution to
$w.p.1$	converges with probability 1

1 SET THEORY

1.1 INTRODUCTION

In everyday language we often hear expressions such as a class of children, a fleet of cars, a crate of soft drinks, etc. All these express the basic idea of a collection of objects with some common properties. The idea of a collection of objects is the basic idea behind the notion of sets. Set theory, in itself, has few direct applications in probability theory. It plays the same role in present day probability as numbers in previous times.

a) It provides a sufficiently sensitive, precise, and flexible language for the accurate expression of probability ideas.
b) It helps to clarify ideas that traditionally seem to be confused in probability.

In this chapter, we shall introduce some basic concepts in set theory that will be sufficient for the understanding of the probability theory presented in this book. Standard symbols for the more important sets of numbers are desirable and we introduce them first.

a) $\mathcal{N}$ = the set of ***natural numbers***: $1, 2, 3, \cdots$
The natural numbers are also called the ***counting numbers***. When the number 0 is added to the set of natural numbers we have the set of ***whole numbers***. Thus, the whole numbers are $0, 1, 2, 3, \cdots$

b) $\mathcal{Z}$ = the set of integers: $\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots$
Clearly, the set of natural numbers $\mathcal{N}$ is contained in $\mathcal{Z}$.

c) $\mathcal{Q}$ = the set of ***rational*** or ***fractional numbers*** $\frac{p}{q}$ where p and q are any integers and $q \neq 0$. Every rational number when expressed as a decimal will be either a terminating[1] or a repeating decimal[2]. An ***irrational number*** is a number that cannot be written as a simple fraction because when written as a decimal, the decimal part is non-terminating, non-repeating decimal. Examples of an irrational number are $\sqrt{2}$ and π.
d) $\mathcal{R}$ = the set of real numbers (collection of rational and irrational numbers).

1.2 SETS

1.2.1 DEFINITION OF SET

Definition 1.1 SET

A set is a collection of well-defined and distinct objects

"Objects" may be things, people or symbols. "Well-defined" means that there must be no doubt whatsoever about whether or not a given item belongs to the set under consideration. "Distinct" is used in the sense that no two identical objects should be contained in the same set.

Example 1.1
Examples of sets include the following:

a) The set of all students in a class.
b) The set of all houses in Ghana.
c) The set of all numbers greater than 1.
d) The set of all months with less than 30 days.

1.2.2 ELEMENTS

Definition 1.2 ELEMENTS

The objects that belong to a set are called its elements

A synonym for elements is members.

Example 1.2
If a list of students is compiled, each student is an element of the set of students.

In general, unless otherwise specified, we denote a set by a capital letter such as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and an element by a lower case letter such as a, b, c. We use the symbol $\in$ to denote membership of a set. Thus the notation

$$\begin{array}{rll} a \in \mathcal{B} & \text{means} & \text{“}a \text{ is an element of the set } \mathcal{B}\text{”;} \\ a \notin \mathcal{B} & \text{means} & \text{“}a \text{ is } \textit{not} \text{ an element of the set } \mathcal{B}\text{”;} \\ a, b \in \mathcal{B} & \text{means} & \text{“both } a \text{ and } b \text{ are elements of the set } \mathcal{B}\text{”;} \\ a, b \notin \mathcal{B} & \text{means} & \text{“both } a \text{ and } b \text{ are } \textit{not} \text{ elements of the set } \mathcal{B}\text{”.} \end{array}$$

Alternatively, $a \in \mathcal{B}$ may be read as "a belongs to the set $\mathcal{B}$."

1.2.3 DESCRIPTION OF SETS

There are three methods commonly used to describe a set:

a) by definition;
b) by the roster method and
c) by the property method.

The ***definition method*** of describing a set has been illustrated in Example 1.1.

The ***roster method*** (or sometimes referred to as the tabular form) specifies a set by actually listing its elements. The simplest notation for a set is to list its members inside the curly brackets {}. The braces are an essential part of the notation, since they identify the contents as a set.

Example 1.3
$\mathcal{A}$ is the set of integers between zero and ten. The set $\mathcal{A}$ may be written as

$$\mathcal{A} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

or simply as

$$\mathcal{A} = \{1, 2, \cdots, 9\}$$

Note

a) The order in which the elements in a set are written does not matter. For example, the set $\mathcal{A} = \{1, 2, 3\}$ is the same as the set $\mathcal{B} = \{2, 1, 3\}$.
b) Each element of a set is listed only once.

Although the roster method is a common practice, the danger exists that the pattern the reader sees may not be the one the writer had in mind. Also, sometimes, it is not possible to use the roster method; for example, the set of all points in a square.

The ***property method*** (sometimes called the *set-builder form* or the *set-generator notation*) specifies a set by stating properties which characterize the elements of the set. A property of membership may be given in words. The following example illustrates the use of the property method.

Example 1.4
The set $\mathcal{A}$ consists of integers between zero and ten. This may be written as: "the set of all elements x such that x is an integer between zero and ten".

In a more precise and compact form, this set may be described as follows:

$$\mathcal{A} = \{x : x \quad \text{is an integer}, \quad 0 < x < 10\}$$

or

$$\mathcal{A} = \{x : x \in Z; \quad 0 < x < 10\}$$

where the colon ":" is read "such that" or "given that".

Note

Sometimes, a vertical line "|" may also be used instead of the colon ":". Thus, the set $\mathcal{A}$ in Example 1.4 could be written as

$$\mathcal{A} = \{x \mid x \quad \text{is an integer}, \quad 0 < x < 10\}$$

1.2.4 FINITE AND INFINITE SETS

Definition 1.3 FINITE AND INFINITE SETS

$\mathcal{A}$ set is finite if it is empty or if it consists of exactly n distinct elements where n is a positive integer; otherwise it is infinite

Example 1.5

a) If $\mathcal{A}$ = the set of the months in the year, then:

$$\mathcal{A} = \{\text{January, February, March}, \cdots, \text{December}\}$$

and, therefore, $\mathcal{A}$ is a finite set.

b) $\mathcal{A}$ = the set of all people living on the earth.
Although it is difficult to count the number of people living on the earth, we can still know the number given sufficient time and resources. Therefore, $\mathcal{A}$ is finite.

c) $\mathcal{A}$ = the set of non-negative integers:
$\mathcal{A} = \{0, 1, 2, \cdots\}$
$\mathcal{A}$ is an infinite set because there is no limit to the elements in the set.

d) $\mathcal{A}$ = {real numbers in the interval $0 \leq x \leq 1$}, that is,
$\mathcal{A} = \{x : 0 \leq x \leq 1\}$
$\mathcal{A}$ is an infinite set because there is an infinite number of points in such an interval, as for example, $\frac{1}{10}, \frac{1}{100}|, \frac{1}{1000}$, and so on.

1.2.5 COUNTABLE AND UNCOUNTABLE SETS

Definition 1.4 COUNTABLE AND UNCOUNTABLE SETS

A set is countable if it is finite or if its elements can be arranged in the form of a sequence, in which case it is said to be *countably infinite* or *denumerable*; otherwise, the set is *uncountable* or *non-denumerable*

Example 1.6
For the information in Example 1.5, determine the sets that are countable, countably infinite and uncountable.

Solution
Example 1.5(a) is a countable set; Example 1.5(b) is a countable set; Example 1.5(c) is a countably infinite set; Example 1.5(d) is an uncountable set.

Cardinality of a set

Definition 1.5 CARDINALITY OF SETS

The cardinality of any set $\mathcal{A}$ is the number of elements in the set $\mathcal{A}$

A synonym for cardinality is the *cardinal number*. The cardinality of, say, set $\mathcal{A}$, is denoted by the symbol $n(\mathcal{A})$ or $|\mathcal{A}|$ or sometimes $\#\mathcal{A}$, and read as "the number of elements in set $\mathcal{A}$".

Note

There is the cardinality of certain infinite sets but we cannot simply use the symbol ∞ to denote the cardinality of any infinite set. The subject of cardinality of infinite sets is important and interesting in its own right but we will not pursue it further here.

Example 1.7
The set $\mathcal{A} = \{1, 2, 7, 9\}$ has cardinal number 4, because the set contains four elements, hence, $n(\mathcal{A}) = 4$.

1.2.6 CONCEPTS IN SETS

Singleton Sets

Definition 1.6 SINGLETON SET

A set which has only one element in it is called a singleton set

A synonym for a singleton set is a *unit set*.

Example 1.8
Consider the set {all months with less than 30 days}. This is {February}. This is a singleton set. Other examples include {2}, {D}, {$\star$} {b}.

Empty Sets

Definition 1.7 EMPTY SET

The set which contains no element at all is called the empty set

Synonym for empty set is *null set* and is denoted by {} or $\emptyset$. Formally, we may define the empty set as

$$\emptyset = \{x \mid x \neq x\}$$

Example 1.9

a) The set of coins with three faces is an empty set since no coin has three faces.
b) The set of months with thirty two days is an empty set since there is no such month.

Non-Empty Sets

Definition 1.8 NON-EMPTY SET

A non-empty set is a set which has at least one element

Example 1.10

The sets {0}, {a}, {15, 6}, $\{\star\}$ are all non-empty sets since each contains at least one element.

Equal Sets

Definition 1.9 EQUAL SETS

Two sets which have exactly the *same elements* are called equal sets

Equality of sets is denoted by "=".

Example 1.11

If

$$\mathcal{A} = \{1, 2, 3, 4\} \quad \text{and} \quad \mathcal{B} = \{2, 4, 3, 1\}$$

then $\mathcal{A} = \mathcal{B}$ because they contain the same elements, namely, $1, 2, 3, 4$.

Unequal Sets

If $\mathcal{A}$ and $\mathcal{B}$ do not have exactly the same elements we say that $\mathcal{A}$ is not equal to $\mathcal{B}$, written as

$$\mathcal{A} \neq \mathcal{B}$$

Example 1.12
If

$$\mathcal{A} = \{1, 2, 3, 4\} \quad \text{and} \quad \mathcal{B} = \{1, 2, 3, 5\}$$

then $\mathcal{A} \neq \mathcal{B}$ because they do not contain exactly the same elements. $4 \in \mathcal{A}$ but $4 \notin \mathcal{B}$; also $5 \in \mathcal{B}$ but $5 \notin A$.

Note

The set {0} is not an empty set because it has zero as its element.

Equivalent Sets

Definition 1.10 EQUIVALENT SETS

Two sets $\mathcal{A}$ and $\mathcal{B}$ are said to be equivalent if they contain the *same number* of elements

Equivalent sets are denoted by "$\Leftrightarrow$", that is, if set $\mathcal{A}$ is equivalent to set $\mathcal{B}$, we write $\mathcal{A} \Leftrightarrow \mathcal{B}$.

Example 1.13
If

$$\mathcal{A} = \{1, 2, 3, 4, 5\} \quad \text{and} \quad \mathcal{B} = \{a, b, c, d, e\}$$

then $\mathcal{A}$ is equivalent to $\mathcal{B}$ because they contain the *same number* of elements.

It is important to remember that ***all equal sets are equivalent but not all equivalent sets are equal***.

Thus, in Example 1.11, the sets $\mathcal{A}$ and $\mathcal{B}$ are equal and, therefore, they are also equivalent sets. However, in Example 1.13, because they contain the same number of elements the two sets are equivalent but they are not equal, because they do not contain the same elements.

Note

The following statements all say the same thing:

a) Sets $\mathcal{A}$ and $\mathcal{B}$ are equivalent
b) Sets $\mathcal{A}$ and $\mathcal{B}$ have the same cardinal number.

Universal Set

In discussing sets in a particular context, all such sets may be viewed relative to some particular set called the universal set.

> **Definition 1.11** UNIVERSAL SETS
>
> The universal set is the set of all elements relevant to a particular problem or discussion

Synonyms for a universal set are the *universe of discourse* (or simply *universe*) or the *population set* and it is denoted as $\mathcal{U}$ or as $\mathcal{E}$ in some texts. In this text, we shall adopt the former notation.

Example 1.14
Examples of universal sets include:

a) The list of all students in a school;
b) The set of real numbers.

1.2.7 SUBSETS

Definition 1.12 SUBSET

If every element of a set $\mathcal{A}$ belongs to a set $\mathcal{B}$, then $\mathcal{A}$ is a subset of $\mathcal{B}$

"$\mathcal{A}$ is a subset of $\mathcal{B}$" is written as:

$$\mathcal{A} \subseteq \mathcal{B}$$

and may also be read "$\mathcal{A}$ is contained in $\mathcal{B}$".
We may also write $\mathcal{A} \subseteq \mathcal{B}$ as

$$\mathcal{B} \supseteq \mathcal{A}$$

and read as "$\mathcal{B}$ contains $\mathcal{A}$", or "$\mathcal{B}$ is a superset of $\mathcal{A}$".

Example 1.15
If

$$\mathcal{A} = \{1, 2, 3, 4, 5\} \quad \text{and} \quad \mathcal{B} = \{1, 2, 3, 4, 5, 6, 7\}$$

then $\mathcal{A} \subseteq \mathcal{B}$ since all elements in $\mathcal{A}$ belong to $\mathcal{B}$ as well.

If at least one element of $\mathcal{A}$ does not belong to $\mathcal{B}$, then $\mathcal{A}$ is *not a subset* of $\mathcal{B}$, and in such a case, we write

$$\mathcal{A} \nsubseteq \mathcal{B} \quad \text{or} \quad \mathcal{B} \nsupseteq \mathcal{A}$$

Example 1.16
If

$$\mathcal{A} = \{1, 2, 3, 4, 5\} \quad \text{and} \quad \mathcal{B} = \{1, 2, 3, 4, 6, 7\}$$

then $\mathcal{A} \nsubseteq \mathcal{B}$ since the element 5 in $\mathcal{A}$ does not belong to $\mathcal{B}$.

Theorem 1.1

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be any sets. Then

a) $\emptyset \subseteq \mathcal{A}$
b) $\mathcal{A} \subseteq \mathcal{A}$
c) If $\mathcal{A} \subseteq \emptyset$ then $\mathcal{A} = \emptyset$
d) If $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{A}$ then $\mathcal{A} = \mathcal{B}$
e) If $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{C}$ then $\mathcal{A} \subseteq \mathcal{C}$

In *Theorem 1.1*, part (a) implies that the null set is a subset of every set; part (b) means that every set is a subset of itself and part (d) is another way of stating equality of two sets $\mathcal{A}$ and $\mathcal{B}$.

Example 1.17
Let $\mathcal{A} = \{1, 2, 3, 4\}$ and $\mathcal{B} = \{1, 2, 3, 4\}$.
Then $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{A}$. Hence $\mathcal{A} = \mathcal{B}$.

Example 1.18
Let $\mathcal{A} = \{1, 2, 3, 4\}$ and $\mathcal{B} = \{1, 2, 3, 4, 5\}$.
Then $\mathcal{A} \subseteq \mathcal{B}$ but $\mathcal{A} \neq \mathcal{B}$ since $\mathcal{B} \nsubseteq \mathcal{A}$.

Proper Subsets

Definition 1.13 PROPER SUBSET

If every element in $\mathcal{A}$ is an element in $\mathcal{B}$, and also $\mathcal{B}$ has at least one other element which is not in $\mathcal{A}$, then $\mathcal{A}$ is called a proper subset of $\mathcal{B}$.

If $\mathcal{A}$ is a proper subset of $\mathcal{B}$, we write

$$\mathcal{A} \subset \mathcal{B} \quad \text{or} \quad \mathcal{B} \supset \mathcal{A}$$

Example 1.19
Let

$$\mathcal{A} = \{1, 2, 3\} \quad \text{and} \quad \mathcal{B} = \{1, 2, 3, 4\}$$

then $\mathcal{A}$ is a proper subset of $\mathcal{B}$. This is because apart from the fact that all the elements in $\mathcal{A}$ are in $\mathcal{B}$, there is at least one element in $\mathcal{B}$, namely, 4, that is not in $\mathcal{A}$, that is. $\mathcal{A} \not\subset \mathcal{B}$. Example 1.15 is also a proper subset.

Improper Subsets

> **Definition 1.14** IMPROPER SUBSET
>
> If every element of set $\mathcal{A}$ is an element of set $\mathcal{B}$, but $\mathcal{B}$ does not have any other elements that are not in $\mathcal{A}$, then set $\mathcal{A}$ is said to be an improper subset of set $\mathcal{B}$

The statement "$\mathcal{A}$ is an improper subset of $\mathcal{B}$" really means that $\mathcal{A} = \mathcal{B}$.

Example 1.20
Let

$$\mathcal{A} = \{1, 2, 3, 4\} \quad \text{and} \quad \mathcal{B} = \{2, 4, 3, 1\}$$

then $\mathcal{A}$ is an improper subset of $\mathcal{B}$, so $\mathcal{A} = \mathcal{B}$.

Note

a) Some authors use the symbol $\subset$ for both proper and improper subsets.
b) If we know only that $\mathcal{A}$ is a subset of $\mathcal{B}$ but do not know if it is proper or improper, then we write $\mathcal{A} \subseteq \mathcal{B}$.
c) We should not confuse the symbol "$\subset$", which stands for "is a proper subset of" with "$\in$" which means "is an element of".

Example 1.21

$$\begin{array}{ll} 2 \in \{1,2,3,4\}, & 2,3 \in \{1,2,3,4\}. \\ \{2\} \subset \{1,2,3,4\}, & \{2,3\} \subset \{1,2,3,4\}. \end{array}$$

1.3 SET OPERATIONS

Given two or more sets, a new set may be derived from them through various operations.

1.3.1 UNION OF SETS

Definition 1.15 UNION OF SETS

The union of $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A} \cup \mathcal{B}$, is the set of all elements which belong to either $\mathcal{A}$ or $\mathcal{B}$ or both.

In symbol, $\mathcal{A} \cup \mathcal{B}$ (read "$\mathcal{A}$ union $\mathcal{B}$" or "$\mathcal{A}$ cup $\mathcal{B}$") may be written as:

$$\mathcal{A} \cup \mathcal{B} = \{x | x \in \mathcal{A} \quad \text{or} \quad x \in \mathcal{B}\}$$

This usage of "or" in the mathematical sense means the same as

$$\mathcal{A} \cup \mathcal{B} = \{x | x \in \mathcal{A} \quad \text{or} \quad x \in \mathcal{B} \quad \text{or} \quad (x \in \mathcal{A} \text{ and } x \in \mathcal{B})\}$$

Example 1.22
Let

$$\mathcal{A} = \{1,2,3,4,5\} \quad \text{and} \quad \mathcal{B} = \{8,9,3,6,7\}$$

Find $\mathcal{A} \cup \mathcal{B}$.

Solution
The union of $\mathcal{A}$ and $\mathcal{B}$ is:

$$\mathcal{A} \cup \mathcal{B} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

When we have many sets, say, $\mathcal{A}_1, \mathcal{A}_2, \cdots$ then

$$\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n = \bigcup_{i=1}^{n} \mathcal{A}_i$$

and

$$\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots = \bigcup_{i=1}^{\infty} \mathcal{A}_i$$

1.3.2 INTERSECTION OF SETS

Definition 1.16 INTERSECTION OF SETS

The intersection of two sets $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A} \cap \mathcal{B}$, is the set of all elements which belong to both $\mathcal{A}$ and $\mathcal{B}$

In symbol, $\mathcal{A} \cap \mathcal{B}$ (read "$\mathcal{A}$ intersection $\mathcal{B}$" or "$\mathcal{A}$ cap $\mathcal{B}$") may be written as:

$$\mathcal{A} \cap \mathcal{B} = \{x | x \in \mathcal{A} \text{ and } x \in \mathcal{B}\}$$

Example 1.23
Refer to Example 1.22. Find $\mathcal{A} \cap \mathcal{B}$.

Solution
The only element that belongs to both $\mathcal{A}$ and $\mathcal{B}$ is 3. Hence

$$\mathcal{A} \cap \mathcal{B} = \{3\}$$

Definition 1.17 DISJOINT SETS

If two sets $\mathcal{A}$ and $\mathcal{B}$ have no elements in common, that is, if $\mathcal{A} \cap \mathcal{B} = \emptyset$, then $\mathcal{A}$ and $\mathcal{B}$ are said to be disjoint

Example 1.24
If

$$\mathcal{A} = \{1, 2, 3, 4, 5\} \quad \text{and} \quad \mathcal{B} = \{a, b, c, d, e\},$$

then

$$\mathcal{A} \cap \mathcal{B} = \emptyset$$

hence $\mathcal{A}$ and $\mathcal{B}$ are disjoint.

1.3.3 COMPLEMENT OF SETS

Definition 1.18 COMPLEMENT OF SETS

The complement of $\mathcal{A}$, denoted by $\mathcal{A}^c$, is the set of all elements in the universal set which are not in $\mathcal{A}$

In symbol, we have

$$\mathcal{A}^c = \{x \mid x \in \mathcal{U}, x \notin \mathcal{A}\}$$

Note

$\mathcal{A}^c$ may also be denoted as $\mathcal{A}'$ or $\overline{\mathcal{A}}$.

Example 1.25
Let $\mathcal{U} = \{1, 2, 3, \cdots, 10\}$ and $\mathcal{A} = \{1, 2, 3, 4, 5\}$
Find $\mathcal{A}^c$.

Solution
$\mathcal{A}^c$ is the set of all elements in $\mathcal{U}$ that are not in $\mathcal{A}$, so that,

$$\mathcal{A}^c = \{6, 7, 8, 9, 10\}$$

Complement of $\mathcal{B}$ Relative to $\mathcal{A}$

Definition 1.19 COMPLEMENT OF $\mathcal{B}$ RELATIVE TO $\mathcal{A}$

If $\mathcal{B} \subseteq \mathcal{A}$, then the set consisting of all elements in $\mathcal{A}$ which do not belong to $\mathcal{B}$ is called the complement of $\mathcal{B}$ relative to $\mathcal{A}$ and is denoted by $\mathcal{B}^c_{\mathcal{A}}$

Note

$\mathcal{B}^c_{\mathcal{A}}$ and $\mathcal{B}$ are disjoint, that is, $\mathcal{B}^c_{\mathcal{A}} \cap \mathcal{B} = \emptyset$

Example 1.26
Let

$$\begin{aligned} \mathcal{A} &= \{1, 2, 3, 4, 5\} \\ \mathcal{B} &= \{1, 2, 3, 6, 7\} \end{aligned}$$

Then

$$\begin{aligned} \mathcal{B}^c_{\mathcal{A}} &= \{4, 5\} \\ \mathcal{B}^c_{\mathcal{A}} \cap \mathcal{B} &= \emptyset \end{aligned}$$

1.3.4 DIFFERENCE OF SETS

Definition 1.20 DIFFERENCE OF SETS

The set consisting of all elements in $\mathcal{A}$ which do not belong to $\mathcal{B}$ is called the difference of $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A} - \mathcal{B}$ or $\mathcal{A} \backslash \mathcal{B}$

In symbol,

$$\mathcal{A}\backslash\mathcal{B} = \{x \mid x \in A, x \notin \mathcal{B}\} = \{x \mid x \in A, x \in \mathcal{B}^c\} = \mathcal{A} \cap \mathcal{B}^c$$

Example 1.27
Let $\mathcal{A} = \{1, 2, 3, 4, 5\}$ and $\mathcal{B} = \{1, 2, 3\}$
Find $\mathcal{A}\backslash\mathcal{B}$.

Solution
$\mathcal{A}\backslash\mathcal{B}$ consists of the elements in $\mathcal{A}$ that are not in $\mathcal{B}$, That is,

$$\mathcal{A}\backslash\mathcal{B} = \{4, 5\}$$

Note

a) If $\mathcal{A} \neq \mathcal{B}$,

$$\mathcal{A} - \mathcal{B} \neq \mathcal{B} - \mathcal{A}$$

If $\mathcal{A}$ and $\mathcal{B}$ are equal (that is, the same set), then

$$\mathcal{A} - \mathcal{B} = \mathcal{B} - \mathcal{A} = \emptyset$$

b) $\mathcal{U}\backslash\mathcal{A} = \mathcal{A}^c$

Symmetric Difference of Sets

Definition 1.21 SYMMETRIC DIFFERENCE OF SETS

The symmetric difference of the sets $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A}\Delta\mathcal{B}$, is the set of elements which belong to exactly one of the two sets

In symbol

$$\begin{aligned} \mathcal{A}\Delta\mathcal{B} &= (\mathcal{A} \cap \mathcal{B}^c) \cup (\mathcal{A}^c \cap \mathcal{B}) \\ &= (\mathcal{A}\backslash\mathcal{B}) \cup (\mathcal{B}\backslash\mathcal{A}) \\ &= (\mathcal{A} \cup \mathcal{B})\backslash(\mathcal{A} \cap \mathcal{B}) \end{aligned}$$

Example 1.28
Refer to Example 1.26. Find $\mathcal{A}\Delta\mathcal{B}$.

Solution

$$\begin{aligned} \mathcal{A} \cup \mathcal{B} &= \{1,2,3,4,5,6,7\} \\ \mathcal{A} \cap \mathcal{B} &= \{1,2,3\} \end{aligned}$$

Hence

$$\mathcal{A} \Delta \mathcal{B} = (\mathcal{A} \cup \mathcal{B}) \backslash (\mathcal{A} \cap \mathcal{B}) = \{4,5,6,7\}$$

1.4 CLASSES OF SETS

Frequently, the elements of a set are sets themselves. For example, each line in a set of lines is a set of points. Usually the word *class* or *family* is used for such a set and the words subclass and sub-family have meanings analogous to subset.

Example 1.29
What are the members of the following class of sets:

$$\mathcal{A} = \{\{1,2\}, \{a,b\}, \{d\}, \{4\}, \{a,3,4\}\}$$

Solution
The members of class $\mathcal{A}$ are the sets

$$\{1,2\}, \{a,b\}, \{d\}, \{4\}, \{a,3,4\}$$

Power Set

Definition 1.22 POWER SET

The power set is the set of all possible subsets of an original set

The power set[3] is denoted as $\mathcal{P}(\mathcal{A})$.

Example 1.30
Let $\mathcal{A} = \{a,b\}$. Find the power set of $\mathcal{A}$.

Solution
$\mathcal{P}(\mathcal{A}) = \{\{a,b\}, \{a\}, \{b\}, \emptyset\}$

Example 1.31
$\mathcal{A} = \{1,2,3\}$. Find the power set of $\mathcal{A}$.

Solution
$\mathcal{P}(\mathcal{A}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$

Note

a) The set $\mathcal{A}$ is a subset of itself.
b) The null set $\emptyset$ is a subset of every set.

Partition of Sets

Definition 1.23 PARTITION OF A SET

A partition of set $\mathcal{A}$ is a subdivision of $\mathcal{A}$ into non-empty subsets which are disjoint and whose union is $\mathcal{A}$

Cells

Definition 1.24 CELLS

The subsets in a partition are called cells

Thus, the class of non-empty sets $\{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_n\}$ forms a partition of the set $\mathcal{A}$ if and only if the following two conditions hold:

a) $\mathcal{P}_i \cap \mathcal{P}_j \quad = \quad \emptyset, \quad \text{if } i \neq j$

b) $\bigcup_{i=1}^{n} P_i \quad = \quad \mathcal{A}$

The sets $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_n$ are the cells.

Example 1.32
Let $\mathcal{A} = \{1, 2, 3, 4, 5, 6\}$. Which of the following form a partition of the set?

a) {{1,3,6}, {2,4}, {2}}
b) {{1,4}, {2}, {3,5,6}}
c) {{1,3}, {2,6}, {5}}
d) {{1,3}, {2,5}, {4,7}, {6}}

Solution
a) This is not a partition of $\mathcal{A}$ since $2 \in \mathcal{A}$ and belongs to both $\{2, 4\}$ and $\{2\}$.
b) This is a partition of the set $\mathcal{A}$ since each element of $\mathcal{A}$ belongs to exactly one cell.
c) This is not a partition of $\mathcal{A}$ since $4 \in \mathcal{A}$ but 4 does not belong to any of the cells.
d) This is not a partition of $\mathcal{A}$ since $7 \in \{4, 7\}$ but $7 \notin \mathcal{A}$.

1.5 LAWS OF ALGEBRA OF SETS

The laws of the algebra of sets are summarized in Table 1.1.

	LAW	UNION	INTERSECTION
1.	Idempotent	$\mathcal{A} \cup \mathcal{A} = \mathcal{A}$	$\mathcal{A} \cap \mathcal{A} = \mathcal{A}$
2.	Commutative	$\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A}$	$\mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A}$

	LAW	UNION	INTERSECTION
3.	Associative	$(\mathcal{A} \cup \mathcal{B}) \cup \mathcal{C} = \mathcal{A} \cup (\mathcal{B} \cup \mathcal{C})$	$(\mathcal{A} \cap \mathcal{B}) \cap \mathcal{C} = \mathcal{A} \cap (\mathcal{B} \cap \mathcal{C})$
4.	Distributive	$\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$	$\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$
5.	Identity	$\mathcal{A} \cup \emptyset = \mathcal{A}$ $\mathcal{A} \cup \mathcal{U} = \mathcal{U}$	$\mathcal{A} \cap \mathcal{U} = \mathcal{A}$ $\mathcal{A} \cap \emptyset = \emptyset$
6.	Complement	$\mathcal{A} \cup \mathcal{A}^c = \mathcal{U}$	$\mathcal{A} \cap \mathcal{A}^c = \emptyset$
7.	De Morgan's	$(\mathcal{A} \cup \mathcal{B})^c = \mathcal{A}^c \cap \mathcal{B}^c$	$(\mathcal{A} \cap \mathcal{B})^c = \mathcal{A}^c \cup \mathcal{B}^c$

Table 1.1 Laws of Set Algebra

Apart from the above seven laws, we have the following ones.

8. $(\mathcal{A}^c)^c = \mathcal{A}; \quad \mathcal{U}^c = \emptyset; \quad \emptyset^c = \mathcal{U}$
9. $\mathcal{A} = (\mathcal{A} \backslash \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B})$
   $\mathcal{A} \cup \mathcal{B} = \mathcal{A} \backslash \mathcal{B}) + \mathcal{B}$
   $(\mathcal{A} \backslash \mathcal{B}) \cap \mathcal{B} = \emptyset$
   $\mathcal{A} \backslash \mathcal{B}) \cap (\mathcal{A} \cap \mathcal{B}) = \emptyset$

Note

The null set $\emptyset$ is the identity for union while $\mathcal{U}$ is the identity for intersection in the sense of Law 5 above.

Many of the laws of the algebra of sets can be generalized to arbitrary unions and intersections. For example,

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$$

and

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$$

Also, De Morgan's laws become

$$\left(\bigcup_{i=1}^{n} A_i \right)^c = \bigcap_{i=1}^{n} A_i^c$$
$$\left(\bigcap_{i=1}^{n} A_i \right)^c = \bigcup_{i=1}^{n} A_i^c$$

Formal proofs of these laws are lengthy. The reader is asked in Exercise 1.19 to illustrate these statements with Venn diagrams discussed in the next section. We must emphasis however, that Venn diagrams are slightly unreliable and are, therefore, not considered sufficient proof of a theorem on set theory. The following theorem will be very useful in Chapter 4.

Theorem 1.2

If $\mathcal{A} \cap \mathcal{B} = \emptyset$ then

$$(\mathcal{A} \cap \mathcal{C}) \cap (\mathcal{B} \cap \mathcal{C}) = \emptyset$$

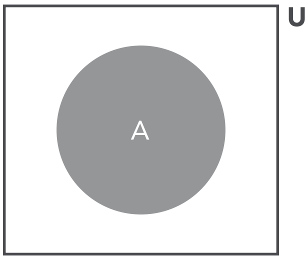
for any set $\mathcal{C}$

The reader will be asked to prove this theorem in Exercise 1.18.

1.6 VENN DIAGRAMS

In our experience with basic statistics, we have probably found that the diagram is often a useful device. In working with sets, diagrams can also be very helpful. Such diagrams are called Venn Diagrams, named after the English mathematician John Venn (1834–1923).

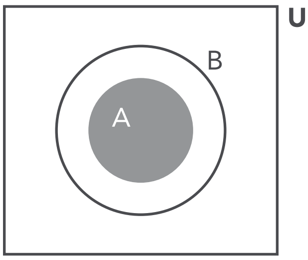
A Venn diagram is a geometrical representation of sets and are useful for visualizing the relationships among sets. A universal set $\mathcal{U}$ is represented geometrically by a rectangle, such as the one shown below. All other sets are drawn as circles within the rectangle.



Set A is shaded

Subsets

Set $\mathcal{A}$ is a proper subset of set $\mathcal{B}$

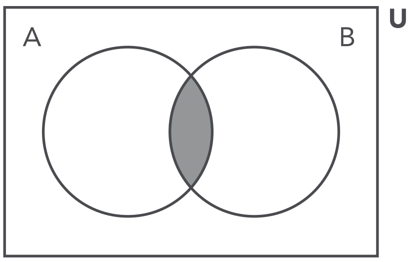


A ⊂ B is shaded

We shall now illustrate set operations using Venn diagrams

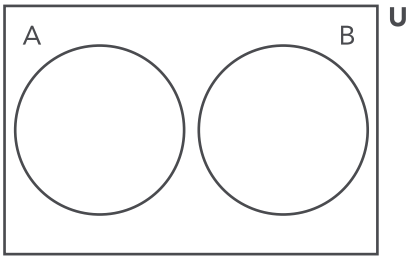
Intersection

a) $\mathcal{A} \cap \mathcal{B} \neq \emptyset$



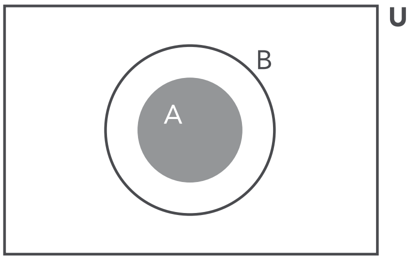
A ∩ B is shaded

b) $\mathcal{A} \cap \mathcal{B} = \emptyset$



Nothing is shaded

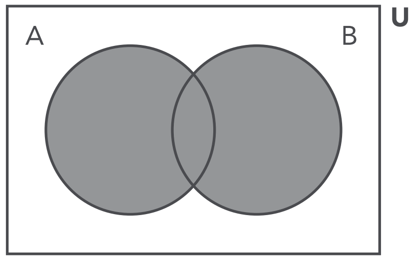
c) $\mathcal{A} \cap \mathcal{B}$ when $\mathcal{A}$ is a subset of $\mathcal{B}$.



A ∩ B is shaded

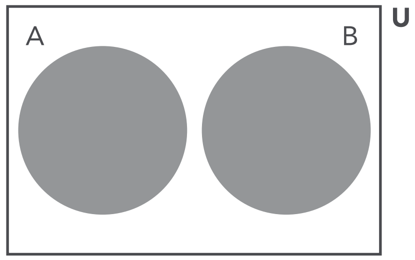
Union

a) $\mathcal{A} \cup \mathcal{B}$ when $\mathcal{A}$ and $\mathcal{B}$ intersect.



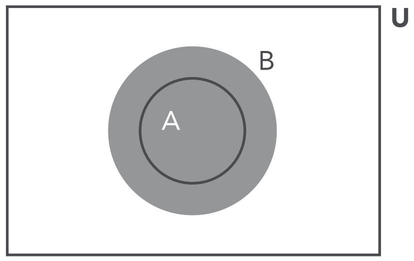
A U B is shaded

b) $\mathcal{A} \cup \mathcal{B}$ when $\mathcal{A}$ and $\mathcal{B}$ are disjoint.



A U B is shaded

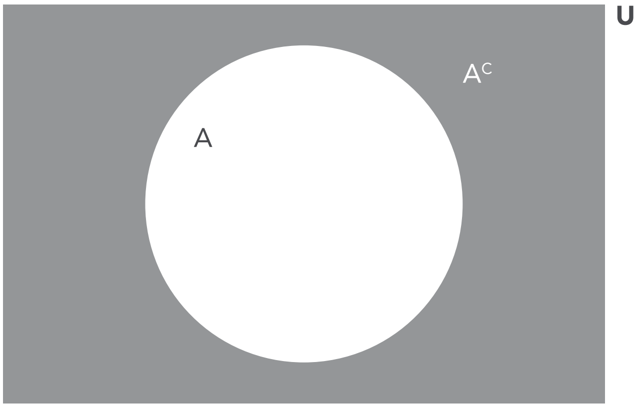
c) $\mathcal{A} \cup \mathcal{B}$ when $\mathcal{A}$ is a subset of $\mathcal{B}$



A U B is shaded

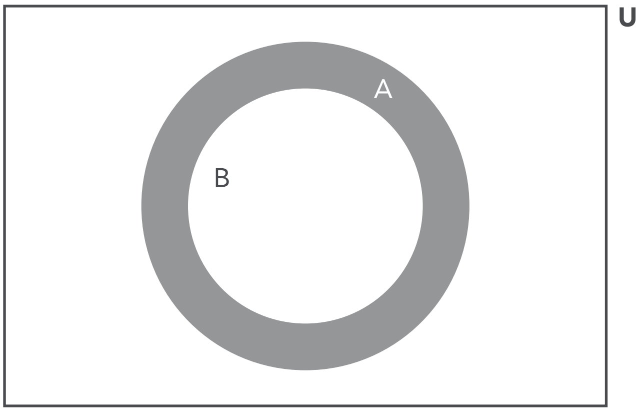
Complement

a) $\mathcal{A}$ complement



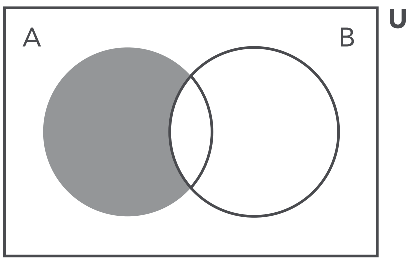
A^C is shaded

b) Complement of $\mathcal{B}$ relative to $\mathcal{A}$



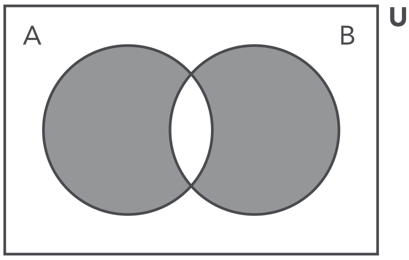
$B^C{}_A$ is shaded

Difference of Sets



A – B is shaded

Symmetric Difference of Sets



A Δ B is shaded

This chapter has reviewed sufficiently the basic concepts of set theory that are crucial in probability theory as presented in this book: definition and manipulation of sets of objects, operations on sets to obtain other sets, special relationships among sets, and so on. It will help us simplify word problems in probability problems. As we shall see in later chapters, the mathematics of probability is expressed most naturally in terms of sets.

EXERCISES

1.1 Let $\mathcal{A}$ be the set of all positive odd numbers less than 10. Describe $\mathcal{A}$ by
(a) the roster method (b) the property method.

1.2 Let $\mathcal{A}$ be the set of real numbers that satisfy the following equation: $2x = 10$. Show how to describe A by

(a) the roster method (b) the property method.

1.3 In Exercise 1.2, is it true that $\mathcal{A} = 5$?

1.4 Let $\mathcal{A} = \{x|x^3 = 8\}$, $\mathcal{B} = \{x|x - 3 = -1\}$. Is $\mathcal{A} = \mathcal{B}$.

1.5 Let $\mathcal{A} = \{x|x^2 + 5x - 6 = 0\}$, $\quad \mathcal{B} = \{x|x^2 = 4\}$,

$C = \{x|x^2 + 1 = 1\}$, $\quad \mathcal{D} = \{x|x \neq x\}$,
$\mathcal{E} = \{x|x^3 = 8; 5x = 15\}$

Determine which of the following statements are true.

$(a)\quad \mathcal{A} \subset \mathcal{B}$ $\qquad (b)\quad \mathcal{C} \subset \mathcal{A}$,
$(c)\quad \mathcal{D} = \emptyset$ $\qquad (d)\quad E = \{\emptyset\} = \{0\}$
$(e)\quad E \supset \mathcal{A}$ $\qquad (f)\quad E = \{0\}$

1.6 Which of these sets are equal?

$(a)\quad \mathcal{A} = \{1, 2, 3, 4\}$, $\quad \mathcal{B} = \{2, 4, 3, 1\}$, $\quad \mathcal{C} = \{4, 1, 3, 2\}$
$(b)\quad \mathcal{A} = \{1, 2, 3\}$, $\quad \mathcal{B} = \{a, b, c\}$
$(c)\quad \mathcal{A} = \{a, b, c\}$, $\quad \mathcal{B} = \{a, c, d\}$, $\quad \mathcal{C} = \{b, a, c\}$
$(d)\quad \mathcal{A} = \{1, 2, 3\}$, $\quad \mathcal{B} = \{1, 2, 3, 4\}$

1.7 In Exercise 1.6, which of the sets are equivalent?

1.8 In Exercise 1.6, which of the sets are equivalent as well as equal?

1.9 Prove that if $\mathcal{A} \subseteq \emptyset$, then $\mathcal{A} = \emptyset$.

1.10 Prove that if $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{C}$, then $\mathcal{A} \subseteq \mathcal{C}$.

1.11 State whether each set is finite or infinite.

a) The letters in English alphabet.
b) The number of rivers on the earth.
c) The numbers which are greater than 10.
d) The lines in a plane.
e) The days of the week.
f) The set $\mathcal{Q}$ of rational numbers.
g) The set $\mathcal{R}$ of real numbers.

1.12 In Exercise 1.11, which of the sets are countable?

1.13 Let

$$\begin{array}{ll} \mathcal{U} = \{1,2,3,4,5,6,7,8,9,10\}, & \mathcal{A} = \{1,2,3,4,5,6\} \\ \mathcal{B} = \{4,5,6,7\}, & \mathcal{C} = \{2,4,6,8,10\} \end{array}$$

Find

(a) $\mathcal{A} \cup \mathcal{B}$	(b) $\mathcal{A} \cap \mathcal{B}$	(c) $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$
(d) $(\mathcal{A} \cap \mathcal{B})^c \cap \mathcal{C}$	(e) $\mathcal{A}^c$	(f) $(\mathcal{A} \cup \mathcal{B})^c$
(g) $(\mathcal{A} \cap \mathcal{B})^c$	(h) $\mathcal{A}^c \cup \mathcal{B}^c$	(i) $\mathcal{A}^c \backslash \mathcal{B}^c$
(j) $\mathcal{A}^c \cap \mathcal{B}$	(k) $\mathcal{A}^c \cup \mathcal{B}$	(l) $\mathcal{A} \backslash \mathcal{B}$

What is the cardinality of each of the above sets?

1.14 Draw a Venn diagram and shade the following when no set is a subset of the other.

(a) $\mathcal{A}^c$	(b) $(\mathcal{A} \cap \mathcal{B})^c$	(c) $\mathcal{A} \cap \mathcal{B} = \mathcal{A}^c$
(d) $\mathcal{A} \cup \mathcal{B} = \emptyset$	(e) $\mathcal{A} \cup \mathcal{B} = \mathcal{A} \cap \mathcal{B}$	

1.15 Prove that if $\mathcal{A} \subseteq \mathcal{C}$ and $\mathcal{B} \subseteq \mathcal{C}$, then $(\mathcal{A} \cup \mathcal{B}) \subseteq \mathcal{C}$.

1.16 Prove that if $\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{A} \subseteq \mathcal{C}$, then $\mathcal{A} \subseteq (\mathcal{B} \cap \mathcal{C})$.

1.17 Show that (a) $\mathcal{A} - \mathcal{B} = \mathcal{A} \cap \mathcal{B}^c$ (b) $\mathcal{A} - \mathcal{B} \subset \mathcal{A} \cup \mathcal{B}$ (c) $\mathcal{A}^c - \mathcal{B}^c = \mathcal{B} \cap \mathcal{A}$.

1.18 Prove Theorem 1.2.

1.19 Use the Venn diagrams to illustrate the laws of the algebra of sets.

1.20 Let the set of classes

$$\mathcal{A} = \{\{a,b,c\},\{d,e\},\{f,g,h\}\}.$$

Determine the correct statements in the following and explain.

(a) $\{a,b,c\} \in \mathcal{A}$	(b) $\{a,b,c\} \subset \mathcal{A}$
(c) $\{\{a,b,c\}\} \subset \mathcal{A}$	(d) $\{\{a,b,c\}\} \in \mathcal{A}$

1.21 Find the power set $\mathcal{P}(\mathcal{A})$ of the set $\mathcal{A} = \{a,b,c\}$.

1.22 Let $\mathcal{A} = \{a,b,c,d\}$. Which of the following are partitions of $\mathcal{A}$?

(a) $\{\{a,b\},\{c\},\{c,d\}\}$	(b) $\{\{a\},\{b,d\},\{c\}\}$
(c) $\{\{a,b,c\},\{d\}\}$	(d) $\{\{a,c\},\{d\},\{b\}\}$
(e) $\{\{a,d\},\{c\}\}$	(f) $\{a,b,c,d\}$

1.23 Find all the partitions of $\mathcal{B} = \{1, 2, 3\}$.

1.24 Prove the associative laws of sets.

1.25 Prove the distributive laws of sets.

1.26 Prove the De-Morgan's laws of sets.

1.27 For any three sets $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, show that

$$\mathcal{A} \cap (\mathcal{B} - \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) - (\mathcal{A} \cap \mathcal{C})$$

1.28 Let $\mathcal{A}$ and $\mathcal{B}$ be any two sets. Prove that

$$(\mathcal{A} - \mathcal{B}) \cup (\mathcal{B} - \mathcal{A}) = (\mathcal{A} \cup \mathcal{B}) - (\mathcal{A} \cap \mathcal{B})$$

2 COUNTING PRINCIPLES

2.1 INTRODUCTION

Counting is the basis of probability and statistics. Often the number of elements in a particular set or the number of possible outcomes of a particular experiment, as we shall see later, is not very large and so direct enumeration or counting is not difficult. However, problems arise in probability where direct counting becomes a practical difficulty. For example, if a team of five players are to be formed with positions from ten players, then each of the choices of five players (elements) of the ten players (in the universal set) is 5-tuple. The total number of all possible subsets in the universal set would be very difficult to list[4]. A more sophisticated way of counting such large numbers of outcomes falls within th realm of a branch of mathematics referred to as *combinatorics* or *combinatorial analysis* or *combinatorial mathematics*.

Combinatorics is the branch of mathematics concerned with problems of selection, arrangement, and operation within a finite or discrete system. It studies, among other things, methods for counting the number of ways in which an action or sequence of actions can be performed. Its main objective is how to count without counting.

The problem of counting begins with drawing r objects[5] of a specified type from a class or group of n objects, a technique known as sampling. The number of ways in which this can be done depends on whether

a) the objects or the classes can be distinguished or not;
b) the order of classes or the order of objects in a class is relevant or not;
c) the objects can be used more than once or not at all;
d) empty classes are allowed or not.

Here in this text, we shall consider only cases having applications in probability problems. Other aspects found in the analysis of algorithms, coding theory, medicine, quality control, communications, agriculture, genetics, quantum mechanics, and so on, are treated in books devoted to combinatorial theory.

2.2 SAMPLING WITH OR WITHOUT REPLACEMENT

Suppose we wish to draw a sample of r objects from a population of n distinct objects. We may take the r objects all at a time or draw one object at a time until the r objects are obtained. If we draw the objects one at a time, we can distinguish between *sampling with replacement* and *sampling without replacement*.

Definition 2.1 SAMPLING WITH REPLACEMENT

If, after an object is drawn and a note is taken of it, it is put back into the population before another object is drawn then it is sampling with replacement

In this case, the same object may appear several times in the sample.

Definition 2.2 SAMPLING WITHOUT REPLACEMENT

If, after an object is drawn, it is put aside before the next one is drawn, until all the r objects are drawn, it is sampling without replacement

In this case, no two objects in the sample may be the same.

When drawing r objects one at a time, the order in which the objects appear in the sample is very important. We may then have to distinguish between *ordered* and *unordered* samples. Consider a population set (or simply the population) of n objects, $a_1, a_2, \cdots, a_n$. We define an *ordered sample* of size r drawn from this population as any ordered arrangement $a_{j1}, a_{j2}, \cdots, a_{jr}$ of r objects. In combinatorics, whether order matters is the principal factor that distinguishes *combinations* and (permutations), the two main techniques for counting. If order is important in the selection of the r objects, it is a permutation problem and if not, a combination problem. These two main techniques for counting are both in turn based on a simple concept known as the *principle of multiplication* which we discuss in the next section.

2.3 ADDITION AND MULTIPLICATION PRINCIPLES OF COUNTING

2.3.1 ADDITION PRINCIPLE OF COUNTING

The most basic of the counting principle is the *addition principle*.

Theorem 2.1 ADDITION PRINCIPLE OF COUNTING

Suppose $\mathcal{A} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_n$, where $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_n$ are disjoint subsets of some finite universal set. Then

$$n(\mathcal{A}) = n(\mathcal{B}_1) + n(\mathcal{B}_2) + \cdots + n(\mathcal{B}_n)$$

Example 2.1
A class of students consists of 6 Ghanaian, 8 Nigerians and 2 Italians. In how many ways can a Ghanaian or an Italian be drawn from the class if none of them has a dual citizenship.

Solution
Let $\mathcal{A}$ denote the set of students who are Ghanaians or Italians;

$\mathcal{B}_1$ denote the set of students who are Ghanaians;
$\mathcal{B}_2$ denote the set of students who are Italians.

Then $\mathcal{B}_1$ and $\mathcal{B}_2$ are disjoint, since no student can be a Ghanaian and an Italian at the same time. Hence

$$\mathcal{A} = 6 + 2 = 8$$

Whenever we have a simple action to perform but that action must satisfy one condition or another where the conditions cannot happen together, we normally use the addition principle.

Theorem 2.2 INCLUSION-EXCLUSION PRINCIPLE

Suppose $\mathcal{A}$ and $\mathcal{B}$ are subsets of some finite universal set. Then

$$n(\mathcal{A} \cup \mathcal{B}) = n(\mathcal{A}) + n(\mathcal{B}) - n(\mathcal{A} \cap \mathcal{B})$$

The inclusion-exclusion principle tells us that when we add together the number of elements in $\mathcal{A}$ and the number of elements in $\mathcal{B}$, the elements in $\mathcal{A} \cap \mathcal{B}$ are counted twice. To find $n(\mathcal{A} \cup \mathcal{B})$, we must add $n(A)$ to $n(B)$ and subtract $n(\mathcal{A} \cap \mathcal{B})$.

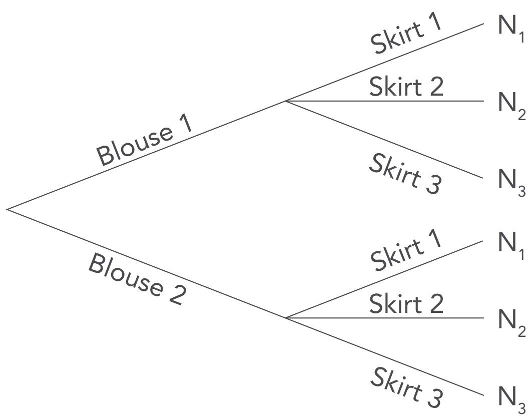
2.3.2 MULTIPLICATION PRINCIPLE OF COUNTING

Sometimes we shall be faced with situations that involve processes consisting of successive actions. In such situations we are more likely to use the multiplication principle. We shall first introduce the tree diagram which is the type of reasoning behind the multiplication principle of counting.

The *tree diagram* is a device used to enumerate all the possible outcomes of sequence of processes where each process can occur in a finite number of ways. It is constructed from left to right. Each complete path is called a branch of the tree. The number of branches at each point correspond to the number of possible outcomes of the next process.

Example 2.2
If a woman has 2 blouses and 3 skirts, how many outfit can she put together.

Solution
The following is the tree diagram. There are 6 outfits she can put together.



Theorem 2.3 MULTIPLICATION PRINCIPLE OF COUNTING

If one thing can be accomplished in n_1 different ways and after this a second thing can be accomplished in n_2 different ways, …, and finally a k^{th} thing can be accomplished in k^{th} ways, then all k things can be accomplished, in specified order, in

$$n_1 \times n_2 \times \cdots \times n_k \text{ different ways}$$

The multiplication principle of counting is also called the *fundamental principle of counting*.

Example 2.3
Use the multiplication principle of counting to solve the problem in Example 2.2.

Solution
There are 2 ways the woman can put on the two blouses and for each of these ways there are 3 ways she can put on the three skirts. Hence, by the fundamental principle of counting, there are $2 \times 3 = 6$ ways of putting outfits together.

Example 2.4
Six dice are rolled. In how many ways may the faces of the dice show up.

Solution
Each face of the die may show up in six different ways so by the fundamental principle of counting there are

$$6 \times 6 \times 6 \times 6 \times 6 \times 6 = 6^6 = 46656 \text{ ways}$$

Example 2.5
Given the numbers, $1, 2, 3, 4,$ how many different numbers of three digits can be formed from them if repetitions are (a) allowed (b) not allowed.

Solution

a) Since repetitions are allowed, each digits can be chosen in four different ways. Hence there are

$$4 \times 4 \times 4 = 64 \text{ such numbers}$$

b) Since repetitions are not allowed, the first digit can be chosen in four ways, the second in three ways, and the remaining digit in two choices. Hence there are

$$4 \times 3 \times 2 = 24 \text{ numbers in all}$$

2.3.3 FACTORIAL

Another concept that we need to know before starting discussions on the main counting techniques is the *factorial*.

> **Definition 2.3** FACTORIAL
>
> Given the positive integer n, the product of all the whole numbers from n down to 1 is called n factorial

Symbolically, n factorial, written $n!$ is defined as

$$n! = n(n-1)(n-2)\cdots 2 \,.\, 1$$

In particular, we see that

$$\begin{aligned} 1! &= 1 \\ 2! &= 2 \,.\, 1 = 2 \\ 3! &= 3 \,.2 \,.\, 1 = 6 \\ 4! &= 4 \,.\, 3.\ 2 \,.\, 1 = 24 \end{aligned}$$

Finally, it can be shown in Chapter 10 that $0! = 1$

Example 2.6
Evaluate 10!.

Solution

$$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3,628,800$$

We may compute further values of factorial by the relation:

$$(n+1)! = (n+1)n! \quad and \quad n! = n(n-1)!$$

For instance

$$\begin{aligned} 10! = 10(10-1)! &= 10 \times 9! \\ &= 10 \times (9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1) \\ &= 10 \times 362,880 = 3,628,800 \end{aligned}$$

For large n, the functions, $n!$ is very large and convenient approximation can be obtained by the Stirling formula, suggested by a Scot Mathematician, James Stirling $(1692 - 1770)$[6]:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

where $\sim$ is read "asymptotically equal" and means that the ratio of the two sides approaches 1 as n approaches infinity; $\pi = 3.14159$ and $e = 2.71828$ is the base of natural (Napierian) logarithms.

For example

$$40! = \sqrt{2(3.14159)(40)}\left(\frac{40}{2.71828}\right)^{40} = 8.142 \times 10^{47}$$

The actual value by the calculator is 8.159×10^{47}.

2.4 PERMUTATIONS: ORDERED SELECTION

2.4.1 n- AND r-PERMUTATIONS

A special case of the fundamental principle of counting occurs when we make successive choices from a single set of n objects. The first choice may be made in n ways, and for each of these ways there are $n-1$ ways for the second choice, then $n-2$ ways for the third, and so on.

Definition 2.4 PERMUTATION
(**Ordered Selection**)

A permutation is an arrangement of objects in a given order

Definition 2.5 n-PERMUTATION

An arrangement of n distinct objects in a given order taking all of them at a time is called an n-permutation

The number of n-permutations is usually denoted by ${}_nP_n$, $P(n, n)$ or ${}^nP_n$.

Theorem 2.4

The number of permutations of n distinct objects is $n!$

Note

Theorem 2.4 may equivalently be stated as "the number of arrangements of n objects in a *row*".

Example 2.7
Consider the set of letters a, b and c. In how many ways can the three letters be arranged taking all of them together. Indicate the arrangements.

Solution
Since order is important in this case, it is a permutation problem. There are 3 letters and all are to be taken together simultaneously so

$${}_3P_3 = 3! = 3 \times 2 \times 1 = 6$$

The arrangements are: $abc, acb, bac, bca, cab, cda$

Definition 2.6 r-PERMUTATION

An arrangement of objects in a given order taking r at a time from a set of n $(r \leq n)$ objects, is called an r-permutations

The number of r-permutations is usually denoted as ${}_nP_r$, $P(n, r)$ or ${}^nP_r$. This definition assumes that the permutations are *without repetition*.

Permutation Without Repetition

Theorem 2.5

The number of permutations or arrangements of r objects *without replacements* from a set of n distinct objects is given by

$${}_nP_r = n(n-1)(n-2)\cdots(n-r+1)$$

or equivalently,

$${}_nP_r = \frac{n!}{(n-r)!}$$

Note

If $r = n$, then ${}_nP_r = {}_nP_n = n!$

Example 2.8
In how many different ways can 4 people be chosen from a set of 6 and be seated in a row of four chairs?

Solution
$n = 6, \ r = 4.$
Hence, the desired different number of ways is

$$\begin{aligned} {}_6P_4 &= \frac{6!}{(6-4)!} \\ &= \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} \\ &= 360 \end{aligned}$$

Permutation With Repetition

In Theorem 2.5, we found out that permutation without repetition gives us the usual formula we know for permutation. We now give the formula for permutation with repetition.

Theorem 2.6

The number of permutations or arrangements of r objects with repetition from n distinct objects is n^r

Example 2.9
In how many different ways can a three-letter word be formed from the letter BEST if any letter may be repeated.

Solution
$n = 4, r = 3.$
Hence, there are $4^3 = 64$ different ways

2.4.2 CIRCULAR PERMUTATIONS

One important variation of r-permutations is when objects are arranged in a circle. Typically in this case the actual positions do not matter, but the relative positions (that is, which objects are next to one another) matter. For example, if six people are sitting in a circle we do not get a new permutation if they all move one position in a clockwise (or anti-clockwise) direction. The first object can be placed anywhere. Starting from this object and moving around clockwise or anti-clockwise there are ${}_nP_{n-1} = (n-1)!$ possibilities.

Theorem 2.7

The number of permutations of n distinct objects in a circle is

$$(n-1)!$$

Example 2.10
Five Executives attend a round-table meeting. How many different arrangements are possible?

Solution
There are $(n-1)! = (5-1)! = 24$ circular permutations.

Example 2.11
Refer to Example 2.10. Suppose each of the Executives was accompanied by the Secretary to take minutes of the meeting.

a) How many arrangements are possible that alternate the Executives and the Secretaries.
b) If a Secretary should sit by his/her Executive, how many arrangements are possible that alternate the Executive and the Secretaries?

Solution
Any person (an Executive or a Secretary) can be placed anywhere at the start.

a) Suppose the first to sit down is an Executive. Then there are $(5-1)! = 4!$ different arrangements for the remaining Executives. The five Secretaries can be seated in the next 5 alternating seats. Thus, there are $5!$ possibilities for them. By the multiplication principle, there are
$$4!5! = 2880 \text{ different arrangements}$$
b) Suppose the first to sit down is an Executive. Then there are $(5-1)!$ different arrangements. There are two ways the first Secretary can sit, either at the left or the right of her Executive. Once she sits, all other places are automatic for the rest of the Secretaries. Hence, there are
$$4!(2) = 48 \text{ possible arrangements}$$

So far, we have been assuming that all objects considered are distinct. What happens if they are not?

Theorem 2.8

The number of permutations (without repetition) of n objects of which n_1 are alike, n_2 are alike, $\cdots$, n_k are alike is

$$M_n = {}_nP_{n_1,n_2,\cdots,n_k} = \frac{n!}{n_1!n_2!\cdots n_{k!}}$$

where $n = n_1 + n_2 + \cdots + n_k$

The number M_n is known as a multinomial coefficient. A particularly important case that arises frequently is when $n = 2$. This is the binomial coefficient (see below).

If $n_1 = n_2 = \cdots = n_k = 1$ (that is, if $k = n$), then M_n is simply $n!$.

Example 2.12
In how many ways can the letters of the word $N\ E\ C\ E\ S\ S\ I\ T\ I\ E\ S$ be arranged.

Solution
The total number of letters in the word is 11. The letters E and S each occurs 3 times. The letter I occurs twice and the letters N, C and T each occurs once. Hence the total number of arrangements is

$$\frac{11!}{3!3!2!1!1!1!} = 554,400$$

Permutations Identities

a) ${}_nP_n = n!$
b) ${}_nP_n = n!$
c) ${}_nP_1 = n$
d) ${}_nP_r = n[P(n-1, r-1)], \quad n > 0, \quad r > 0$

2.5 COMBINATIONS: UNORDERED SELECTION

In a permutation, we are interested in the order of arrangement of the objects. Thus ab is a different permutation from ab. In some problems, the order in which objects are arranged is not important. We may be interested only in selecting or choosing objects without regard to order.

Definition 2.7 COMBINATION
(**Unordered Selection**)

A combination is a selection of objects without regard to their order

Definition 2.8 r-COMBINATION

The r-combination is the total number of combinations of a set of n objects taken r at a time, $n \geq r$

Similar to permutations, there are two types of combinations, ***without repetition*** and ***with repetition***.

Combination Without Repetitions

Definitions 2.9 COMBINATION WITHOUT REPETITION

The number of combinations of n distinct objects, taken r at a time $(n \geq r)$ *without repetitions* is the number of arrangements that can be made from the n given objects, each arrangement containing r different objects and no two arrangements containing exactly the same objects

Example 2.13
How many combinations are there of three letters a, b, c taken two letters at a time, without repetitions?

Solution
There are 3 such combinations: ab, ac, bc.

Suppose 3 letters are chosen from 4 letters. If order is important then it is a permutation problem. The number of permutations of 3 letters chosen from 4 letter is

$$_4P_3 = \frac{4!}{(4-3)!} = 4 \times 3 \times 2 = 24$$

However, any set of 3 letters will appear 6 (or 3!) times in the list of all possible arrangements. For instance, the set containing letters a, b, and c will appear as abc, acb, bac, cab and cba. Hence the total number of ways of selecting 3 letters is

$$\frac{4!}{3!(4-3)!} = \frac{4 \times 3 \times 2 \times 1}{3!1!} = 4$$

In general, the number of ways of selecting r objects from n distinct objects where order is not important is

$$_nC_r = \frac{_nP_r}{r!}$$

When this formula is expanded, it will give us the following Theorem.

Theorem 2.9

The number of combinations of a set of n distinct objects taken r at a time *without repetition*, $n \geq r$ is given by

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

for $0 \leq r \leq n$

$\binom{n}{r}$ is also called the binomial coefficient for reasons that will become clear in Section 2.6. Usually, when we talk of r-combination, we refer to combination *without repetition*. The number of r-combinations is usually denoted by ${}_nC_r$, $C(n, r)$ or by $\binom{n}{r}$ and read as "n combination r" or "n choose r". Throughout in this text we shall adopt the notation $\binom{n}{r}$.

Example 2.14

A school basketball squad for the inter-school competition has ten players. The coach must select a team for the first tournament.

a) How many different teams of five players can be constituted for this tournament?
b) If, in constituting the team, the coach also has to designate positions, how many different teams of five players can be constituted?

Solution

a) Here, we are not interested in the positions each of the five players in the team will take. It is, therefore, a problem of combination, and

$$\begin{aligned} \binom{10}{5} &= \frac{10!}{5!(10-5)!} \\ &= \frac{10 \times 8 \times 7 \times 6 \times 5!}{5 \times 4 \times 3 \times 2 \times 1 \times 5!} \\ &= 252 \quad \text{combinations.} \end{aligned}$$

b) Since the order counts in this case, the problem is one of permutation Hence,

$$\begin{aligned} {}_{10}P_5 &= \frac{10!}{(10-5)!} \\ &= \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5!}{5!} \\ &= 30,240 \quad \text{permutations.} \end{aligned}$$

Example 2.15

A Committee of 5 is to be formed from 12 men and 8 women. In how many ways can the Committee be chosen so that there are 3 men and 2 women on it.

Solution

Number of ways of choosing 3 men from 12 is

$$\binom{12}{3} = \frac{12!}{3!9!} = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = 220$$

Number of ways of choosing 2 women from 8 is

$$\binom{8}{2} = \frac{8!}{2!6!} = \frac{8 \times 7}{2 \times 1} = 28$$

By the multiplication principle, the total number of ways of forming the Committee is

$$\binom{12}{3} \times \binom{(8}{2} = 220 \times 28 = 6160$$

Combination With Repetition

Definition 2.10 COMBINATION WITH REPETITION

The number of combinations of n distinct objects, taken r at a time, *with repetitions* is the number of arrangements that can be made up of the r objects chosen from the given objects, each being used as often as desired

Example 2.16
How many combinations are there of three letters a, b, c taken two letter at a time, with repetitions.

Solution
There are 6 such combinations ab, ac, bc, aa, bb, cc.

Theorem 2.10

The number of different combinations of a set of n distinct objects taken r at a time *with repetitions* $(n \geq r)$, is given by

$$\binom{n+r-1}{r}$$

Note

In general,

$$\binom{n+r}{r} = \binom{n+r}{n}$$

That is, the formula is symmetric in n and r. Hence,

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$$

Example 2.17
Refer to Example 2.14. Use Theorem 2.10 to rework it.

Solution

$n = 3,\ \ r = 2.$

Hence

$$\binom{3+2-1}{2} = \binom{4}{2} = 6$$

Theorem 2.10 can be used to solve one peculiar type of counting problems, that is, the selection of an r-element subset of a set of size n. The objects were considered to be distinct with $n \geq r$.

Suppose there is a set of n elements, $S = \{s_1, s_2, \cdots, s_n\}$ and r elements have to be selected from S, where each element can be selected an arbitrary number of times but the order of selection does not matter. Then there are

$$\binom{r+n-1}{r}$$

different selections that can be made.

In this particular type of problems, r can be greater than n.

Example 2.18
Three types of beverages, tea, coffee and cocoa, are to be served at the Academic Board meeting. There are 12 members present at the meeting. How many different beverage orders are possible.

Solution
Here the set S contains three elements, namely, tea, coffee and cocoa; and $r = 12$ selections are to be made from S (the twelve members, each selecting one kind of beverage). Thus $n = 3$ and $r = 12$

Hence there are

$$\binom{12+3-1}{12} = \binom{14}{12} = 91 \text{ different possibilities}$$

One way of stating this problem is "distribute r similar balls in n distinct boxes". If, in placing the balls in the boxes, *no box should be empty*, then there are

$$\binom{r-1}{n-1}$$

distinguishable ways of doing that (see Exercise 2.29).

Combination Identities

a) $\binom{n}{0} = 1$

b) $\binom{n}{1} = n$

c) $\binom{n}{n} = 1$

d) $\binom{0}{0} = 1$

e) $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$

f) $\binom{n}{r} = 0, \quad \text{if} \;\; r > n \;\; \text{or} \;\; r < 0$

g) $\binom{n}{r} = \binom{n}{n-r}$

h) $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}, \quad r = 1, 2, \cdots, n-1$

i) $\binom{n}{r} = \binom{n+1}{r+1} - \binom{n}{r+1}, \quad r = 1, 2, \cdots, n-1$

j) $\binom{n}{r} = \frac{n}{n-r}\binom{n-1}{r}, \quad \text{if} \quad r < n$

k) $\binom{n}{r} = \frac{n-r+1}{r}\binom{n}{r-1}, \quad \text{if} \quad r > 0$

l) $\binom{n}{r} = \frac{n}{r}\binom{n-1}{r-1}, \qquad n > 0, \;\; r > 0$

In Table 2.1, we have summarized what we call the basic rules of counting.

Counting Procedure	Repetition	Without Repetition
Permutation (ordered)	n^r	$\dfrac{n!}{(n-r)!}$
Combination (unordered)	$\binom{n+r-1}{r}$	$\dfrac{n!}{r!(n-r)!}$

In addition to the basic rules summarized in Table 2.1, the number of ordered pairs that can be formed when there are m choices for the first element and n choices for the second is mn.

2.6 BINOMIAL THEOREM

We shall be concerned with the expansion of $(1+x)^n$ in a series of terms each of which is a power of x multiplied by a quantity depending on n.

2.6.1 BINOMIAL THEOREM FOR NON-NEGATIVE INTEGER

When n is small, it is possible to expand the expression $(1+x)^n$ by multiplying out.

Example 2.19

$n = 0, \;\; (1+x)^0 = 1,$

$n = 1, \;\; (1+x)^1 = 1 + x,$

$n = 2, \;\; (1+x)^2 = (1+x)(1+x)^1 = 1 + 2x + x^2$

$n = 3, \;\; (1+x)^3 = (1+x)(1+x)^2 = 1 + 3x + 3x^2 + x^3$

$n = 4, \;\; (1+x)^4 = (1+x)(1+x)^3 = 1 + 4x + 6x^2 + 4x^3 + x^4$

$n = 5, \;\; (1+x)^5 = (1+x)(1+x)^4 = 1 + 5x + 10x^2 + 10^3 + 5x^4 + x^5$

Similarly, to obtain the expansion $(1 + x)^6$ we multiply the expansion of $(1 + x)^5$ by $(1 + x)$. This process could be continued indefinitely. The expansions on the right hand side of Example 2.19 have coefficients which correspond to the symmetrical array:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

known as *Pascal's triangle*, named after Blaise Pascal (1623–1662), a French mathematician, and one of the founders of the science of probability. There are two obvious characteristics of the Pascal's triangle:

a) Each row starts and terminates with a one;
b) Each number in Pascal's triangle with the exception of the two 1's on each row, is the sum of the two numbers immediately above it.

Example 2.20
Construct the row of the Pascal triangle corresponding to $n = 6$

Solution

$$1 \quad (1+5) \quad (5+10) \quad (10+10) \quad (10+5) \quad (5+1) \quad 1$$

That is

$$1 \qquad 6 \qquad 15 \qquad 20 \qquad 15 \qquad 6 \qquad 1$$

The process of multiplying out, as in Example 2.16, or using Pascal's triangle is not a satisfactory method of obtaining the expansion of $(1+x)^n$ for large values of n. A better method is to use what is called the *Binomial theorem*.

Theorem 2.11 BINOMIAL THEOREM

If n is a positive integer, then

$$(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r$$

When Theorem 2.11 is expanded, we obtain

$$\begin{aligned}
(1+x)^n &= 1 + \frac{nx}{a} + \frac{n(n-1)x^2}{1.2} + \frac{n(n-1)(n-2)x^3}{1.2.3} + \cdots + x^n \\
&= \binom{n}{0} x^0 + \binom{n}{1} x^1 + \binom{n}{2} x^2 + \binom{n}{3} x^3 + \cdots + \binom{n}{n} x^n \\
&= \sum_{r=0}^{n} \binom{n}{r} x^r
\end{aligned}$$

Some Identities of Sums of Binomial Coefficients

a) $\displaystyle\sum_{r=0}^{n} \binom{n}{r} = 2^n$

b) $\displaystyle\sum_{r=0}^{n} \binom{n}{r}^2 = \binom{2n}{n}$

c) $\displaystyle\sum_{r=0}^{n} \binom{p}{r}\binom{q}{n-r} = \binom{p+q}{n}$

d) $\displaystyle\sum_{r=0}^{n}\binom{r+1}{k}=\binom{n+k+1}{k+1}$

e) $\displaystyle\sum_{r=0}^{n}(-1)^r\binom{n}{r}=0$

f) $\displaystyle\sum_{r=0}^{n}\binom{n}{r}(a-1)^r=a^n$

g) $\displaystyle\sum_{r=0}^{n}r\binom{n}{r}=n2^{n-1}$

Example 2.21
Use the binomial theorem to expand (a) $(1+x)^5$, (b) $(1-y)^5$

Solution

a) From Theorem 2.11,

$$\begin{aligned}(1+x)^5 &= \binom{5}{0}x^0+\binom{5}{1}x^1+\binom{5}{2}x^2+\cdots+\binom{5}{5}x^5\\ &= 1+5x+10x^2+10x^3+5x^4+x^5\end{aligned}$$

as in Example 2.19.

b) If we replace x in (a) by $-y$, the result follows immediately. Thus

$$\begin{aligned}(1-y)^5 &= [1+(-y)]^5\\ &= 1+5(-y)+10(-y)^2+10(-y)^3+5(-y)^4+(-y)^5\\ &= 1+5(-y)+10(-y)^2+10(-y)^3+5(-y)^4+(-y)^5\\ &= 1-5y+10y^2-10y^3+5y^4-y^5\end{aligned}$$

A more general form for the binomial theorem when n is a positive integer is

$$\begin{aligned}(y+z)^n &= \binom{n}{0}y^nz^0+\binom{n}{1}y^{n-1}z^1+\binom{n}{2}y^{n-2}z^2+\cdots\\ &\quad+\binom{n}{n-1}y^1z^{n-1}+\binom{n}{n}y^0z^n\\ &= \sum_{r=0}^{n}\binom{n}{r}y^{n-r}z^r\end{aligned}$$

We re-emphasize the properties of the expansion of $(y+z)^n$ written in descending order with respect to y.

a) There are $n+1$ terms;
b) Coefficients of the first and last terms are 1;
c) Coefficients of the terms are symmetrical with respect to the middle term (or terms);
d) Sum of the exponents of y and z in n and the exponents of z is one less than the term's position number, that is, first term, second term, and so on;
e) Powers of y decrease while powers of z increase

Example 2.22
Determine the coefficients of x^8 in $\left(x^2+\dfrac{1}{x}\right)^{10}$.

Solution

$$\begin{aligned}\left(x^2+\frac{1}{x}\right)^{10} &= x^{20}\left[1+\left(\frac{1}{x}\right)^3\right]^{10} \\ &= x^{20}\left[1+\binom{10}{1}\left(\frac{1}{x}\right)^3+\binom{10}{2}\left(\frac{1}{x}\right)^6+\binom{10}{3}\left(\frac{1}{x}\right)^9\right. \\ &\qquad \left.+\binom{10}{4}\left(\frac{1}{x}\right)^{12}+\cdots\right]\end{aligned}$$

Therefore the coefficient is $\binom{10}{4}=210$.

It is usual, when expanding by the binomial theorem for large values of n, to leave the coefficients in the form $\binom{n}{r}$ or in factorial form, rather than evaluating them explicitly, since the numbers may become very large.

2.6.2 BINOMIAL THEOREM FOR RATIONAL INDEX

In the binomial theorem both n and r of the binomial coefficient

$$\binom{n}{r}=\frac{n!}{r!(n-r)!} \qquad ...(i)$$

are required to be nonnegative integers. But even though it does not make sense for negative integers of the expression on the right-hand of (i) the same cannot be said about the expression on the left-hand side.

Let α be any real number. We can still write

$$\binom{\alpha}{r}=\frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-r+1)}{r!}$$

where r is still a nonnegative integer. If r is not an integer we shall never use the symbol $\binom{\alpha}{r}$.

Some Identities of Binomial Coefficients

For any real number α and $r\geq 0$ an integer,

a) $\binom{\alpha}{0}=1$ (by convention)

b) $\binom{\alpha}{r}=0$ (if $\alpha<r$)

c) $\begin{pmatrix} \alpha \\ r \end{pmatrix} = (-1)^r \begin{pmatrix} -\alpha + r - 1 \\ r \end{pmatrix}$

d) $\begin{pmatrix} -\alpha \\ r \end{pmatrix} = (-1)^r \begin{pmatrix} \alpha + r - 1 \\ r \end{pmatrix}$

Example 2.23
Evaluate the following

$$\text{a)} \begin{pmatrix} -5 \\ 4 \end{pmatrix} \quad (b) \begin{pmatrix} -2 \\ 5 \end{pmatrix} \quad (c) \begin{pmatrix} 1/3 \\ 4 \end{pmatrix} \quad (d) \begin{pmatrix} -1/2 \\ 3 \end{pmatrix} \quad (e) \begin{pmatrix} 4 \\ 6 \end{pmatrix}$$

Solution

a) $\begin{pmatrix} -5 \\ 4 \end{pmatrix} = \dfrac{(-5)(-5-1)(-5-2)(-5-3)}{4!} = 70$

b) $\begin{pmatrix} -2 \\ 5 \end{pmatrix} = \dfrac{(-2)(-2-1)(-2-2)(-2-3)(-2-4)}{4!} = -6$

c) $\begin{pmatrix} 1/3 \\ 4 \end{pmatrix} = \dfrac{(1/3)(-1/3-1)(-1/3/2)(-1/3-3)}{4!} = \dfrac{55}{972}$

d) $\begin{pmatrix} -1/2 \\ 3 \end{pmatrix} = \dfrac{(1-1/2)(-1/2-1)(-1/2-2)}{3!} = -\dfrac{5}{16}$

e) $\begin{pmatrix} 4 \\ 6 \end{pmatrix} = 0, \text{ (since } 4 < 6)$

Theorem 2.12 GENERALIZED BINOMIAL THEOREM

If α is any real number, then

$$(1+x)^\alpha = \sum_{r=0}^{\infty} \begin{pmatrix} \alpha \\ r \end{pmatrix} x^a$$

for all $|x| < 1$

Expanding Theorem 2.12 we obtain

$$\begin{aligned} (1+x)^\alpha \quad &= \begin{pmatrix} \alpha \\ 0 \end{pmatrix} x^0 + \begin{pmatrix} \alpha \\ 1 \end{pmatrix} x^1 + \begin{pmatrix} \alpha \\ 2 \end{pmatrix} x^2 + \begin{pmatrix} \alpha \\ 3 \end{pmatrix} x^3 + \cdots \\ &\quad + \begin{pmatrix} \alpha \\ n \end{pmatrix} x^n + \cdots \\ &= 1 + \alpha x + \frac{\alpha(\alpha-1)}{21} x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{31} x^3 + \cdots \\ &\quad + \frac{\alpha(\alpha-1)(\alpha-2)(\alpha-n+1)}{n!} x^n + \cdots \end{aligned}$$

Theorem 2.12 is also called *Newton's Binomial Expansion* and the coefficient $\begin{pmatrix} \alpha \\ r \end{pmatrix}$ as the *Newton's binomial coefficient* or the *generalized binomial coefficient*.

Example 2.24

Use the generalized binomial theorem to expand $(1+x)^{\frac{1}{2}}$ up to the term in x^3 and determine $(1.08)^{\frac{1}{2}}$ up to the term in x^3 to 5 decimal places.

Solution

Here, $\alpha = \frac{1}{2}$ so that

$$\begin{aligned} (1+x)^{\frac{1}{2}} &= 1 + \frac{\frac{1}{2}}{1!}x + \frac{\frac{1}{2}(\frac{1}{2}-1)}{2!}x^2 + \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{3!}x^3 + \cdots \\ &= 1 + \frac{\frac{1}{2}}{1!}x + \frac{\frac{1}{2}(-\frac{1}{2})}{2!}x^2 + \frac{\frac{1}{2}(1-\frac{1}{2})(-\frac{3}{2})}{3!}x^3 + \cdots \\ &= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \cdots \end{aligned}$$

To obtain an approximate value for $(1.08)^{\frac{1}{2}}$ up to the term in x^3, we need to set $x = 0.08$ in the expansion for $(1+x)^{\frac{1}{2}}$. That is

$$\begin{aligned}(1+0.08)^{\frac{1}{2}} &= 1+\frac{1}{2}(0.08)-\frac{1}{8}(0.08)^2+\frac{1}{16}(0.08)^3\\ &= 1+0.04-0.0008+0.000032 = 1.03923 \quad \text{(5 decimal places)}\end{aligned}$$

Note that the expansion contains an infinite number of terms when α is not a positive integer, and the restrictions $|x|<1$ is necessary to ensure that the terms decrease in value sufficiently as quickly as the number of terms increases. Since it is not possible to write down all the terms in the expansion, it is usual to present the first few terms, the number of terms being chosen to give the required accuracy. If we substitute $(-x)$ for x and $(-\alpha)$ for α in Theorem 2.12, the result is the next theorem.

Theorem 2.13

For any real number α (but not a positive integer),

$$(1-x)^{-\alpha} = \sum_{r=0}^{\infty}\binom{-\alpha}{r}(-x)^r$$

for all $|x|<1$

Example 2.25
Expand $(1+x)^{-2}$ up to the term in x^4 and state its general form.

Solution

$$\begin{aligned}(1+x)^{-2} &= \binom{-2}{0}x^0+\binom{-2}{1}x^1+\binom{-2}{2}x^2+\binom{-2}{3}x^3+\cdots\\ &= 1+\frac{-2}{1!}x+\frac{(-2)(-2-1)}{2!}x^2+\frac{(-2)(-2-1)(-2-2)}{3!}x^3\\ &\quad +\frac{(-2)(-2-1)(-2-2)(-2-4)}{4!}x^4+\cdots\\ &\quad +(-1)^n\frac{\alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)}{n!}x^n+\cdots\\ &= 1-2x+\frac{6}{2!}x^2-\frac{24}{3!}x^3+\frac{120}{4!}x^4+\cdots\\ &= 1-2x+3x^2-4x^3+5x^4+\cdots\end{aligned}$$

Example 2.26

Use the generalized binomial theorem to expand $(1-x)^{-\frac{1}{3}}$ up to the term in x^3 and state its general terms.

Solution

$$\begin{aligned}(1-x)^{-\frac{1}{3}} &= 1+\frac{-\frac{1}{3}}{1!}(-x)+\frac{-\frac{1}{3}(-\frac{1}{3}-1)}{2!}(-x)^2+\frac{-\frac{1}{3}(-\frac{1}{3}-1)(-\frac{1}{3}-2)(-x)^3}{3!}+\cdots \\ &= 1+\frac{1}{3}x+\frac{1.4}{3^2}.\frac{x^2}{2!}+\frac{1.4.7}{3^3}.\frac{x^3}{31}+\cdots+\frac{1.4.7\cdots(3n-2)}{3^n}.\frac{x^n}{n!}+\cdots\end{aligned}$$

Expansion of $(y+z)^\alpha$

In order to expand $(y+z)^\alpha$ when α is a real number it is necessary to write

$$y+z=y\left(1+\frac{z}{y}\right)$$

So that

$$(y+z)^\alpha=y^\alpha(1+\frac{z}{y})^\alpha=y^\alpha(1-x)^\alpha$$

where $x=\dfrac{z}{y}$ with $|x|<1$

Example 2.27

Find the expansion of the expression $\dfrac{1}{(4-x)^{\frac{1}{2}}}$ up to the term in x^3.

Solution

We may write $4-x$ as $4-x=4\left(1-\dfrac{x}{4}\right)$

Let $y=-\dfrac{x}{4}$. Then $4-x=4(1+y)$

so that $(4-x)^{\frac{1}{2}}=(4)^{\frac{1}{2}}(1+y)^{\frac{1}{2}}$. Now

$$\frac{1}{(4-x)^{\frac{1}{2}}}=\frac{1}{2}(1+y)^{-\frac{1}{2}}$$

But

$$\begin{aligned}(1+y)^{\frac{1}{2}} &= 1+-\frac{\frac{1}{2}}{1!}y+\frac{-\frac{1}{2}(-\frac{3}{2})}{2!}y^2+\frac{-\frac{1}{2}(-\frac{3}{2})(-\frac{5}{2})}{3!}y^3+\cdots \\ &= 1-\frac{1}{2}y+\frac{3}{8}y^2-\frac{5}{16}y^3+\cdots,\quad |y|<1\end{aligned}$$

Replacing y by $-\dfrac{x}{4}$, we obtain

$$\begin{aligned}
\frac{1}{(4-x)^{\frac{1}{2}}} &= \frac{1}{2}\left(1-\frac{x}{4}\right)^{-\frac{1}{2}} \\
&= \frac{1}{2}\left(1-\frac{1}{8}x+\frac{3}{128}x^2-\frac{5}{1025}x^3+\cdots\right), \quad |x|<4
\end{aligned}$$

Expansion of $\left(y+\dfrac{z}{w}\right)^{\alpha}$

The following theorem provides a general formula for determining any term of the binomial expansion.

Theorem 2.14

For any real number α,

$$\left(y+\frac{z}{w}\right)^{\alpha} = \sum_{k=0}^{\alpha}\binom{a}{k} y^k \left(\frac{z}{w}\right)^{\alpha-k}$$

Example 2.28
Use Theorem 2.14 to expand $\left(x - \frac{1}{x^2}\right)^4$

Solution

$$\left(x - \frac{1}{x^2}\right)^4 = \left[x + \left(-\frac{1}{x^2}\right)\right]^4 = \sum_{k=0}^{4} \binom{4}{k} x^k \left(-\frac{1}{x^2}\right)^{4-k}$$

For $k = 0$, $\quad \frac{4!}{0!4!} x^0 \left(-\frac{1}{x^2}\right)^{4-0} = \frac{1}{x^8}$

For $k = 1$, $\quad \frac{4!}{1!3!} x^1 \left(-\frac{1}{x^2}\right)^{4-1} = -4\frac{1}{x^5}$

For $k = 2$, $\quad \frac{4!}{2!2!} x^2 \left(-\frac{1}{x^2}\right)^{4-2} = 6\frac{1}{x^2}$

For $k = 3$, $\quad \frac{4!}{3!1!} x^3 \left(-\frac{1}{x^2}\right)^{4-3} = 4x$

For $k = 4$, $\quad \frac{4!}{4!0!} x^4 \left(-\frac{1}{x^2}\right)^{4-4} = x^4$

Therefore

$$\left(x - \frac{1}{x^2}\right)^4 = x^{-8} - 4x^{-5} + 6x^{-2} - 4x + x^4$$

2.7 MULTINOMIAL THEOREM

The multinomial theorem is a generalization of the binomial theorem to polynomials with any number of terms.

Theorem 2.15 MULTINOMIAL THEOREM

Let n be any integer greater than 0. For all $x_1, x_2, \cdots, x_k \quad k \geq 1$

$$(x_1 + x_1 + \cdots + x_k)^n = \sum \binom{n}{n_1, n_2, \cdots, n_k} x_1^{n_1}, x_2^{n_2}, \cdots, x_k^{n_k}$$

where the sum is over all possible integer values of $n_1, n_2, \cdots, n_k$ such that the constraint $n_1 + n_2 + \cdots + n_k = n$ is satisfied.

The multinomial formula may be written as

$$(x_1 + x_1 + \cdots + x_k)^n = \sum \binom{n}{n_1, n_2, \cdots, n_k} \prod_{i=1}^{k} x_i^{n_i}$$

where

$$\binom{n}{n_1, n_2, \cdots, n_k}$$

is the multinomial coefficient.

Some Identities of Multinomial Coefficients

a) The multinomial coefficient can be expressed

   i) in factorials:

$$\binom{n}{n_1, n_2, \cdots, n_k} = \frac{n!}{n_1, n_2, \cdots, n_k}$$

   ii) as a product of binomial coefficients:

$$\binom{n}{n_1, n_2, \cdots, n_k} = \binom{n_1}{n_1}\binom{n_1+n_2}{n_2}\cdots\binom{n_1+n_2+\cdots+n_k}{n_k}$$

b) For $k = 2$,

$$\binom{n}{n_1, n_2} = \binom{n}{n_1} = \binom{n}{n_2}$$

c) For $k = n$,

$$\binom{n}{1, 1, \cdots, 1} = n!$$

d) The number of terms in the multinomial expansion is $\left(k+n-1 \right)$

e) $\displaystyle\sum_{r=1}^{n} \binom{n}{n_1, n_2, \cdots, n_k} = k^n$

Example 2.29

Use the multinomial theorem to expand $(x+y+z)^2$

Solution

We expect to have $\binom{3+2-1}{2} = 6$ terms of the form $x^{n_1},\ y^{n_2},\ z^{n_3}$, where $n_1+n_2+n_3=2$. Therefore

$$
\begin{aligned}
(x+y+z)^2 &= \frac{2!}{2!0!0!}x^2y^0z^0 + \frac{2!}{1!1!0!}x^1y^1z^0 + \frac{2!}{1!0!1!}x^1y^0z^1 + \\
&\quad \frac{2!}{0!1!1!}x^0y^1z^1 + \frac{2!}{0!\ 2!0!}x^0y^2z^0 + \frac{2!}{0!0!2!}x^0y^0z^2 \\
&= x^2y^0z^0 + 2x^1y^1z^0 + 2x^1y^0z^1 + 2x^0y^1z^1 \\
&\quad + x^0y^2z^0 + x^0y^0z^2 \\
&= x^2 + 2xy + 2xz + 2yz + y^2 + z^2
\end{aligned}
$$

The multinomial coefficients have a direct combinatorial interpretation, as

a) the number of n-letter words formed with r distinct letters used $n_1, n_2, \cdots, n_r$ times;
b) the number of ways of distributing n distinct objects into k distinct boxes, with n_1 objects in the first box, n_2 objects in the second box, and so on.
c) the number of ways to split n distinct objects into r distinct groups, of sizes $n_1, n_2, \cdots, n_r$, respectively.

We have now gone through some basic mathematics that we obviously require to understand and solve probability problems. If we have understood the materials in this part we are then ready for the rest of the book, especially Parts 2 and 3.

EXERCISES

2.1 In Ghana, vehicle license numbers follow the system of two letters followed by four digits and then by a letter. How many license plate are possible.

2.2 How many six-digit telephone numbers are possible if the first digit cannot be a 0.

2.3 Without using a calculator, find the values of

a) ${}_4P_1$ (b) ${}_8P_5$ (c) ${}_8P_3$ (d) ${}_{10}P_{10}$

2.4 How many different three-digit numbers can be formed using $0, 1, 2, 3, 4$ (excluding numbers which begin with 0) if

a) no digits are to be repeated;
b) repetitions are allowed

2.5 Find the number of ways of arranging 6 different books on a shelf

2.6 Refer to Exercise 2.5. Suppose the books consist of two Mathematics books, three Statistics books, and 1 Physics book. In how many ways can these books be arranged on a shelf it
a) the different books on a subject are not distinguished?
b) two particular statistics books will be together?
c) books on the same subject are placed together?

2.7 Prove the permutation identities in Section 2.4.

2.8 A lady has 8 house plants. In how many ways can she arrange 6 of them on a line a window still?

2.9 The Management Committee of a certain football team consists of nine members. If 5 are to fill the positions of the president, vice-president, secretary, treasurer and publicity officer, how many different slates of officers are possible?

2.10 There are 25 entrants in a gymnastics competition. In how many different way can the gold, silver and bronze medals be won?

2.11 In how many ways could we arrange four guests, in a row of four numbered chairs at a concert. Show all the arrangements.

2.12 In how many ways can 3 boys and 3 girls sit in a row
a) all of them can sit anywhere?
b) the boys and the girls are each to sit together?
c) only the boys must sit together?
d) no two people of the same sex are allowed to sit together?

2.13 Refer to Exercise 2.11. Suppose two of the guests are not on taking terms, what different arrangements are possible with the two
a) sitting together?
b) not sitting together?

2.14 In how many ways could we sit four couples in a row if
a) no husband should sit by the wife?
b) a husband should sit by the wife?

2.15 The Managing Director of a reputable company wants to fill four vacant positions in four regional capitals, namely, Kumasi, Tamale, Koforidua and Ho. There is a pool of ten officers at the headquarters in Accra from which to fill these positions. In how many different ways can this be done?

2.16 Find the number of different ways in which the letters of the following words can be arranged:
(a) $N\ U\ M\ B\ E\ R$; (b) $P\ O\ S\ S\ I\ B\ L\ E$;
(c) $P\ E\ P\ P\ E\ R$; (d) $S\ T\ A\ T\ I\ S\ T\ I\ C\ S$.

2.17 Refer to Exercise 2.16 (d). Rework it if all the three T's should be together.

2.18 A child has 12 blocks of which 6 are black, 4 are red, 1 is white and 1 is blue. If the child puts blocks in a line, how many arrangement are possible?

2.19 Six people, A, B, C, D, and F are to be seated round a circular table.
a) How many ways are there of achieving this?
b) How many ways are there of achieving this if $\mathcal{A}$ refuses to st beside B?

2.20 In how many ways could we arrange four couples in a circle if
a) each person can be seated anywhere?
b) a husband should sit by the wife?
c) no husband should sit by the wife?
d) the men and women alternate?
e) husbands and wives alternate?

2.21 Evaluate

$$(a)\ \binom{7}{4} \quad (b)\ \binom{5}{4} \quad (c)\ \binom{6}{1} \quad (d)\ \binom{8}{0} \quad (e)\ \binom{8}{6}$$

$$(f)\ \binom{9}{4} \quad (g)\ \binom{50}{4} \quad (h)\ \binom{9}{0} \quad (i)\ \binom{50}{46} \quad (j)\ \binom{4}{5}$$

2.22 The Reverend Minister of church always insists that every member shakes hands with every other member exactly once. On one particular Sunday, ninety members were present in the church. How many handshakes occurred.

2.23 In how many ways can
a) 3 representatives be chosen from 20 students?
b) 11 players be selected from 12 footballers?

2.24 A Committee of 6 is to be formed from1 3 men and 7 women. In how many ways can the Committee be selected given that it must have

a) 4 men and 2 women?

b) at least one member of each sex?

2.25 The Managing Director of a company wishes to form a Committee consisting of 4 members from a pool of 10 Executives. How many different Committees can be formed?

2.26 16 people, 4 from each of 4 groups $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$, have to select 6 of their members to represent them on Committee. How many selections can be made if

a) each group must be represented?

b) no group can be more than two representatives?

2.27 The Managing Director of a company decides on one Monday to accompany one of his ten Executives on each of the remaining four days of the week to negotiate contracts. To decide whom to accompany, he writes the name of each of the Executives on a separate card. He puts all the cards in a small container, mixes then up and draws one. In how many different sequence are there when by a sequence we mean a list of four names, the first being the name of the individual he accompanies in Tuesday, the second the individual he accompanies on Wednesday, and so on?

2.28 There are four female and eight male senior officers of an establishment. The Chief Executive plans to form a Committee on which there are two females and four males. How many different Committees can be formed for this compositions?

2.29 How many different ways are there to place 10 indistinguishable balls in four boxes?

2.30 Roll five dices once. How many different outcomes are there if the dice are (a) distinguishable (b) indistinguishable?

2.31 Construct the row of the Pascal triangle corresponding to

(a) $n = 7$, (b) $n = 8$

2.32 Use the binomial theorem to expand

(a) $(1+x)^6$ (b) $(1+2x)^6$

(c) $(1-x)^6$ (d) $(2-x)^6$

(e) Use the first three terms of (d) to estimate $(1.95)^6$.

2.33 Use the binomial theorem to expand

(a) $(x+y)^4$, (b) $(x-y)^3$
(c) $(x+3y)^4$, (d) $(3x+2y)^3$

2.34 Use the binomial theorem to expand

(a) $\left(x+\dfrac{1}{x}\right)^4$ (b) $\left(x-\dfrac{1}{x^2}\right)^4$ (c) $\left(2x^2-\dfrac{3}{x}\right)$

2.35 Determine the coefficient of

(a) x^4 in $(1-x)^7$, (b) x^5 in $(1-x)^7$
(c) x^9 in $(2-x)^{12}$, (d) x^3 in $(1+2x)^{20}$
(e) x^0 in $\left(\dfrac{1}{x}-2x\right)^8$ (f) x^{-3} in $\left(\dfrac{3}{x^3}-\dfrac{x^2}{2}\right)$

2.36 Without expanding, find the term of $(2x+3y)^6$ involving x^2y^4.

2.37 Find the term involving y^9 in the expansion of $(x+y^3)^7$.

2.38 Use the binomial theorem to find the first four terms, and give the range of x for which the full expansion is valid, for:

(a) $(1+x)^{\frac{1}{3}}$, (b) $(1-x)^{\frac{1}{3}}$, (c) $(1+2x)^5$
(d) $(1+2x)^{-4}$, (e) $(4-x)^{\frac{1}{2}}$, (f) $(4-x)^{-5}$
(e) $(1+x)^{-1}$, (h) $(1-x)^{-2}$, (i) $(1+x)^{-\frac{1}{4}}$

2.39 Use the expansion obtained in (e) in Exercise 2.38 above to evaluate $(3.8)^{\frac{1}{2}}$ to four places of decimals.

2.40 Use the binomial theorem to evaluate $(1.2)^4$.

2.41 Shoe that $\begin{pmatrix} -\alpha \\ r \end{pmatrix} = (-1)^r \begin{pmatrix} \alpha+r-1 \\ \alpha-1 \end{pmatrix}$

2.42 Use Theorem 2.14 to rework part (a) and (c) of Exercise 2.34.

2.43 Use the multinomial theorem to expand the following

(a) $(x+y+z)^3$ (b) $(x-y+1)^2$

3 BASIC CONCEPTS IN PROBABILITY

3.1 INTRODUCTION

In this chapter, we shall introduce certain terms, which are used in a technical sense in probability problems. These include the concepts of "experiment", "outcome", "sample space" and "event". It is important that their meanings be well understood. Many of them are used in a nontechnical sense in everyday language and we are, therefore, cautioned to master their technical meanings and put aside their ordinary usage. We would then be equipped in developing an understanding of the more important concepts in probability in the last part of this chapter and in the subsequent ones.

3.2 EXPERIMENTS

3.2.1 DEFINITION OF CONCEPTS

Uncertainty refers to the outcome of some process of change. If a process of change can lead to two or more possible results, the results are said to be uncertain.

Definition 3.1 EXPERIMENT

An experiment $\mathcal{E}$ is an act or a process specially set up or occurring naturally which leads to some well-defined outcomes or results

Example 3.1
If a coin is tossed[7] we observe whether a head or a tail is obtained. We might describe this as the experiment of tossing a coin. Another example of an experiment is rolling a die.

Definition 3.2 TRIAL

Each repetition of an experiment is a trial

Example 3.2
If a coin is tossed four times, each single toss is a trial.

Definition 3.3 OUTCOME OF EXPERIMENT

An outcome of an experiment is the result when the experiment is performed once

Example 3.3
A die is rolled once and a 5 showed on the face. The outcome is the number 5.

Definition 3.4 EQUALLY LIKELY OUTCOMES

When any one outcome of an experiment has the same chance of occurrence as any other outcome when the experiment is performed, then outcomes are said to be equally likely

Example 3.4
When a die is tossed once the outcomes 1, 2, 3, 4, 5, 6 are all equally likely as long as the die is fair.

Definition 3.5 SIMPLE OUTCOME

A simple outcome is an outcome which cannot be further decomposed

Synonym for a simple outcome is an ***elementary outcome***.

Example 3.5
If a die is thrown once there are six possible outcomes: 1, 2, 3, 4, 5, 6. Each of these outcomes cannot be further subdivided. Hence each outcome is simple.

Example 3.6
Consider an experiment $\mathcal{E}$ of tossing a coin twice and observing the sequence of heads $\mathcal{H}$ and tails $\mathcal{T}$. The four possible simple outcomes are

$$\mathcal{TT} \quad \mathcal{TH} \quad \mathcal{HT} \quad \mathcal{HH}$$

where $\mathcal{TH}$ for example, indicates a tail on the first throw and a head on the second throw.

Note

The experiment of tossing a coin twice is the same as tossing two identical coins once.

Definition 3.6 COMPOSITE OUTCOMES

A composite outcome is an outcome which can be further broken down into simple outcomes

A synonym for a composite outcome is a ***compound outcome***.

Example 3.7
Refer to Example 3.5. There can be two possible outcomes: odd numbers 1, 3, 5 or even numbers 2, 4, 6. Each of these outcomes can further be subdivided into simple outcomes 1, 3 and 5 and 2, 4 and 6. Hence, the two outcomes "odd" and "even" are composite outcomes.

3.2.2 TYPES OF EXPERIMENTS

There are, essentially, two types of experiments:

a) Deterministic experiments,
b) Random experiments.

Definition 3.7 DETERMINISTIC EXPERIMENT

An experiment is deterministic if its observed result is not subject to chance.

In a deterministic experiments, if the experiment is repeated a number of times under exactly the same conditions, we expect the same result.

Example 3.8
If we measure the distance in, say, kilometres between town A and town B many times under the same conditions, we expect to have the same result.

Definitions 3.8 RANDOM EXPERIMENT
An experiment is random if its outcomes are uncertain

A synonym for a random experiment is a *stochastic experiment*.

If a random experiment is repeated under identical conditions, the outcomes may be different as there may be some random phenomena or chance mechanism at work affecting the outcome.

Example 3.9
Tossing a coin or rolling a die is a random experiment since in each case, the process can lead to more than one possible outcome. In the case of the coin tossing experiment, the result will come up tail ($\mathcal{T}$) or head ($\mathcal{H}$); and in the case of the die, the result will be one of the numbers 1, 2, 3, 4, 5 and 6.

Example 3.10
Picking a ball from a box containing 50 numbered balls is a random experiment since the process can lead to one of the many possible outcomes, that is, any of the 50 balls may be chosen.

3.2.3 PROPERTIES OF RANDOM EXPERIMENTS

Random experiments have the following three properties:

a) The experiment should be repeatable under identical conditions;
b) The outcome on any particular trial is variable, dependent on some chance or random mechanism;
c) If the experiment is repeated a large number of times, then some statistical regularity becomes apparent in the outcomes obtained. This means that, as an experiment is repeated a large number of times, the proportion of trials that a particular event $\mathcal{A}$ is observed tends to some limit.

3.3 SAMPLE SPACES

3.3.1 DEFINITION OF SAMPLE SPACES

Definition 3.9 SAMPLE SPACE

A sample space $\mathcal{S}$ is the set of all possible outcomes of some given random experiment $\mathcal{E}$

Note

A sample space is synonymous to *outcome set* or an *outcome space*.

The following three examples are experiments related to the coin, the die and cards. For each of them we shall define the sample space. Such and similar experiments are performed even more often in probability textbooks than they are in real life.

Example 3.11
A coin is tossed once. There are only two possible outcomes. Either it falls Head, $\mathcal{H}$ or Tail, $\mathcal{T}$, so the sample space is

$$\mathcal{S} = \{\mathcal{H}, \mathcal{T}\}$$

Example 3.12
A die is rolled once. There are six faces of a die so the sample space is

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

Example 3.13
A deck of cards has fifty-two cards. There are four suits: clubs ($\clubsuit$), hearts ($\heartsuit$), diamond ($\diamondsuit$) and spade ($\spadesuit$). Hearts and diamond are red; clubs and spade are black. Each suit has thirteen cards. The picture cards for each of the four suits are the Jack ($\mathcal{J}$), the Queen ($\mathcal{Q}$), and the King ($\mathcal{K}$). Hence there are 12 picture cards in the deck of 52 cards. In addition, each of the suits has an Ace ($\mathcal{A}$) so there are 4 Aces in the deck.

An experiment may have several sample spaces depending on the problem of interest. This is illustrate in the following examples.

Example 3.14
Toss a coin three times and observe the sequence of heads and tails.

Solution
Let $\mathcal{H}$ be Head and $\mathcal{T}$ be Tail. Then the sample space $\mathcal{S}$ would be

$$\{\mathcal{HHH},\quad \mathcal{HHT},\quad \mathcal{TTT},\quad \mathcal{TTH},\quad \mathcal{THH},\quad \mathcal{HTH},\quad \mathcal{THT},\quad \mathcal{HTT}\}$$

where $\mathcal{THT}$, for example, indicates a tail on the first throw, a head on the second throw and a tail on the third throw.

Example 3.15
Toss a coin three times and observe the total number of heads that occur.

Solution
When a coin is tossed three times, it is likely none of them will show a head ($\mathcal{TTT}$) or only one coin will show a head (either $\mathcal{HTT}$ or $\mathcal{THT}$ or $\mathcal{TTH}$) or all the three will show a head ($\mathcal{HHH}$). Hence the sample space $\mathcal{S}$ in this case is

$$\mathcal{S} = \{0,\ 1,\ 2,\ 3\}$$

Definitions 3.10 SAMPLE POINT

A particular outcome in the sample space $\mathcal{S}$ is called a sample point

The number of sample points in $\mathcal{S}$ may be denoted as $n(\mathcal{S})$ and is called the ***size of*** $\mathcal{S}$ or ***cardinality of*** $\mathcal{S}$. In Example 3.12, there are six sample points in the sample space, hence $n(\mathcal{S}) = 6$.

Example 3.16
Throw two dice once[8] and observe the numbers that appear on their faces.

a) Describe a suitable sample space.
b) How many sample points are there?

Solution
Let the pair of integers (i, j) $(1 \leq i, j \leq 6)$ represent the numbers that appear on Die 1 and Die 2 respectively, where the dice are numbered arbitrary. For example, the pair (2, 3) means that the number 2 appear on Die 1 and 3 on Die 2. In order that we may not miss any pair, it is always advisable to fix the number that appears on the face of one of the dice while varying the number on the other die.

a) All the possible pairs are presented in the table below.

	Die 2					
Die 1	1,1	1,2	1,3	1,4	1,5	1,6
	2,1	2,2	2,3	2,4	2,5	2,6
	3,1	3,2	3,3	3,4	3,5	3,6
	4,1	4,2	4,3	4,4	4,5	4,6
	5,1	5,2	5,3	5,4	5,5	5,6
	6,1	6,2	6,3	6,4	6,5	6,6

b) There are 36 sample points, that is, $n(S) = 36$.

3.3.2 KINDS OF SAMPLE SPACE

If we classify sample spaces according to the number of points that they contain, then there are three distinct kinds:

a) finite sample space,
b) countably or denumerably infinite sample space,
c) uncountable or nondenumerably infinite sample space.

Definition 3.11 FINITE SAMPLE SPACE

If a sample space has a finite number of points, it is called a finite sample space

Example 3.17
Toss a coin twice and count the number of heads. The sample space which is $\mathcal{S} = \{0, 1, 2\}$ is finite.

Definitions 3.12 COUNTABLY INFINITE SAMPLE SPACE

If a sample space has as many point as there are natural numbers $1, 2, 3, \cdots$, then it is called a countably infinite sample space

A synonym for countably infinite sample space is *denumerable sample space*.

Example 3.18
Toss a coin until a head appears and then count the number of times the coin was tossed. A coin may be tossed once and a head appears, or it may require two, three, …, fiftieth, …, and for all we know it may require thousands of tosses before a head appears. Not knowing how many times we may have to toss the coin, it is appropriate in an example like this to take as the sample space the whole set of natural numbers, of which there is a countable infinity before a head appears, and so on. Thus, the sample space S is

$$S = \{1,\, 2,\, 3, \cdots\}$$

which is the set of all natural numbers and is countably infinite.

Definitions 3.13 UNCOUNTABLY INFINITE SAMPLE SPACE

If a sample has as many points as there are in some interval on the x-axis, it is called an uncountably infinite sample space

A synonym for uncountably infinite sample space is *nondenumerable sample space*.

Example 3.19
The interval

$$[0, 1] = \{0 \leq x \leq 1\}$$

is an uncountably infinite sample space.

Sample spaces may also be distinguished according to whether they are *discrete* or *continuous*.

Definition 3.14 DISCRETE SAMPLE SPACE

A sample space which is finite or countably infinite is often called a discrete sample space.

Elements in a discrete sample space can be separated and counted. It can be finite or countably infinite

Example 3.20
The number of students in a school can be considered a finite outcome of a discrete space. The first success in trials (Example 3.18) can be a countably infinite outcome of discrete sample space.

Definition 3.15 CONTINUOUS SAMPLE SPACE

A sample space which is uncountably infinite is called a nondiscrete or continuous sample space.

Continuous sample spaces have elements that cannot be separated and counted. They are required whenever the outcomes of experiments are measurements of physical properties such as length, temperature, time, weight, mass, etc, which are measured on a continuous scale. The number of sample points is always uncountably infinite.

Example 3.21
The time it takes for light-bulb to burn out. The outcome of this experiment could be any real number from zero to a certain upper limit (such as 100,000 hours or more).

3.4 EVENTS

3.4.1 DEFINITION OF CONCEPTS

The term "event" is vital in understanding the basics of probability and it is important that after understanding the concept of "sample space", we grasp this concept.

Definition 3.16 EVENTS

An event is a subset of the sample $\mathcal{S}$

We denote an event by a capital letter, such as, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, etc.

Example 3.22
Roll a die once. Write down the following events as sets:

a) the number 4;
b) a number greater than 4;
c) an odd number.

Solution
Let $\mathcal{A}$ represent an event. Then

a) The event of rolling a 4 can be satisfied by only one outcome, the 4 itself, hence

$$\mathcal{A} = \{4\}$$

b) The event of rolling a number greater than 4 can be satisfied by any one of two outcomes: the number 5 or 6, hence

$$\mathcal{A} = \{5, 6\}$$

c) The event of rolling an odd number can be satisfied by any one of three outcomes: the number $1, 3,$ or 5 hence

$$\mathcal{A} = \{1, 3, 5\}$$

Note

a) An event may be defined also as a subcollection of the outcomes of an experiment.
b) If the outcome ω of an experiment is an element of event $\mathcal{A}$, that is, $\omega \in \mathcal{A}$, then the event $\mathcal{A}$ is said to occur. It is important to remember this idea. The set of outcomes not in $\mathcal{A}$ is called the complement (or *negation*) of $\mathcal{A}$, and is denoted by $\overline{\mathcal{A}}$ or $\mathcal{A}'$ or A^c (see Chapter 1).
c) If a sample space contains n sample points, then there are a total of 2^n subsets of event (see Chapter 1).

The empty set $\emptyset$ and the sample space $\mathcal{S}$ itself are particular events. The sample space $\mathcal{S}$ is called the "sure" or "certain" or "definite" event since an element of $\mathcal{S}$ is bound to occur in one trial of the experiment. An event containing no outcome is an "impossible" event and is denoted as $\emptyset$. An element of $\emptyset$ cannot occur in any of the trials of the experiment. Obviously, therefore, $\emptyset = \mathcal{S}^c$.

Definitions 3.17 SIMPLE EVENT

An event consisting of a single outcome is called a simple event

Synonyms for simple event are ***elementary event*** of ***fundamental event***. The letter e with a subscript will be used to denote a simple event or the corresponding sample point.

Example 3.23
Toss a die once. The individual event

$$\begin{array}{llllll} e_1: & \text{observe a 1;} & e_2: & \text{observe a 2;} & e_3: & \text{observe a 3} \\ e_4: & \text{observe a 4;} & e_5: & \text{observe a 5;} & e_6: & \text{observe a 6} \end{array}$$

are each a simple event.

Definitions 3.18 STRING EVENT

An event which involves a sequence of elementary events is called a string event

Example 3.24
Toss a coin twice and obtain, say, 1 head and 1 tail. This involves a sequence of two elementary events of tossing a head and a tail.

Note

In a string event, the order is of crucial importance. Thus, the events $\mathcal{HT}$ is not the same as $\mathcal{TH}$ when a coin is tossed twice.

> **Definitions 3.19** COMPOSITE EVENTS
>
> An event consisting of more than one outcome is called a composite event

Synonyms for a composite event is a ***compound event*** or ***compound outcome*** or simply an ***event***.

Example 3.25
Throw a balance die once. Is the event $\mathcal{A} = \{\text{number} \leq 2\}$ an example of a composite event? Explain.

Solution
Let e_i denote a simple event i $(i = 1, 2, \cdots, 6)$. Then

$$\mathcal{A} = \{e_2, e_3, e_4, e_5, e_6\}$$

is a composite event. This is because the set $\mathcal{A}$ can be decomposed further into the following simple events: e_2, e_3, e_4, e_5, e_6.

3.4.2 OPERATIONS ON EVENTS

Since events are sets, it is clear that statements concerning events can be translated into the language of set theory and conversely. In particular, we have an algebra for events corresponding to the algebra of sets. We can, therefore, combine events using the various set operations. Thus if $\mathcal{A}$ and $\mathcal{B}$ are events in a sample space $\mathcal{S}$, then

a) $\mathcal{A} \subset \mathcal{B}$ denotes outcomes in "$\mathcal{A}$ implies $\mathcal{B}$";
b) $\mathcal{A} \cup \mathcal{B}$ denotes outcomes in "either $\mathcal{A}$ or $\mathcal{B}$ (or both)", their *union*;
c) $\mathcal{A} \cap \mathcal{B}$ denotes outcomes in "both $\mathcal{A}$ and $\mathcal{B}$", their *intersection*;
d) $\mathcal{A} \Delta \mathcal{B}$ denotes outcomes in "either $\mathcal{A}$ or $\mathcal{B}$, but not both", their *symmetric difference*;
e) $\overline{\mathcal{A}}$ is the event "not $\mathcal{A}$" (that is, event $\mathcal{A}$ does not occur), its *complement*;
f) $\mathcal{A}/\mathcal{B}$ denotes outcomes in "$\mathcal{A}$ but not $\mathcal{B}$", their *difference*.

Example 3.26
A die is rolled once and the numbers on its face recorded.

a) List the elements of the following events:

$\mathcal{A} = \{\text{even number occurs}\};$
$\mathcal{B} = \{\text{odd number occurs}\};$
$\mathcal{C} = \{\text{prime number occurs}\};$
$\mathcal{D} = \{\text{positive integer between 0 and 7 occurs}\};$
$\mathcal{E} = \{\text{7 occur}\}.$

b) Describe the events

$$\mathcal{A} \cup \mathcal{B}, \quad \overline{\mathcal{A}}, \quad \mathcal{A} \cup \mathcal{C}, \quad \mathcal{A} \cap \mathcal{B}, \quad \mathcal{A}/\mathcal{C}, \quad \mathcal{A} \Delta \mathcal{C}$$

Solution
The sample space consists of the six possible numbers

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

a) The events $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{E}$ are as follows:

$$\mathcal{A} = \{2, 4, 6\}, \quad \mathcal{B} = \{1, 3, 5\}, \quad \mathcal{C} = \{2, 3, 5\}, \quad \mathcal{D} = \mathcal{S}, \quad \mathcal{E} = \emptyset$$

b) The set $\mathcal{A} \cup \mathcal{B}$ is the event that either even number occurs or odd number occurs:

$$\mathcal{A} \cup \mathcal{B} = \{2, 4, 6, 1, 3, 5\} = \mathcal{S}$$

The set $\overline{\mathcal{A}}$ is the event that an even number does not occur:

$$\overline{\mathcal{A}} = S \backslash A = \{1,\ 3,\ 5\}$$

The set $\mathcal{A} \cap \mathcal{C}$ is the event that the numbers that occur are even as well as prime numbers:

$$\mathcal{A} \cap \mathcal{C} = \{2\}$$

The set $\mathcal{A} \cap \mathcal{B}$ is the event that the numbers that occur are even as well as odd numbers:

$$\mathcal{A} \cap \mathcal{B} = \emptyset$$

The set $\mathcal{A} \backslash \mathcal{C}$ is the event that the numbers that occur are either even and not prime numbers:

$$\mathcal{A} \backslash \mathcal{C} = \{4,\ 6\}$$

The set $\mathcal{A} \Delta \mathcal{C}$ is the event that the numbers that occur are either even numbers or prime numbers but not both:

$$\mathcal{A} \Delta \mathcal{C} = \{3, 4, 5, 6\}$$

Note

If $\mathcal{A}_1$, $\mathcal{A}_2, \cdots, \mathcal{A}_n$ is a sequence of sets, then:

$$\bigcup_{i=1}^{n} \mathcal{A}_\rangle = \text{the event that at least one of the events } A_i \text{ occurred;}$$

$$\bigcap_{i=1}^{n} \mathcal{A}_\rangle = \text{the event that all the events } A_i \text{ occurred.}$$

3.4.3 MUTUALLY EXCLUSIVE AND COLLECTIVELY EXHAUSTIVE EVENTS

> **Definitions 3.20** MUTUALLY EXCLUSIVE EVENTS
>
> Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be mutually exclusive if they cannot occur together

That is, two events, $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive if the occurrence of $\mathcal{A}$ implies the non-occurrence of $\mathcal{B}$ and vice versa. Synonyms for mutually exclusive events are ***disjoint events***, ***incompatible events*** or ***non-overlapping events***.

$\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive if and only if $\mathcal{A} \cap \mathcal{B} = \emptyset$.

Example 3.27
When a die is rolled once, the numbers 4 and 5 cannot occur together and hence the event $\mathcal{A} = \{4\}$ and $\mathcal{B} = \{5\}$ are mutually exclusive.

Definitions 3.21 MUTUALLY INCLUSIVE EVENTS

Two events $\mathcal{A}$ and $\mathcal{B}$ are called mutually inclusive if they can occur together

A synonym of mutually inclusive events is ***overlapping events*** or ***compatible events***.

$\mathcal{A}$ an $\mathcal{B}$ are mutually inclusive if and only if $\mathcal{A} \cap \mathcal{B} \neq \emptyset$.

Example 3.28
Suppose

$$S = \{x|x \ \in N \ \text{ and } x < 8\}$$

Which of the following events are mutually inclusive:

a) $\mathcal{A} = \{1, 2, 3\}$ and $\mathcal{B} = \{4, 5, 6\}$
b) $\mathcal{C} = \{x|x$ is an even number, $3 < x < 7\}$ and $\mathcal{D} = \{4, 6, 7\}$

Solution
a) $\mathcal{A}$ and $\mathcal{B}$ are not mutually inclusive (they are mutually exclusive) since they cannot occur together.
b) $\mathcal{C}$ and $\mathcal{D}$ are mutually inclusive since they overlap:

$$\mathcal{C} \cap \mathcal{D} = \{4, 6\} \neq \emptyset$$

Definition 3.22 COLLECTIVELY EXHAUSTIVE EVENTS

Two or more events defined on the same sample space are said to be collectively exhaustive if their union is equal to the sample space $\mathcal{S}$

In order words, n events are said to be collectively exhaustive if their union equals the sample space, that is

$$\mathcal{A}_1 \cup A_2 \cup \cdots \cup A_n = S$$

Example 3.29
When a die is thrown once, the events $1, 2, 3, 4, 5, 6$ are collectively exhaustive because their union equals the sample space.

Note

a) If two or more events are mutually exclusive, they cannot occur together, so that at most one of the events will occur.
b) If two or more events are collectively exhaustive, then at least one of them will occur.
c) If a set of events is mutually exclusive and collectively exhaustive, then exactly one of the events will occur.
d) For any two events, $\mathcal{A}$ and $\mathcal{B}$, defined over the same sample space $\mathcal{S}$,

i) $(\mathcal{A} \cap \mathcal{B})$ and $(\mathcal{A} \cap \overline{\mathcal{B}})$,
ii) $(\overline{\mathcal{A}} \cap \mathcal{B})$ and $(\mathcal{A} \cap \mathcal{B})$,
iii) $(\overline{\mathcal{A}} \cap \mathcal{B})$ and $(\mathcal{A} \cap \overline{\mathcal{B}})$,

are mutually exclusive and collectively exhaustive

Definition 3.23 PARTITION OF SAMPLE SPACE

The events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ form a partition of the same sample space $\mathcal{S}$, if

a) $\mathcal{A}_i \neq \emptyset$ for all $i = 1, 2, \cdots n$
b) $\mathcal{A}_1 \cap \mathcal{A}_j = \emptyset$ for all $i \neq j; \quad i, j = 1, 2, \cdots n$
c) $\bigcap_{i=1}^{n} A_i = \mathcal{S}$

In other words, the n events $\mathcal{A}_1, A_2, \cdots, A_n$ form a partition of the sample space $\mathcal{S}$ if the n events are (a) nonempty, (b) mutually exclusive and (c) collective exhaustive. Fig. 3.1 is an example of a partition.

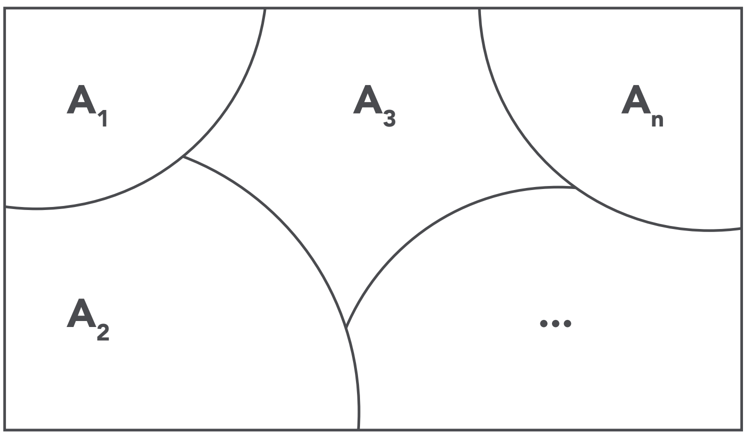


Fig. 3.1 Partition of Sample Space

Example 3.30

A coin is tossed three times. Partition the sample space according to the number of heads in the outcome.

Solution

The sample space is

$$\mathcal{S} = \{\mathcal{TTT},\ \mathcal{TTH},\ \mathcal{THT},\ \mathcal{THH},\ \mathcal{HTH},\ \mathcal{HHT},\ \mathcal{HHH}\}$$

The partitions are

$$\begin{aligned} \mathcal{A}_1 &= \{\mathcal{HHH}\} \\ \mathcal{A}_2 &= \{\mathcal{HHT},\ \mathcal{HTH},\ \mathcal{THH}\} \\ \mathcal{A}_3 &= \{\mathcal{TTH},\ \mathcal{THT},\ \mathcal{HTT}\} \\ \mathcal{A}_4 &= \{\mathcal{TTT}\} \end{aligned}$$

These subsets satisfy the definition of a partition, namely, no nonempty event, no overlapping events and all the possible events add up to the sample space $\mathcal{S}$.

3.4.4 INDEPENDENT EVENTS

Definition 3.24 INDEPENDENT EVENTS

Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if the occurrence (or non-occurrence) of one of them is not affected by the occurrences (or non-occurrence) of the other

Example 3.31

Toss two coins. The events "Head" on the first coin and "Tail" on the second coin are independent.

Definition 3.25 DEPENDENT EVENTS

Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be dependent if the occurrence (or non-occurrence) of one is affected by the occurrence (or non-occurrence) of the other

Example 3.32

A box contains two red pens and three blue pens. Two pens are picked at random successively. The events "blue pen" in the second picking and the "red pen picked in the first round" are dependent.

Properties of Independent Events

Property 1

In a collection of independent events, any of the events may be replaced by its complement and the resulting collection of events is still independent[9].

That is, if $\mathcal{A}$ and $\mathcal{B}$ defined over the same sample space $\mathcal{S}$ are independent events, then

a) $\mathcal{A}$ and $\overline{\mathcal{B}}$ are independent;
b) $\overline{\mathcal{A}}$ and $\mathcal{B}$ are independent;
c) $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$ are independent.

Property 2

If $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ are independent events, then

a) $\mathcal{C}$ and $\mathcal{A} \cup \mathcal{B}$ are independent
b) $\mathcal{C}$ and $\mathcal{A} \cap \mathcal{B}$ are independent
c) $\mathcal{C}$ and $\mathcal{A} \backslash \mathcal{B}$ are independent

That is, if $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ are independent, then $\mathcal{C}$ will be independent of any event that can be formed from $\mathcal{A}$ and $\mathcal{B}$ using set union, intersection and complementation.

3.5 CONCEPT OF PROBABILITY

3.5.1 WHAT IS PROBABILITY?

The word "probability" is frequently used in everyday speech. We say, for instance, "It will probably rain tomorrow" or "He is probably guilty of the offence" or "The train will probably be late". What does the word "probable" mean? Here are a few examples of the definition from the point of view of philosophy.

> The "probably" is something which lies mid-way between truth and error (Thomasius).
> An assertion, of which the contrary is not completely self-contradictory or impossible is called probable (Reimarus).
> That which, if it were held as truth would be more than half certain, is called probable (Kent).

We shall not concern ourselves with such philosophical definitions.

Some phenomena can give different results when repeated trials of an experiment are performed. If a die is thrown in the air, the fact that it will come down, is deterministic but the observation of say, a "6", is uncertain (synonyms are random or nondeterministic). Probability is concerned with the study of such non-deterministic experiments.

The origin of the theory of probability goes back to the middle of the seventeenth century and is connected to the mathematician Fermat (1601–1665), B. Pascal (1623–1662), and Huygens (1629–1695)[10]. Among later scholars who contributed significantly to the development of the theory of probability included Jakob Bernoulli (1654–1705), Abraham De Moivre (1667–1754), Thomas Bayes (1702–1761), P. Laplace (1749–1827), S.D. Poisson (1781–1840), Karl Gauss (1856–1922), P.L. Chebyshev (1821–1894), A.A. Markov (1856–1922), A.M. Lyapunov (1857–1918), A. Khinchin (1894–1959) and A.N. Kolmogorov (1903–1987).

Historically, probability originated from the study of games of chance and early applications of the theory of probability were in such games. In the middle of the seventeenth century, a French courtier, the Chavelier de Méré wanted to know how to adjust the stakes in gambling so that in the long run, the advantage would be his. He presented his problem to Blaise Pascal, his countryman. It was in the correspondence between Pascal and Pierre Fermat, another French mathematician (a friend of Pascal's father), that the theory of probability has its beginning. Many of the probability calculations were, therefore, based on objects of gambling: the coin, the die, and the cards. Even though the use of probability in gambling today is just one of its minor applications, the use of such objects is a convention.

In the previous chapter, these objects have been used in some examples. In most probability problems, unless otherwise specified, a coin and a die are considered to be fair and a deck of cards are assumed to be well shuffled.

Definition 3.26 PROBABILITY OF EVENTS

The probability of a given event is an expression of the likelihood of the occurrence of the event.

Historically, there are three main schools of thought in defining and interpreting the probability of an event: the classical approach, the frequency approach, and the subjective approach. The first two are referred to as objective approaches.

3.5.2 CLASSICAL DEFINITION OF PROBABILITY

The classical approach to defining probability is a mathematical way of treating or measuring uncertainty. As stated earlier, much of this theory was developed out of attempts to solve problems related to games of chance, such as, the rolling of dice, tossing of coins and drawing of cards. It is based on the simple assumption that outcomes of a random experiment are equally likely and there is no reason to prefer any one of the possible outcomes to the other. This argument was called the ***principle of insufficient reason*** by Jakob Bernoulli. By this principle, all the possible outcomes of the sample space $\mathcal{S}$ should have the same weights or probabilities of occurring. Such a finite probability space[11] is usually called an ***equiprobable space***.

Definition 3.27 PROBABILITY
(CLASSICAL APPROACH)

If there are a possible outcomes favourable to the occurrence of an event $\mathcal{A}$, and b possible outcomes unfavourable to the occurrence of $\mathcal{A}$, and all possible outcomes are equally likely and mutually exclusive, then the probability that event $\mathcal{A}$ will occur is

$$P(A) = \frac{a}{a+b} = \frac{\text{Number of } \textbf{Favourable} \text{ outcomes}}{\text{Number of all } \textbf{Possible} \text{ outcomes}}$$

or equivalently

If a statistical experiment or simply an experiment $\mathcal{E}$ can lead to n mutually exclusive and equally likely simple outcomes, and if a of these outcomes have attribute $\mathcal{A}$, then

$$P(A) = \frac{a}{n}$$

To explain how the classical probability formula arises, suppose that $\mathcal{S}$ contains n points. Then the probability of each point is $\frac{1}{n}$. Suppose also that the event $\mathcal{A}$ contains a points. Then, its probability is

$$a \times \frac{1}{n} = \frac{a}{n}$$

That is

$$P(\text{event}) = \frac{\text{Number of outcomes comprising the event}}{\text{Total number of equally likely outcomes in the sample space}}$$

Example 3.33

In a well-shuffled deck of cards, a card is selected at random. What is the probability of (a) a queen? (b) a 5? (c) a spade? (d) a picture card?

Solution
The sample space for selecting a well-shuffled deck of cards was first provided in Example 3.13. There are 52 possible outcomes.

a) There are four queens in a deck of 52 cards, hence

$$P(\mathcal{A}) = \frac{4}{52}$$

b) There are four 5's in a deck of 52 cards, hence

$$P(5) = \frac{4}{52}$$

c) There are 13 spades in the deck, hence

$$P(\text{spade}) = \frac{13}{52}$$

d) There are 12 picture cards, hence

$$P(\text{picture cards}) = \frac{12}{52}$$

Example 3.34
A pair of dice is rolled once. Find the probability of rolling
(a) a sum of 7 (b) a sum of 7 or 11 (c) a double.

Solution
The sample space for rolling a pair of dice was first provided in Example 3.16. There are 36 possible outcomes.

a) There are six ways to roll a sum of 7:

$$(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)$$

Hence

$$P(\text{sum of } 7) = \frac{6}{36}$$

b) There are eight outcomes that will give a sum of 7 or 11:

$$(1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(6,5),(5,6)$$

Hence

$$P(\text{sum of 7 of 11}) = \frac{8}{36}$$

c) There are six ways to roll a double:

$$(1,1),\ (2,2),\ (3,3),\ (4,4),\ (5,5),\ (6,6).$$

Hence

$$P(\text{double}) = \frac{6}{36}$$

Definition 3.28 PROBABILITY CALCULATION
(**In Terms of Combinatorial Analysis**)

If there are n objects, m of which are of a type $\mathcal{A}$. The probability of selecting r objects of that type $\mathcal{A}$ is

$$P(\mathcal{A}) = \frac{\begin{pmatrix} m \\ r \end{pmatrix}}{\begin{pmatrix} n \\ r \end{pmatrix}}$$

Example 3.35
In a well-shuffled deck of playing cards, four cards are drawn at random. What is the probability that they are all hearts.

Solution
The number of all possible outcomes is the number of ways we can select 4 cards from 52 cards, namely

$$\begin{pmatrix} 52 \\ 4 \end{pmatrix}$$

The number of favourable outcomes is the number of ways we can select 4 hearts from a total of 13 hearts, namely

$$\begin{pmatrix} 13 \\ 4 \end{pmatrix}$$

The probability of selecting 4 hearts is

$$\frac{\begin{pmatrix} 13 \\ 4 \end{pmatrix}}{\begin{pmatrix} 52 \\ 4 \end{pmatrix}} = \frac{11}{4165}$$

Example 3.36
A box contains 10 marbles of which 6 are red and 4 are blue. Two marbles are chosen at random. Find the probability that both are red.

Solution

There are $\begin{pmatrix} 10 \\ 2 \end{pmatrix} = 45$ ways to choose 2 marbles from 10 marbles. 6 of the marbles are red so there are $\begin{pmatrix} 6 \\ 2 \end{pmatrix} = 15$ ways to choose red marbles. Hence

$$\begin{aligned} P(A) &= \frac{\text{number of ways 2 red marbles can be chosen}}{\text{number of ways two marbles can be chosen}} = \frac{\begin{pmatrix} 6 \\ 2 \end{pmatrix}}{\begin{pmatrix} 10 \\ 2 \end{pmatrix}} \\ &= \frac{15}{45} \end{aligned}$$

Because the classical approach (when it is applicable) permits determination of probability values before any experiment has been performed or before any sample events are observed, it has also been called a *priori approach*. We might adapt a well-known idiom and describe the classical probability as "being wise before the event".

The classical probability is objective in the sense that it is based on deductions from a set of assumptions. It is very useful not only in games of chance but also in a great variety of situations where gambling devices are used to make random selection.

Disadvantages of the Classical Approach

a) *Circulation of Definition*
   The term equally likely used in the definition actually means *equally probable*. Thus, in the definition, use is made of the concept which itself is to be defined.

b) *Limited Applicability*
   The classical approach is all very well when dealing with the games-of-chance type of problems, but its field of application is limited.

   i) Even in the theory of games, although the outcome is unpredictable, some results are not equally likely. For instance, if we are tossing a coin which is weighted so that a head (H) is three times as likely to appear as a tail (T), then a head is much more a likely outcome of a throw than a tail.
   ii) Furthermore, there are many more situations in real-world where the possibilities that arise cannot be considered as equally likely. For example, the probability that a ninety-year-old will live the next twenty years is not as equally likely as a twenty-year-old person surviving the next twenty years.

c) *Restriction of Definition*
The classical approach to probability can only be applied to situations where

i) we can determine *a priori*, or
ii) without experimentation, the probability that an event will occur.

In more complex situations, such as the ones we often meet in the real-world, we often cannot assign probabilities *a priori* and the classical approach cannot be used. Again, the real-world has not such thing as a perfectly balanced coin, perfect die and so on, and therefore, the assumption of perfection will cause slightly wrong probability assumptions.

3.5.3 FREQUENCY DEFINITION OF PROBABILITY

The relative frequency approach overcomes the disadvantages of the classical approach by using the relative frequencies of past occurrences as probabilities. This approach claims that the only valid procedure for determining event probabilities is through repetitive experiments.

Definition 3.29 PROBABILITY
(Relative Frequency Approach)

Suppose that an experiment is repeated a large number of times under the same conditions with each of the n trials being independent of one another. If an event $\mathcal{A}$ is observed to occur in $n(\mathcal{A})$ of the trials, the probability of the event $\mathcal{A}$ is

$$R(\mathcal{A}) = \frac{n(\mathcal{A})}{n} = \frac{m}{n}$$

where $m = n(\mathcal{A})$ is the number of times event $\mathcal{A}$ occurs

Sometimes, if we want to explicitly indicate that the event $\mathcal{A}$ depends on the number of trials, we write it as $R_n(A)$. The different probabilities for different number of trials stabilize or approach a limit as the number of trials or experiments increases. A mathematician may write this as

$$\lim_{n\to\infty} R_n = \lim_{n\to\infty} \frac{n(\mathcal{A})}{n} = P(\mathcal{A})$$

where "$\lim\limits_{n\to\infty}$" means "when the number of trials increases without bound".

This type of "limit" is called a *probability limit* and $P(\mathcal{A})$ is a unique number to which the outcome ultimately settles. That is, when n is very large, this ratio may be taken as an approximation to the "true" probability function that underlies the experiment in the sense that $R_n(\mathcal{A})$ will become close to $P(\mathcal{A})$. This assertion is not about mathematical limits but about empirical phenomena. Of course, there is no way actually to repeat an experiment an infinite number of times to evaluate the limit, but the intuitive idea is that such a limit must exist.

It is obvious that with the relative frequency approach, there is just no way of estimating the probability without empirically drawing a sample or performing an experiment and it is for this reason that this approach has also been called the ***empirical approach***. It is also sometimes called a *posteriori approach* because probability values are determined only after events are observed. We may, therefore, describe it as "being wise after the event". The relative-frequency definition is also objective because the probability of an event is determined by repeated empirical observations.

Note

We must keep in mind that strictly speaking $\frac{n(\mathcal{A})}{n}$ is only an estimate of $P(\mathcal{A})$.

Properties of the Frequency Approach

Property 1

$$0 \le R_n(\mathcal{A}) \le 1$$

where
$R_n(\mathcal{A}) = 0,$ if $\mathcal{A}$ occurs in none of the n trials;
$R_n(\mathcal{A}) = 1$ if $\mathcal{A}$ occurs in all the n trials.

This property is intuitively clear. If $n(\mathcal{A})$ is the number of trials in which $\mathcal{A}$ occurs out of n trials, then

$$0 \le n(\mathcal{A}) \le n$$

and we have

$$0 = \frac{0}{n} \le R_n(\mathcal{A}) = \frac{n(\mathcal{A})}{n} \le \frac{n}{n} = 1$$

and the result follows.

Property 2

a) If $\mathcal{A}$ and $\mathcal{B}$ are any two events, then

$$n(\mathcal{A} \cup \mathcal{B}) = n(\mathcal{A}) + n(\mathcal{B}) - n(\mathcal{A} \cap \mathcal{B})$$

(by the inclusive-exclusive principle in Theorem 2.2), so that

$$R(\mathcal{A} \cup \mathcal{B}) = R(\mathcal{A}) + R(\mathcal{B}) - R(\mathcal{A} \cap \mathcal{B})$$

b) In particular, if events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive, so that $n(\mathcal{A} \cap \mathcal{B}) = \emptyset$, then

$$n(\mathcal{A} \cap \mathcal{B}) = n(\mathcal{A}) + n(\mathcal{B})$$

and hence

$$R(\mathcal{A} \cup \mathcal{B}) = R(\mathcal{A}) + R(\mathcal{B})$$

Example 3.37
A coin is tossed 1,000 times and tail comes up 508 times. Estimates the probability of the tail.

Solution

$n = 1,000 \quad n(\mathcal{A}) = 5508$

Hence

$$P(\text{Tail}) = \frac{n(\mathcal{A})}{n} = \frac{508}{1,000} = 0.508$$

The ratio $\dfrac{508}{1000}$ is rather an estimate but we would be very reluctant to state that the probability of getting a tail on that coin is 0.508, because if we had stopped at some other number of trials the ratio would have been different. What we could say is that if we toss the coin many times it will come up tail 50.8% of the times. Intuitively, we feel that as the number of trials increases, the relative frequency will settle down to some stable value, greater than zero and less than unity.

Disadvantages of the Frequency Approach

a) The relative frequency definition of probability assumes that the conditions of the experiments are reproducible This assumption may not be valid if "time" is one of the factors to be considered.
b) Different probabilities (relative frequencies) for different numbers of trials or experiments will be obtained in an attempt to estimate the probability of an event.

c) The expression "large number" of trials in the definition is vague and this creates even more practical problems. In an attempt to reach this number, we may face the problem of cost, in terms of money and time. This may discourage the good-intended experimenter to use an "insufficient" number of trials.
d) The concepts of the relative frequency approach in defining probability is somewhat intuitive which is a plus for understanding but it is a minus for a rigorous development.

In spite of these limitations, the frequency approach provides a link between the outside world and the mathematical theory of probability we study. It enables us to construct probability models of natural phenomena and study them. If past is any guide to the future and if by and large nature is stable, the method of assigning probabilities to real-life situations should be satisfactory.

3.5.4 SUBJECTIVE DEFINITION OF PROBABILITY

The subjective approach to probability was brought to the attention of statisticians by Professor Savage. It is an expression of our judgement of the likelihood of the occurrence of an event.

Sometimes, we attempt to quantify the possibility of an event happening. When we do so, we are only expressing our opinions based on our feelings and convictions of the situation, our experiences and backgrounds and our degree of rational belief.

Definition 3.30 PROBABILITY
(Subject Approach)

Subjective probability is the degree of rational belief by an individual that the event will occur, based on all evidence available to that individual.

Equivalently,
The subjective probability of $\mathcal{A}$ is a measure of confidence that a reasonable person assigns to the event $\mathcal{A}$

In this definition, it is obvious that a person assigns a probability to an event in accordance with his or her subjective assessment, taking into consideration all his or her knowledge and experience; and because the probability value is a personal judgment, the subjective approach has been called the *personalistic approach* to defining probability or simply ***personal probability***.

As an example, we might say that we are "almost sure" that it will rain now; in which case "almost sure" is a non-quantitative measure of subjective probability. We might wish to quantify the subjective probability and say, perhaps, "I am 90 percent certain that it will rain now", but the 90 percent is still a degree of belief or subjective probability.

As can be seen, this subjective approach may be applied especially

a) where there is little or no direct evidence (either the events have not yet occurred, or have occurred only once) so that there is no choice but to consider collateral (indirect) evidence, educated guesses, and perhaps intuition and other subjective factors;
b) in situations that do not require
   i) an experiment with a large number of trials, or
   ii) the assumption of statistical regularity.
c) in situations where from experience, there has been massive evidence in favour of the event; for example, a person who has seen the sun appearing in the sky everyday will conclude subjectively that the probability that the sun will appear everyday is one.

This approach to probability, also known as Bayesian approach, has been developed relatively recently and is related to Bayesian decision analysis. Although the subjective view of probability has enjoyed increased attention over the years, it has not been fully accepted by statisticians who have traditional orientations.

Disadvantages of the Subjective Approach

a) The subjective approach allows an individual to act rationally in accord with his or her beliefs and this is sometimes a disadvantage of the method. The assigned probabilities are formulated taking into account all the prejudices of the individual. Therefore, different people faced with the same situation may come up with completely different probabilities.
b) Since the assigned probabilities are in a sense non-objective, subjective probability is considered to be inappropriate in scientific studies.

3.5.5 COMPARISON OF THE THREE APPROACHES TO PROBABILITY

The three distinct approaches to probabilities, namely, the classical, relative frequency and subjective approaches, raise interesting philosophical problems. We know that $P(\mathcal{A}) = 0$ can never be less than "zero" nor exceed "one" but do $P(\mathcal{A}) = 0$ imply impossibility and $P(\mathcal{A}) = 1$ absolute certainty? Let us now look at probability values as interpreted in these approaches.

$P(\mathcal{A}) = 0$

To the classical probabilist, this means that event $\mathcal{A}$ ***cannot occur***. However, to the frequency probabilist (frequentist), event $\mathcal{A}$ ***has never occurred***. Of course this does not mean it cannot occur in the future. The subjective probabilist (subjectivist) will say ***I think that the event*** $\mathcal{A}$ ***will not occur*** and which also does not mean that it cannot occur.

$P(\mathcal{A}) = 1$

To the classical probabilist, it means that event $\mathcal{A}$ ***must always occur***. To the frequentist, it means that event $\mathcal{A}$ ***has always occurred***, which does not imply that it must occur in the future. The subjectivist will say that ***I think that event*** $\mathcal{A}$ ***will occur*** and this also does not imply it must occur.

$0 < P(\mathcal{A}) < 1$

The classical probabilist will say that event $\mathcal{A}$ ***is expected to occur*** $n(\mathcal{A})$ ***out of*** n ***equally likely outcomes***, the frequentist will say that event $\mathcal{A}$ ***has occurred*** $n(\mathcal{A})$ ***times out of*** n ***trials*** and the subjectivist thinks that event $\mathcal{A}$ ***will happen*** $n(\mathcal{A})$ ***times out of*** n ***trials***.

Without doubt, the empirical and subjective approaches are more interesting and more useful than *a priori* approach. However, in an introductory text such as this, it is preferable to concentrate attention on classical probability, and unless we state to the contrary, an *a priori* approach is used.

3.5.6 MODERN DEFINITION OF PROBABILITY

Even though three different conceptual approaches have been developed for defining and interpreting the probability of an event $\mathcal{A}$, at the mathematical level, there is no any disagreement about the properties of a probability function[12] and the calculation of probabilities of events. A mathematically useful treatment of probability should lay a common ground so that every one would be speaking the same language. However the mathematical advances in probability theory were relatively limited and difficult to establish on a firm basis until the Russian mathematician A.N. Kolmogorov, in 1933, laid down as axiomatic approach to it.

Kolmogorov gave a simple set of three axioms[13] or postulates which a probability function is assumed to obey. This approach is what is considered the *modern approach* or *axiomatic definition* of probability.

Definition 3.31 PROBABILITY
(Axiomatic Approach)

Let $\mathcal{S}$ be a sample space whose members $e_1, e_2, \cdots$ are elementary events, $\mathcal{A} \subset \mathcal{S}$. Let $P: \; P(\mathcal{S}) \rightarrow [0,\, 1]$. Then the function $P(\mathcal{A})$ is said to be a probability function if it satisfies the following three axioms:

Axiom 1 ***(Non-negativity)***
For every event $\mathcal{A}$
$$0 \leq P(\mathcal{A}) \leq 1$$

Axiom 2 ***(Normed)***
$$P(\mathcal{S}) = 1$$

Axiom 3 ***(Finite Additivity)***
If $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events $(\mathcal{A} \cap \mathcal{B} = \emptyset)$, then
$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$$

Note

Recall that $[0,\, 1]$ denotes $\{x \;\in\; R | 0 \leq x \leq 1\}$

Axiom 1 states that the probability of an event $\mathcal{A}$ exists and can never be less than zero (which corresponds to an "impossibility" of an event) nor greater than one (which corresponds to a "certainty" of an event).

Axiom 2 states that the events $\mathcal{A}_i's$ comprising the sample space must be exhaustive: it must be a certainty that at least one of the mutually exclusive events in the sample space will take place in each trial of the experiment (and when all are taken together their total probability is 1).

Axiom 3 implies that if the probability can be assigned to each of the sample points in $\mathcal{S}$, then we can obtain the probability of any event defined on $\mathcal{S}$ merely by summing the separate probabilities of all the sample points that are members of the event set under consideration.

These three axioms satisfy our intuitive notions of what we mean by a probability function or probability measure and seem motivated by the properties of the relative frequency of an event. The axiomatic approach launched probability as a separately identifiable subfield of mathematics and provided a mathematical foundation to the classical theory of probability, hence this approach is sometimes called the ***mathematical approach to defining probability***.

Example 3.38
Suppose in a certain experiment, the sample space $\mathcal{S}$ consists of 3 elements: $S = \{e_1, e_2, e_3\}$. Which of the following functions defines a probability function of $\mathcal{S}$?

a) $P(e_1) = 1/5, \quad P(e_2) = 1/2, \quad P(e_3) = 3/10$
b) $P(e_1) = 3/4, \quad P(e_2) = 1/4, \quad P(e_3) = 1/2$
c) $P(e_1) = 3/4, \quad P(e_2) = -1/4, \quad P(e_3) = 1/2$
d) $P(e_1) = 1/4, \quad P(e_2) = 1/3, \quad P(e_3) = 0$

Solution

a) The function does define a probability function on $\mathcal{S}$ since each value is non-negative and the sum of the values is one; that is $\frac{1}{5} + \frac{1}{2} + \frac{3}{10} = 1$
b) The function does not define a probability function on $\mathcal{S}$ since the sum of the values on the points is greater than; $\frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12} > 1$
c) The function does not define a probability function on $\mathcal{S}$, since $P(e_2) = -\frac{1}{2}$, a negative number.
d) The function does not define a probability function $\mathcal{S}$ since the sum of the values is less than 1; $\frac{1}{4} + \frac{1}{3} = \frac{7}{12} < 1$

This chapter has laid a good foundation upon which we can understand the basic theory of probability which is developed systematically in the next chapter. It is advisable that the reader goes through most of the exercises to ensure that the materials presented in this chapter are well assimilated.

EXERCISES

3.1 A fair die is thrown twice. List the elements of the following events:

$$
\begin{aligned}
\mathcal{A} &= \{\text{the sum of the two faces is 3}\}\\
\mathcal{B} &= \{\text{the sum of the two faces is at least 8}\}\\
\mathcal{C} &= \{\text{the sum of the two faces is not 8}\}\\
\mathcal{D} &= \{\text{both dice show the same number}\}\\
\mathcal{E} &= \{\text{both faces are divisible by 3}\}\\
\mathcal{F} &= \{\text{at least one of the faces is devisible by 3}
\end{aligned}
$$

3.2 Refer to Exercise 3.1. Express explicitly the event that

a) $\mathcal{A}$ and the complement of E occur;
b) only D occurs;
c) $\mathcal{D}$ and $\mathcal{C}$ occur;
d) $\mathcal{D}$ and the complement of $\mathcal{C}$ occur;
e) $\mathcal{A}$ or $\mathcal{B}$ occurs.

3.3 Let $\mathcal{A}, B$ and $\mathcal{C}$ be three events. For each of the following statements, write in set notation the event specified by the statement:

a) Only $\mathcal{A}$ occurs:
b) Exactly one of $\mathcal{A}$, B, C occurs
c) At least one of the events occurs;
d) Both $\mathcal{A}$ and $\mathcal{B}$ but not $\mathcal{C}$ occur;
e) Exactly two of the events occur;
f) At least two of the events occur;
g) All three occur;
h) Either $\mathcal{A}$ or $\mathcal{B}$ occurs, but not $\mathcal{C}$;
i) $\mathcal{C}$ occurs and either $\mathcal{A}$ or $\mathcal{B}$, but not both

3.4 Refer to Exercise 3.3. Draw a Venn diagram and on it shade the area corresponding to the event specified by the statement in set notation:

3.5 Two dice are thrown once. Let

$\mathcal{E} =$ the event that the sum of the numbers that appear on the faces of the dice is odd;
$\mathcal{F} =$ at least one of the dice shows 1, and
$\mathcal{G} =$ the sum is 5.

List the elements of the event

$$\mathcal{E}\cup\mathcal{F}, \quad \mathcal{E}\cap\mathcal{F}, \quad \mathcal{F}\cap\mathcal{G}, \quad \mathcal{E}\cap\mathcal{F}^c, \text{ and } \mathcal{E}\cap\mathcal{F}\cup\mathcal{G}$$

3.6 A box contains 3 marbles, 1 red, 1 greed and 1 blue. Consider an experiment that consists of taking 1 marble from the box, then replacing it in the box and drawing a second marble from the box. Describe the sample space.

3.7 Repeat the experiment in Exercise 3.6 but this time the second marble is drawn without first replacing the first marble.

3.8 A box contains 10 pairs of shoes. If 8 shoes are randomly selected, what is the probability that there will be

a) no complete pair;
b) exactly one complete pair.

3.9 If 4 married couples are arranged in a row, find the probability that no husband sits next to his wife.

3.10 If 4 married couples are arranged in a circle, find the probability that

a) the husbands sit by their wives;
b) the men and the women are seated alternative.

3.11 Eight married couples enter a supermarket. If the owner of the supermarket selects 2 persons at random to give them free lottery coupons find the probability that

a) they are man and wife;
b) one is a male and the other is a female.

3.12 Refer to Exercise 3.11. If 6 persons are chosen at random, find the probability that

a) three married couples are chosen;
b) no married couple is among the 6 persons.

3.13 Refer to Exercise 3.11. If 16 persons are divided into a 8 pairs, find the probability that

a) each pair is a married couple;
b) each pair contains a male and a female.

3.14 Refer to Exercise 2.6. Find the probability that books on the same subject are together.

3.15 A bus starts with 6 people and stop at 10 different stops. Assuming that passengers are equally likely to depart at any stop, find the probability that,

a) no two passengers leave at the same bus stop;
b) at least one person leaves at each stop;
c) all passengers leave at the same stop.

3.16 At Kwame's birthday party, there were 22 people present. Find the probability that

a) two of them have the same birthday;
b) someone has the same birthday as Kwame.

3.17 A table consists of four digit random numbers. What is the probability that four consecutive random digits are all different.

3.18 Two balls are drawn with replacement from a box containing 3 white and 2 black balls. Calculate the probability that

a) both balls drawn will be of the same colour;
b) at least one of balls drawn will be white.

3.19 A box contains 40 good and 10 defective fuses. If 10 fuses are selected what is the probability that they will all be good.

3.20 Suppose a sample space $\mathcal{S}$ consists of four elements $\mathcal{S} = \{e_1, e_2, e_3, e_4\}$. Which of the following functions defines a probability space on $\mathcal{S}$?

a) $P(e_1) = 2/7, \quad P(e_2) = 5/21. \quad P(e_3) = 1/3, \quad P(e_4) = 1/7$
b) $P(e_1) = 2/5, \quad P(e_2) = 1/2, \quad P(e_3) = -1/2, \quad P(e_4) = 3/5$
c) $P(e_1) = 2/7, \quad P(e_2) = 5/21. \quad P(e_3) = 1/3, \quad P(e_4) = 1/7$
d) $P(e_1) = 0, \quad P(e_2) = 0, \quad P(e_3) = 1, \quad P(e_4) = 0$
e) $P(e_1) = 1/8, \quad P(e_2) = 1/3, \quad P(e_3) = 1/2, \quad P(e_4) = 1/4$

3.21 Suppose $\mathcal{S} = \{e_1, e_2, e_3\}$. Let P be a probability function on $\mathcal{S}$.

a) Find $P(e_1)$ if $P(e_2) = \frac{1}{5}$ and $P(e_3) = \frac{1}{3}$;
b) Find $P(e_1)$ and $P(e_2)$ if $P(e_3) = \frac{1}{3}$ and $P(e_2) = 3P(e_1)$;
c) Find $P(e_1)$ if $P(\{e_2, e_3\}) = \frac{2}{3}$ and $P(e_3) = \frac{1}{4}$

3.22 A coin is weighted so that a head $\mathcal{H}$ is three times as likely to appear as a tail $\mathcal{T}$. Find $P(\mathcal{T})$ and $P(\mathcal{H})$.

3.23 Three athletes, Yaa, Ama, and Afua run a 100 metre race. Yaa is twice as likely to win as Ama and Ama is thrice as likely to win as Afua.

a) What are their respective probabilities of winning;
b) What is the probability that Yaa or Afua wins.

3.24 Three men, Fosu, Yeboah, and Abu and two women, Attaa and Adjoa are in a spelling competition. Those of the same sex have equal probabilities of winning, but each woman is twice as likely to win as any man.

a) Find the probability that a woman wins the competition;
b) If Fosu and Attaa are married, find the probability that one of them wins the competition.

3.25 If n distinguishable balls are distributed at random into r boxes, what is the probability that box 1 has exactly j balls, $0 \le j \le n$.

3.26 A box has b black balls and r red balls. Their colours are distinguishable but balls of the same colours are not distinguishable. Balls are drawn from the box one at a time without replacement. Find the probability that the first black ball selected is drawn at the n^{th} trial.

3.27 Repeat Exercise 3.26 for the case when the first ball is not replaced.

3.28 Suppose r objects are drawn from a set of n objects without replacement. Find the probability that k given objects are selected.

3.29 Suppose n objects are permuted at random among themselves. Find the probability that k specified objects occupy k specified positions.

3.30 Two dice are used, each loaded so that the probabilities of throwing 1, 2, 3, 4, and 6 are

$$\frac{1-x}{6},\ \frac{1+2x}{6},\ \frac{1-x}{6},\ \frac{1+x}{6},\ \frac{1-2x}{6},\ \frac{1+x}{6}$$

respectively. Compute the probability that, in one rolling of two such dice, we obtain a total of (a) a seven (b) a six.

3.31 Refer to Exercise 2.11. What is the probability that the two guests who are not on talking terms will sit next to each other.

3.32 Refer to Exercise 2.16 (d). Find the probability that all T's are together.

4 BASIC PROBABILITY LAWS AND THEOREMS

4.1 INTRODUCTION

From the three axioms stated in Definition 3.31, other laws and properties of probability can be established. We shall in this section introduce some of these properties via a series of basic theorems. These theorems together with the basic definitions given here are the basis for computing probabilities.

Theorem 4.1

If $\emptyset$ is the empty set, then

$$P(\emptyset) = 0$$

Proof
Let $\mathcal{S}$ be the sample space. Then

$$\begin{aligned}
\mathcal{S} &= \mathcal{S} \cup \emptyset \qquad \text{(from the Identity Law in Table 1.1)} \\
P(\mathcal{S}) &= P(\mathcal{S} \cup \emptyset) \\
&= P(\mathcal{S}) + P(\emptyset) \qquad \text{(from Axiom 3 of Definition 3.31)}
\end{aligned}$$

since $\mathcal{S}$ and $\emptyset$ are disjoint. Subtracting $P(\mathcal{S})$ from both sides, we get

$$P(\emptyset) = 0$$

This theorem states that if the event set is the null set, then the event is an "impossibility" and the probability of occurrence is zero. This is obvious. The null set contains no sample point; hence no weight can be assigned to it.

Example 4.1
A fair die is tossed once. What is the probability of obtaining a 7?

Solution
The sample space for this experiment is

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

Let $\mathcal{A}$ be "the event of obtaining a 7". Then

$$\mathcal{A} = \emptyset$$

since there is no such number in $\mathcal{S}$. Hence

$$P(\mathcal{A}) = P(\emptyset) = 0$$

4.2 ADDITION LAW OF PROBABILITY

4.2.1 GENERAL ADDITION LAW OF PROBABILITY

Theorem 4.2

If $\mathcal{A}$ and $\mathcal{B}$ are events defined over the same sample space $\mathcal{S}$ and if they overlap ($\mathcal{A} \cap \mathcal{B} \neq \emptyset$) then the probability that either $\mathcal{A}$ or $\mathcal{B}$ (or both) will occur is the sum of their separate probabilities less the probability of their joint occurrence:

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$$

Proof
From Section 1.5, $\mathcal{A} \cup \mathcal{B}$ can be decomposed into the mutually exclusive events $\mathcal{A} \backslash \mathcal{B}$ and $\mathcal{B}$:

$$\mathcal{A} \cup \mathcal{B} = (\mathcal{A} \backslash \mathcal{B}) \cup \mathcal{B}$$

so that

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A} \backslash \mathcal{B}) + P(\mathcal{B}) \qquad (i)$$

Again from Section 1.5,

$$\mathcal{A} = (\mathcal{A} \backslash \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B})$$

Therefore,

$$P(\mathcal{A}) = P(\mathcal{A} \backslash \mathcal{B}) + P(\mathcal{A} \cap \mathcal{B})$$

so that

$$P(\mathcal{A} \backslash \mathcal{B}) = P(\mathcal{A}) - P(\mathcal{A} \cap \mathcal{B}) \qquad (ii)$$

Substituting $P(\mathcal{A}\backslash\mathcal{B})$ in (ii) into (i), we obtain

$$\begin{aligned} P(\mathcal{A}\cup\mathcal{B}) &= [P(\mathcal{A}) - P(\mathcal{A}\cap\mathcal{B})] + P(\mathcal{B}) \\ &= P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A}\cap\mathcal{B}) \end{aligned}$$

Aliter

Let us represent $(\mathcal{A}\cup\mathcal{B})$ in the diagram presented in Figure 4.1.

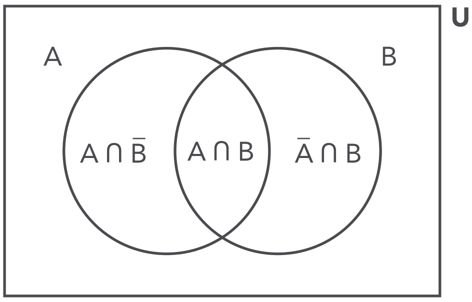


Fig. 4.1 Joint Events of $\mathcal{A}$ and $\mathcal{B}$

We can observe from Fig. 4.1 that

$$\mathcal{A} = (\mathcal{A}\cap\mathcal{B}) \cup (\mathcal{A}\cap\overline{\mathcal{B}})$$

so that[14]

$$P(\mathcal{A}\cup\mathcal{B}) = P(\mathcal{A}\backslash\mathcal{B}) + P(\mathcal{B}) \qquad (i)$$

Similarly,

$$P(\mathcal{B}) = P(\mathcal{A}\cap\mathcal{B}) + P(\overline{\mathcal{A}}\cap\mathcal{B}) \qquad (ii)$$

Now, adding (i) and (ii) we obtain

$$P(\mathcal{A}) + P(\mathcal{B}) = P(\mathcal{A}\cap\mathcal{B}) + P(\mathcal{A}\cap\overline{\mathcal{B}}) + P(\mathcal{A}\cap\mathcal{B}) + P(\overline{\mathcal{A}}\cap\mathcal{B}) \qquad (iii)$$

We can again observe from Fig. 4.1 that

$$(\mathcal{A}\cup\mathcal{B}) = (\mathcal{A}\cap\overline{\mathcal{B}}) + (\mathcal{A}\cap\mathcal{B}) + (\overline{\mathcal{A}}\cap\mathcal{B})$$

so that

$$P(\mathcal{A}+\mathcal{B}) = P(\mathcal{A}\cap\overline{\mathcal{B}}) + P(\mathcal{A}\cap\mathcal{B}) + P(\overline{\mathcal{A}}\cap\mathcal{B}) \qquad (iv)$$

The expression at the right-hand side of Eq. (iv) is the same as the last three terms of the expression on the right-hand side of Eq. (iii).

Therefore, Eq. (iii) can be written as

$$P(\mathcal{A}) + P(\mathcal{B}) = P(\mathcal{A} \cap \mathcal{B}) + P(\mathcal{A} + \mathcal{B})$$

Hence,

$$P(\mathcal{A} + \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$$

Example 4.2
The student union of the Regent University of Science and Technology decided to elect a representative from five of its members to represent it on an Exchange Programme Committee. Profiles of the five students were as follows: a male Accounting student, a male Theology student, a female Psychology student, a female Computer Science student, and a male Engineering student.

The Union decided to elect the representative by drawing a name from a hat. What is the probability that the representative will be either a female or from the Faculty of Science and Engineering.

Solution
Let the event $\mathcal{A}$ represent a female student and event $\mathcal{B}$ represent a student from the Faculty of Science and Engineering. Then

$$n(\mathcal{S}) = 5, \quad n(\mathcal{A}) = 2, \quad n(\mathcal{B}) = 2, \quad n(\mathcal{A} \cap \mathcal{B}) = 1$$

Hence the desired probability is

$$\begin{aligned} P(\mathcal{A} \cup \mathcal{B}) &= P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B}) \\ &= \frac{2}{5} + \frac{2}{5} - \frac{1}{5} = \frac{3}{5} \end{aligned}$$

Corollary 4.1
If $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are three events defined on the same sample space, $\mathcal{S}$, then

$$\begin{aligned} P(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}) &= P(\mathcal{A}) + P(\mathcal{B}) + P(\mathcal{C}) - P(\mathcal{A} \cap \mathcal{B}) - P(\mathcal{A} \cap \mathcal{C}) \\ &\quad - P(\mathcal{B} \cap \mathcal{C}) + P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) \end{aligned}$$

Example 4.3
A school offers its 200 students three science subjects: Mathematics $(\mathcal{M})$ Physics $(\mathcal{P})$ and Chemistry $(\mathcal{C})$ The following are data on the number of students who are offered the various subjects:

$$\begin{aligned} &n(\mathcal{M}) = 60, \quad (\mathcal{P}) = 40, \quad (\mathcal{C}) = 30, \quad (\mathcal{M} \cap \mathcal{P}) = 10, \quad (\mathcal{M} \cap \mathcal{C}) = 5, \\ &(\mathcal{P} \cap \mathcal{C}) = 3, \quad (\mathcal{M} \cap \mathcal{P} \cap \mathcal{C}) = 1. \end{aligned}$$

A student is selected at random from this school, what is the probability that he is studying at least one of the three science subjects.

Solution

$$\begin{aligned} P(\mathcal{M} \cup \mathcal{P} \cup \mathcal{C}) &= P(\mathcal{M}) + P(\mathcal{P}) + P(\mathcal{C}) - P(\mathcal{M} \cap \mathcal{P}) \\ &\quad - P(\mathcal{M} \cap \mathcal{C}) - P(\mathcal{P} \cap \mathcal{C}) + P(\mathcal{M} \cap \mathcal{P} \cap \mathcal{C}) \\ &= \frac{60}{200} + \frac{40}{200} + \frac{30}{200} - \frac{10}{200} - \frac{5}{200} - \frac{3}{200} + \frac{1}{200} = \frac{113}{200} \end{aligned}$$

Corollary 4.2

For any events $\mathcal{A}_1,\ \mathcal{A}_2,\ \ldots,\ \mathcal{A}_n$ defined over the same sample space, $\mathcal{S}$,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} \mathcal{A}_i\right) &= \sum_{i=1}^{n} P(\mathcal{A}_i) - \sum_{\substack{i=1 \\ i<j}}^{n} \sum_{j=1}^{m} P(\mathcal{A}_i \cap \mathcal{A}_j) \\ &\quad + \sum_{\substack{i=1 \\ i<j<k}}^{n} \sum_{j=1}^{m} \sum_{k=1}^{r} P(\mathcal{A}_i \cap \mathcal{A}_j \cap \mathcal{A}_k) - \cdots \\ &\quad + (-1)^{n-1} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_n) \end{aligned}$$

That is, the sum of the second term on the right hand side is over all distinct pairs of sets, that of the third over all distinct triples of sets, and so forth.

This corollary can be proved by the principle of mathematical induction; by applying Theorem 4.2 to

$$P(\mathcal{A} \cup \mathcal{Z})$$

where $\mathcal{Z} = \mathcal{B} \cup \mathcal{C}$.

Example 4.4

St. Augustine College is presenting six athletes for the forthcoming Inter-School Games and each of athletes has been promised a laptop computer by the school's authority if at least one of them wins a Gold medal. Find the probability that each athlete will be given a car after the games if

a) each of the athletes has a probability of winning of $\frac{2}{5}$;

b) the probability that the first athlete would win is $\frac{1}{4}$, the second is $\frac{2}{5}$, the third is $\frac{1}{5}$, the fourth is $\frac{2}{7}$, the fifth is $\frac{1}{6}$, and the sixth is $\frac{4}{9}$.

Solution

Let $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_6$ be the event that the 1^{st}, 2^{nd}, $\cdots$, 6^{th} athlete wins a Gold medal. We are required to calculate $P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_6)$.

a) Since all athletes have an equal probability of winning a Gold medal,

$$P(A_1) = P(A_2) = P(A_3) = P(A_4) = P(A_5) = P(A_6) = \frac{2}{5}$$

(There are $\binom{6}{1}$ in all);

$$P(\mathcal{A}_1 \cap \mathcal{A}_2) = P(\mathcal{A}_1 \cap \mathcal{A}_3) = \cdots = P(\mathcal{A}_5 \cap \mathcal{A}_6) = P(A_1)P(A_2) = \left(\frac{2}{5}\right)^2$$

(There are $\binom{6}{2}$ pairs in all);

$$\begin{aligned}&P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) = P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_4) = \cdots = P(\mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6)\\ &= P(A_1)P(A_2)P(A_3) = \left(\frac{2}{5}\right)^3\end{aligned}$$

(There are $\binom{6}{3}$ triples in all);

$$\vdots$$

$$\begin{aligned}P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) &= P(A_1)P(A_2)P(A_3)P(A_4)P(A_5)P(A_6)\\ &= \left(\frac{2}{5}\right)^6\end{aligned}$$

Hence

$$
\begin{aligned}
&P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_6) \\
&= \binom{6}{1} P(\mathcal{A}_1) - \binom{6}{2} P(\mathcal{A}_1 \cap \mathcal{A}_2) + \binom{6}{3} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) \\
&- \binom{6}{4} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3 \cap \mathcal{A}_4) + \binom{6}{5} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5) \\
&- \binom{6}{6} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5) \cap \mathcal{A}_6 \\
&= \binom{6}{1}\left(\frac{2}{5}\right) - \binom{6}{2}\left(\frac{2}{5}\right)^2 + \binom{6}{3}\left(\frac{2}{5}\right)^3 - \binom{6}{4}\left(\frac{2}{5}\right)^4 + \binom{6}{5}\left(\frac{2}{5}\right)^5 - \binom{6}{6}\left(\frac{2}{5}\right)^6 \\
&= \frac{12}{5} - \frac{60}{25} + \frac{160}{125} - \frac{240}{625} + \frac{192}{3125} - \frac{64}{15625} = 0.953344
\end{aligned}
$$

b) $P(A_1) = \frac{1}{4}, \quad P(A_2) = \frac{2}{5}, \quad P(A_3) = \frac{1}{5}, \quad P(A_4) = \frac{2}{7}, \quad P(A_5) = \frac{1}{6},$

$$P(A_6) = \frac{4}{9}. \qquad (i)$$

Since the events $A_i, \ (i = 1, 2, \cdots, 6)$ are not mutually exclusive we use Corollary 4.2:

$$
\begin{aligned}
P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_6) \quad = \quad & P(\mathcal{A}_1) + P(\mathcal{A}_2) + \cdots + P(\mathcal{A}_6) - P(\mathcal{A}_1 \cap \mathcal{A}_2) \\
& - P(\mathcal{A}_1 \cap \mathcal{A}_3) - \cdots - P(\mathcal{A}_5 \cap \mathcal{A}_6) \\
& + P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) + P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_4) \\
& + \cdots + P(\mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) \\
& - P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) \qquad (ii)
\end{aligned}
$$

Substituting the probabilities from (i) into (ii), the reader should verify that the result is

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_6) = \frac{37}{42}$$

We realise that this approach is indeed cumbersome. A shorter approach will be shown in Section 4.4 when we discuss independence.

Theorem 4.2, together with its corollaries, is called ***the general law (rule) of addition***, because it can be applied to any events.

If the events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive, the last term in Theorem 4.2, $P(\mathcal{A} \cap \mathcal{B}) = \emptyset$, and Theorems 4.2 and Axiom 3 in Definition 3.31 are the same. Thus, the formula in Theorem 4.2 is true whether or not $\mathcal{A}$ and $\mathcal{B}$ are disjoint (mutually exclusive).

4.2.2 SPECIAL ADDITION LAW OF PROBABILITY

Theorem 4.3

If two events, $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive, then the probability that both events will occur together is the sum of their individual probabilities:

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$$

Corollary 4.3
For a set of mutually exclusive events, $\mathcal{A}_i$, $i = 1, 2, \cdots, n$, the probability of occurrence of any of the $\mathcal{A}_i$ is the sum of their individual probabilities:

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup ... \cup \mathcal{A}_n) = P(\mathcal{A}_1) + P(\mathcal{A}_2) + ... + P(\mathcal{A}_n),$$

or simply

$$P\left(\bigcup_{i=1}^{n} \mathcal{A}_i\right) = \sum_{i=1}^{n} P(\mathcal{A}_i)$$

Theorem 4.3, together with its corollary (Corollary 4.3), is sometimes called ***the special law (rule) of addition***, since it is concerned with the special case when the events are mutually exclusive.

Example 4.5
A box contains 40 identical bulbs of which 10 are red, 25 are black and 5 white. A ball is selected at random from the box. What is the probability that it is red or black.

Solution
Let event $\mathcal{R}$ be selecting a ***red ball***, $\mathcal{B}$ selecting a ***black ball***, $\mathcal{W}$ selecting a ***white ball***. Now

$$n(\mathcal{R}) = 10; \quad n(\mathcal{B}) = 25; \quad n(\mathcal{W}) = 5; \quad n = 40.$$
$$P(\mathcal{R}) = \frac{10}{40}; \quad P(\mathcal{B}) = \frac{25}{40}; \quad P(\mathcal{W}) = \frac{5}{40}$$

Hence the probability of selecting a red or a black ball is

$$\begin{aligned} P(\mathcal{R} \cup \mathcal{B}) &= P(\mathcal{R}) + P(\mathcal{B}) \\ &= \frac{10}{40} + \frac{25}{40} \\ &= \frac{7}{8} \qquad (\text{Since } \mathcal{R} \cap \mathcal{B} = \emptyset) \end{aligned}$$

4.2.3 PROBABILITY OF MUTUALLY EXCLUSIVE EVENTS OF "AT LEAST", "AT MOST" AND "OR"

The understanding of the probability with the concepts of ***"at least"***, ***"at most"*** and ***"or"*** in mutually exclusive events is very important in probability. We shall realise from the example that follows that they all use the idea of the special law of addition.

Example 4.6
A die is rolled once. What is the probability that the outcome will be:

a) a 3;
b) less than 3;
c) at least 3;
d) more than 3;
e) at most 3;
f) either 4 or 6.

Solution

Let represent the outcome on the die by x. The probability of each of the outcomes is

$$P(x) = \frac{1}{6}$$

Then:

a) The probability that the outcome will be a 3 is:

$$P(x = 3) = \frac{1}{6}$$

b) The probability that the outcome will be less than 3 is:

$$\begin{aligned} P(x < 3) &= P(x = 1) + P(x = 2) \\ &= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \end{aligned}$$

c) The probability that the outcome will be at least 3 is:

$$\begin{aligned} P(x \geq 3) &= P(x = 3) + P(x = 4) + P(x = 5) + P(x = 6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} \end{aligned}$$

d) The probability that the outcome will be more than 3 is:

$$\begin{aligned} P(x > 3) &= P(x = 4) + P(x = 5) + P(x = 6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} \end{aligned}$$

e) The probability that the outcome will be at most 3 is:

$$\begin{aligned} P(x \leq 3) &= P(x = 1) + P(x = 2) + P(x = 3) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} \end{aligned}$$

f) The probability that the outcome will be either 4 or 6 is:

$$\begin{aligned} P(x = 4 \text{ or } x = 6) &= P(x = 4) + P(x = 6) \\ &= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} \end{aligned}$$

Theorem 4.4

If $\mathcal{A} \subseteq \mathcal{B}$, then

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{B})$$

Proof
If $\mathcal{A} \subseteq \mathcal{B}$, then

$$\mathcal{A} \cap \mathcal{B} = \mathcal{A}$$

so that Theorem 4.2 gives

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A})$$

from which the result follows.

4.3 LAW OF COMPLEMENTATION

There is a simple relationship between the probability of an event happening $P(\mathcal{A})$ and the probability of it not happening, $P(not\ \mathcal{A})$. This relationship is termed the *law of complementation.*

Theorem 4.5

Let $\overline{\mathcal{A}}$ be the complement of $\mathcal{A}$ with respect to the same sample space $\mathcal{S}$, then

$$P(\overline{\mathcal{A}}) = 1 - P(\mathcal{A})$$

Proof
The sample space $\mathcal{S}$ can be decomposed into mutually exclusive events $\mathcal{A}$ and $\overline{\mathcal{A}}$, that is,

$$\mathcal{S} = \mathcal{A} \cup \overline{\mathcal{A}}$$

Hence,

$$P(\mathcal{S}) = P(\mathcal{A} \cup \overline{\mathcal{A}}) = 1$$

Since $\mathcal{A}$ and $\overline{\mathcal{A}}$ are mutually exclusive,

$$P(\mathcal{A} \cup \overline{\mathcal{A}}) = P(\mathcal{A}) + P(\overline{\mathcal{A}})$$

Hence

$$P(\mathcal{A}) + P(\overline{\mathcal{A}}) = 1$$

giving

$$P(\overline{\mathcal{A}}) = 1 - P(\mathcal{A})$$

Example 4.7
Suppose a fair die is rolled twice. What is the probability of not getting a sum of five.

Solution
Let $\mathcal{A}$ be "the event of getting a sum of five". Then "the event of *not* getting a sum of five" is $\overline{\mathcal{A}}$, the complement of $\mathcal{A}$.

The sample space of this experiment has 36 equally likely outcomes (Refer to Example 3.17). Among them are 4 points that correspond to the event "getting a sum of five", namely,

$$(1,4),(4,1),(2,3),(3,2)$$

where the first number represents the number shown on the first die and the second number the one shown on the second die.

Hence

$$P(\mathcal{A}) = \frac{4}{36}$$

The probability of *not getting a sum of five* is, therefore

$$\begin{aligned} P(\overline{\mathcal{A}}) &= 1 - P(\mathcal{A}) \\ &= 1 - \frac{1}{9} = \frac{8}{9} \end{aligned}$$

4.4 INEQUALITY OF PROBABILITIES

There are a number of inequalities of probabilities but in this text, we shall state only two of them.

Theorem 4.6

If $\mathcal{A}$ and $\mathcal{B}$ are any events defined on a sample space $\mathcal{S}$ and if $\mathcal{A} \subseteq \mathcal{B}$, then

$$P(\mathcal{A}) \leq P(\mathcal{B})$$

Proof

We may decompose $\mathcal{B}$ into two mutually exclusive events as follows:

$$\mathcal{B} = \mathcal{A} \cup (\mathcal{B} \cap \overline{\mathcal{A}})$$

Hence

$$\begin{aligned} P(\mathcal{B}) &= P(\mathcal{A}) + P(\mathcal{B} \cap \overline{\mathcal{A}}) \\ &\geq P(\mathcal{A}) \qquad [\text{since } P(\mathcal{B} \cap \overline{\mathcal{A}}) \geq 0] \end{aligned}$$

Note

Theorem 4.6 is intuitively appealing, for it says that if $\mathcal{B}$ must occur whenever $\mathcal{A}$ occurs then $\mathcal{B}$ is at least as possible as $\mathcal{A}$.

Boole's Inequality

Theorem 4.7

The probability of the occurrence of at least one of the two events $\mathcal{A}$ and $\mathcal{B}$ never exceeds the sum of the probabilities of these events:

$$P(\mathcal{A} \cup \mathcal{B}) \leq P(\mathcal{A}) + P(\mathcal{B})$$

Proof

Since $P(\mathcal{A} \cap \mathcal{B}) \geq 0$, the result follows from Theorem 4.2.

This inequality can easily be extended to any number of events, as the following corollary shows.

Corollary 4.4
The probability of the occurrence of at least one of several events never exceeds the sum of the probabilities of these events:

$$P\left(\bigcup_{i=1}^{n} \mathcal{A}_i\right) \leq \sum_{i=1}^{n} P(\mathcal{A}_i)$$

In the case of three events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$,

$$P(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}) \leq P(\mathcal{A}) + P(\mathcal{B}) + P(\mathcal{C})$$

The equality sign in Corollary 4.4 holds only in the case when each pair of the given events is mutually exclusive.

We shall now consider some additional rules of probability, introducing first what is called conditional probability.

4.5 CONDITIONAL PROBABILITY

4.5.1 DEFINITION OF CONDITIONAL PROBABILITY

So far, in defining the probability of an event $\mathcal{A}$ we have assumed that the outcome corresponds to some point in the given sample space $\mathcal{S}$. Thus, when we use the symbol $P(\mathcal{A})$ for the probability of $\mathcal{A}$, we really mean the probability of $\mathcal{A}$ given some sample space $\mathcal{S}$. We use the notation $P(\mathcal{A}|\mathcal{S})$ to make clear that we are referring to a particular sample space $\mathcal{S}$ and we read as "the conditional probability of $\mathcal{A}$ relative to $\mathcal{S}$". Of course, we usual use the conventional notation $P(\mathcal{A})$ whenever the choice of $\mathcal{S}$ is clearly understood.

Suppose we have some additional information that the outcome of a trial is contained in a subset $\mathcal{B}$ of the sample space $\mathcal{S}$, with $P(\mathcal{B}) \neq 0$. The knowledge of the occurrence of the event $\mathcal{B}$ in effect reduces the original sample space $\mathcal{S}$ to one of its subsets and this may change the probability of the occurrence of the event $\mathcal{A}$. The resulting probability is what is known as *conditional probability.*

We introduce the definition and use of conditional probability with an example.

Example 4.8
A class consists of 15 Science students and 25 Arts students. A student has broken a chair.

a) Find the probability that Kofi, who is one of the students in the class, broke the chair.
b) If it is known that the student who broke the chair is a Science student, and Kofi is a Science student, what is the probability that it was Kofi?

Solution

a) $n(\mathcal{S}) = 15 + 25 = 40.$

$$P(\text{Kofi broke the chair}) = \frac{1}{40}$$

since each of the 40 students has an equal chance of breaking the chair.

b) Given the extra information that the student who broke the chair is a Science student reduces the sample space to only the science students so that n(Science students) $= 15$. Therefore,

$$P(\text{Kofi broke the chair}) = \frac{1}{15}$$

This is usually written as: P(Kofi broke the chair $|$ a Science student has broken the chair) and described as "the probability that Kofi broke the chair given that a Science student has broken the chair". This is an example of a *conditional probability*.

Definition 4.1 CONDITIONAL PROBABILITY
(Conceptual Definition)

Let $\mathcal{A}$ and $\mathcal{B}$ be two events in the same sample space $\mathcal{S}$ with $P(\mathcal{B}) > 0$. The probability assigned to $\mathcal{A}$ or that would be assigned to $\mathcal{A}$ when it is known that $\mathcal{B}$ has already occurred is called the conditional probability of $\mathcal{A}$ given $\mathcal{B}$ and is denoted by $P(\mathcal{A}|\mathcal{B})$

We now introduce a mathematical definition of conditional probability.

Definition 4.2 CONDITIONAL PROBABILITY
(Formal Definition)

Let $\mathcal{A}$ and $\mathcal{B}$ be two events. The conditional probability of $\mathcal{A}$ given $\mathcal{B}$ is given by

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})}, \qquad P(\mathcal{B}) > 0$$

Note

a) $P(\mathcal{A}|\mathcal{B})$ is undefined if $P(\mathcal{B}) = 0$.
b) Instead of using the longer statement given in the formal definition (Definition 4.2), we generally call $P(\mathcal{A}|\mathcal{B})$ "the probability of $\mathcal{A}$ given $\mathcal{B}$".
c) Order is of no significance in the intersection set, since $\mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A}$. This property of intersection yields the following results:

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{B} \cap \mathcal{A})$$

d) In some books the notation $\mathcal{AB}$ is used for $\mathcal{A} \cap \mathcal{B}$. In such books,

$$\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_n$$

may be represented by

$$\mathcal{A}_1\mathcal{A}_2 \cdots \mathcal{A}_n$$

and

$$P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_n)$$

is represented by

$$P(\mathcal{A}_1)P(\mathcal{A}_2) \cdots P(\mathcal{A}_n)$$

The conditional probability satisfies all the axioms required for it to be a probability function (see Definition 3.31).

Let us consider the two extreme cases, namely, when the probability of an event is 0 or 1

a) If $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive, then $\mathcal{A} \cap \mathcal{B} = \emptyset$ and hence

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} = 0$$

This is intuitively clear. If $\mathcal{A}$ and $\mathcal{B}$ are disjoint events, then whenever $\mathcal{B}$ occurs $\mathcal{A}$ cannot occur and so $P(\mathcal{A}|\mathcal{B}) = 0$

b) If $\mathcal{B} \subseteq \mathcal{A}$, then

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} = \frac{P(\mathcal{B})}{P(\mathcal{B})} = 1$$

That is, if $\mathcal{B}$ is contained in $\mathcal{A}$, then whenever $\mathcal{B}$ occurs $\mathcal{A}$ must certainly occur.

c) If $\mathcal{A} \subseteq \mathcal{B}$, then

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} = \frac{P(\mathcal{A})}{P(\mathcal{B})}$$

Example 4.9
A fair die is thrown once.

a) What is the probability that the number obtained is greater than 2?
b) If we are told that the throw resulted in an even number, find the probability that the number is greater than 2.

Solution
The sample space is

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

a) Let $\mathcal{A}$ denote the event "greater than 2", then

$$\mathcal{A} = \{3, 4, 5, 6\}$$

Hence,

$$P(\mathcal{A}) = \frac{4}{6}$$

b) Let $\mathcal{B}$ be the event "even number occurred". Then $\mathcal{B} = \{2, 4, 6\}$.
We are required to calculate $P(\mathcal{A}|\mathcal{B})$.
Now,

$$\begin{aligned} P(\mathcal{B}) &= \frac{3}{6} \\ \mathcal{A} \cap \mathcal{B} &= \{4, 6\} \\ P(\mathcal{A} \cap \mathcal{B}) &= \frac{2}{6} \end{aligned}$$

Hence,

$$P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} = \frac{2/6}{3/6} = \frac{2}{3}$$

4.5.2 GRAPHICAL INTERPRETATION OF CONDITIONAL PROBABILITY

Both probability and area can be thought of as measures of size. As can be seen in Fig. 4.2, $P(\mathcal{A}|\mathcal{B})$, in a certain sense, measures the probability of $\mathcal{A}$ relative to the reduced sample space $\mathcal{B}$. It measures the size of $P(\mathcal{A} \cap \mathcal{B})$ in comparison to $P(\mathcal{B})$.

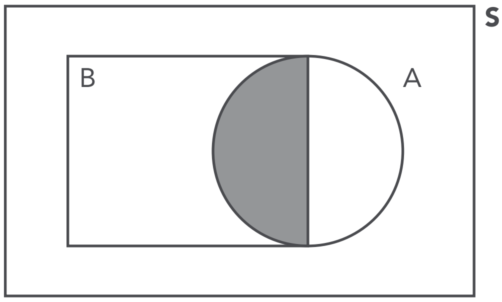


Fig. 4.2 $\mathcal{A} \cap \mathcal{B}$ with $\mathcal{B}$ viewed as a new Sample Space is shaded

Geometrically, the analogous idea would be to compare the area of $\mathcal{A} \cap \mathcal{B}$ to the area of $\mathcal{B}$. In both the probability setting and the geometric figure we are in the same sense measuring what fraction of $\mathcal{B}$ happens to lie also in the set $\mathcal{A}$. Hence, we can view $\mathcal{B}$ as our new sample space, that is, knowing the outcomes in $\mathcal{B}$ reduces our sample space from $\mathcal{S}$ to $\mathcal{B}$.

Theorem 4.8

Let $\mathcal{S}$ be a finite equiprobable sample space with events $\mathcal{A}$ and $\mathcal{B}$. Then

$$P(\mathcal{A}|\mathcal{B}) = \frac{n(\mathcal{A} \cap \mathcal{B})}{n(\mathcal{B})}, \qquad n(\mathcal{B}) > 0$$

Proof
From the relative frequency definition of probability (Definition 3.29),

$$\begin{aligned} P(\mathcal{A} \cap \mathcal{B}) &= \frac{n(\mathcal{A} \cap \mathcal{B})}{n(\mathcal{S})} \\ P(\mathcal{B}) &= \frac{n(\mathcal{B})}{n(\mathcal{S})} \end{aligned}$$

so that

$$\begin{aligned} P(\mathcal{A}|\mathcal{B}) &= \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} \\ &= \frac{n(\mathcal{A} \cap \mathcal{B})}{n(\mathcal{S})} \Big/ \frac{n(\mathcal{B})}{n(\mathcal{S})} \\ &= \frac{n(\mathcal{A} \cap \mathcal{B})}{n(\mathcal{B})} \end{aligned}$$

The intuitive idea behind this theorem using equiprobable measure is that, if we want to find the probability of an event $\mathcal{A}$ and we have no additional information, we must find the size of the sample space $\mathcal{S}$, the size of $\mathcal{A}$ and divide the latter by the former quantity. But suppose we have more information, namely, that another event, $\mathcal{B}$, has definitely occurred. Then we no longer need to count all elements in $\mathcal{S}$, but only those in $\mathcal{B}$. Also, we do not need to count every element in $\mathcal{A}$, but only those in $\mathcal{A} \cap \mathcal{B}$. Hence, with additional information, we can obtain

$$\frac{n(\mathcal{A} \cap \mathcal{B})}{n(\mathcal{B})}$$

This is the conditional probability of $\mathcal{A}$ given $\mathcal{B}$; that is, $P(\mathcal{A}|\mathcal{B})$ is the measure of $\mathcal{A} \cap \mathcal{B}$ divided by the measure of $\mathcal{B}$.

4.5.3 USES OF CONDITIONAL PROBABILITY

The conditional probability plays a major role in probability. It has given birth to rules in probability theory which are of great theoretical and practical importance. These are the multiplication rule, the total probability law and the Bayes' theorem.

4.6 MULTIPLICATION LAW OF PROBABILITY

Strangely enough, one of the easiest and most useful applications of conditional probability is in the "computation of unconditional probabilities" which is the multiplication law of probability.

4.6.1 GENERAL MULTIPLICATION LAW

Theorem 4.9

If $\mathcal{A}$ and $\mathcal{B}$ are two events in the same sample space $\mathcal{S}$, then the probability of the joint occurrence of $\mathcal{A}$ and $\mathcal{B}$ is given by

$$\begin{aligned} P(\mathcal{A} \cap \mathcal{B}) &= P(\mathcal{B})P(\mathcal{A}|\mathcal{B}), \qquad \text{if } P(\mathcal{B}) \neq 0 \\ &= P(\mathcal{A})P(\mathcal{B}|\mathcal{A}), \qquad \text{if } P(\mathcal{A}) \neq 0 \end{aligned}$$

That is, the multiplication rule states that the probability of the simultaneous occurrence of two events equals the product of the probability of the first event and the conditional probability of the second event given that the first event has already occurred.

Proof
The first relation is obtained from the conditional probability formula in Definition 4.2, by cross-multiplying. The second relation follows from the first by interchanging the letters $\mathcal{A}$ and $\mathcal{B}$ and using the fact that $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{B} \cap \mathcal{A})$

Aliter
Refer to Fig 4.2. Let $\mathcal{S}$ have finite number of points. We define $n(\mathcal{A}|\mathcal{B})$ as the number of elements x in $\mathcal{A}$ such that x also belongs to the reduced sample space $\mathcal{B}$. By this definition, $n(\mathcal{A}|\mathcal{B}) = n(\mathcal{A} \cap \mathcal{B})$. Let

$$n(\mathcal{S}) = n, \quad n(\mathcal{B}) = m, \quad n(\mathcal{A}|\mathcal{B}) = k, \quad n(\mathcal{A} \cap \mathcal{B}) = k$$

Then the probability that the event $\mathcal{B}$ will happen is

$$P(\mathcal{B}) = \frac{m}{n}$$

The conditional probability that the event $\mathcal{A}$ will happen given that $\mathcal{B}$ has already happened is

$$P(\mathcal{A}|\mathcal{B}) = \frac{k}{m}$$

The probability that the events $\mathcal{A}$ and $\mathcal{B}$ will happen simultaneously is

$$\begin{aligned} P(\mathcal{A} \cap \mathcal{B}) &= \frac{k}{n} \\ &= \frac{m}{n} \cdot \frac{k}{m} \\ &= P(\mathcal{B})P(\mathcal{A}|\mathcal{B}) \end{aligned}$$

Example 4.10

A box contains 10 balls, of which 6 are red and 4 are blue. If 2 balls are randomly selected from the box without replacement, what is the probability that both are red?

Solution

We may think of the balls as being drawn one at a time. (This just means that we are going to label them "first" and "second". It does not really matter at all whether the balls are drawn one at a time or all together.)

Let $\mathcal{A}$ be the event that the first ball drawn is 'red', and
$\mathcal{B}$, the event that the second is 'red'.

We are required to calculate $P(\mathcal{A} \cap \mathcal{B})$.
The probability that the first ball drawn is red is

$$P(\mathcal{A}) = \frac{6}{10}$$

After the first ball has been drawn, we shall be left with 9 balls from which to draw the second ball. If we know that the first ball drawn is red, then what it means is that there are 5 red balls left in the box, one of which might be drawn second time, so

$$P(\mathcal{B}|\mathcal{A}) = \frac{5}{9}$$

Hence,

$$\begin{aligned} P(\mathcal{A} \cap \mathcal{B}) &= P(\mathcal{A})P(\mathcal{B}|\mathcal{A}) \\ &= \left(\frac{3}{5}\right)\left(\frac{5}{9}\right) = \frac{1}{3} \end{aligned}$$

Corollary 4.5

The probability of the simultaneous occurrence of three events, $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ is given by[15]

$$P(\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}) = P(\mathcal{A})P(\mathcal{B}|\mathcal{A})\,P(\mathcal{C}|\mathcal{A}\cap\mathcal{B})$$

where $P(\mathcal{A}) \neq 0$ and $P(\mathcal{A}\cap\mathcal{B}) \neq 0$

Proof

By the associative law,

$$\begin{aligned}
\mathcal{A}\cap\mathcal{B}\cap\mathcal{C} &= (\mathcal{A}\cap\mathcal{B})\cap\mathcal{C} \\
P(\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}) &= P[(\mathcal{A}\cap\mathcal{B})\cap\mathcal{C}] \\
&= P(\mathcal{A}\cap\mathcal{B})\cap P(\mathcal{C}|\mathcal{A}\cap\mathcal{B}) \\
&= P(\mathcal{B})P(\mathcal{A}|\mathcal{B})\cdot P(\mathcal{C}|\mathcal{A}\cap\mathcal{B})
\end{aligned}$$

Example 4.11

In a consignment of 40 manufactured items, 8 are known to be defective. Suppose three items are drawn at random without replacement. What is the probability that all three in the sample are defective?

Solution

Let $\mathcal{A}_i$ be the event "getting a defective on the i^{th} draw". Then
The probability of selecting a defective item in the 1^{st} draw is:

$$P(\mathcal{A}_1) = \frac{8}{40}$$

The probability of selecting a defective item in the 2^{nd} draw is:

$$P(\mathcal{A}_2|\mathcal{A}_1) = \frac{7}{39}$$

The probability of selecting a defective item in the 3^{rd} draw is:

$$P(\mathcal{A}_3|\mathcal{A}_1\cap\mathcal{A}_2) = \frac{6}{38}$$

Hence, the desired probability is

$$\begin{aligned}
P(\mathcal{A}_1\cap\mathcal{A}_2\cap\mathcal{A}_3) &= P(\mathcal{A}_1)P(\mathcal{A}_2|\mathcal{A}_1)P(\mathcal{A}_3|\mathcal{A}_1\cap\mathcal{A}_2) \\
&= \left(\frac{8}{40}\right)\left(\frac{7}{39}\right)\left(\frac{6}{38}\right) \\
&= \frac{7}{1235}
\end{aligned}$$

This corollary can be extended by mathematical induction as follows:

Corollary 4.6
For any events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n, \quad \mathcal{A}_i > 0$,

$$\begin{aligned} P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_n) &= P(\mathcal{A}_1)P(\mathcal{A}_2|\mathcal{A}_1)P(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2) \\ &\quad \cdots P(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cdots \cap \mathcal{A}_{n-1}) \end{aligned}$$

Theorem 4.9, together with its corollaries, is called the ***general multiplication law (rule)*** or the ***chain rule*** or ***Bayes' sequential formula***.

Using Combinatorial Analysis

The application of the multiplication rule in solving probability problems may sometimes be tedious or confusing. An easier approach is the application of a method of first principles, the combinatorial analysis. We should not allow the clumsiness of the factorials to frighten us. The arithmetic is not too arduous as we cancel the factorials and then use the calculator.

Example 4.12
Refer to Example 4.11.

a) Solve the question using the combinatorial method.
b) Calculate the probability that the sample contains just one defective

Solution
The sample space for this problem is the set of all possible 3-tuples of defective items that could be selected from 40 items so that the sample space consists of $\binom{40}{3}$ equally likely simple events.

a) There are 8 defective items and so 3-tuples of defective items can be selected in $\binom{8}{3}$ number of ways.

$$P(\text{all three are ddfective items}) = \frac{\binom{8}{3}}{\binom{40}{3}} = \frac{7}{1235}$$

b) The sample of 3 items would contain 1 defective and 2 non-defective items. Now, the defective item can be chosen from 8 such ones in $\binom{8}{1}$ ways, and the 2 non-defective items can be drawn from 32 such ones in $\binom{32}{2}$ ways. Hence, the total number of ways of choosing 3 items in which exactly 1 is defective is $\binom{8}{1}\binom{32}{2}$

Hence

$$P(\text{exactly } 1 \text{ defective item}) = \frac{\binom{8}{1}\binom{32}{2}}{\binom{40}{3}} = \frac{496}{1235}$$

The preceding 'combinatorial' solution comes under what is generally called the Hypergeometric Probability Distribution which will be taken up in the next volume.

4.6.2 SPECIAL MULTIPLICATION LAW

Theorem 4.10

If $\mathcal{A}$ and $\mathcal{B}$ are two independent events in the same sample space $\mathcal{S}$, then the probability of the joint occurrence of $\mathcal{A}$ and $\mathcal{B}$ is given by

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$$

Corollary 4.7
For a set of n independent events, $\mathcal{A}_i,\ i = 1, 2, \cdots, n,$ the probability of the joint occurrence of $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ is the product of their individual probabilities:

$$P(\mathcal{A}_1 \cap \mathcal{A}_2 \cap ... \cap \mathcal{A}_n) = P(\mathcal{A}_1)P(\mathcal{A}_2)\cdots P(\mathcal{A}_n),$$

or simply

$$P\left(\bigcap_{i=1}^{n} \mathcal{A}_i\right) = \prod_{i=1}^{n} P(\mathcal{A}_i)$$

Theorem 4.10, together with its corollary, is called *special multiplication rule* because it is only applicable to the case when the events are independent discussed in the sequel.

Example 4.13
Two dice are thrown once. What is the probability that the first die will show a 4 and the second one will show a 6?

Solution
Let
$\mathcal{A}$ be the event that the first die will show a 4;
$\mathcal{B}$ be the event that the second die will show a 6.
Then

$$P(\mathcal{A}) = \frac{1}{6} \quad \text{and} \quad P(\mathcal{B}) = \frac{1}{6}$$

and since the two events are independent,

$$\begin{aligned} P(\mathcal{A} \cap \mathcal{B}) &= P(\mathcal{A})P(\mathcal{B}) \\ &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \end{aligned}$$

4.7 TOTAL PROBABILITY LAW

4.7.1 APPLICATION OF MULTIPLICATION AND ADDITION LAWS

The addition and multiplication laws are the basic tools of probability. The application of these laws may be demonstrated in the total probability law, also called ***the formula of incompatible and exhaustive causes***.

To discuss total probability law, we would need the following theorem.

Theorem 4.11

If $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_n$ form a partition of the sample space $\mathcal{S}$, then for any event $\mathcal{B} \subseteq \mathcal{S}$ and $P(\mathcal{B}) > 0$

$$\begin{aligned} P(\mathcal{B}) &= P(\mathcal{A}_1 \cap \mathcal{B}) + P(\mathcal{A}_2 \cap \mathcal{B}) + \cdots + P(\mathcal{A}_n \cap \mathcal{B}) \\ &= \sum_{i=1}^{n} P(\mathcal{A}_i \cap \mathcal{B}) \end{aligned}$$

Proof

Fig. 4.3 represents a partition of the sample space $\mathcal{S}$:

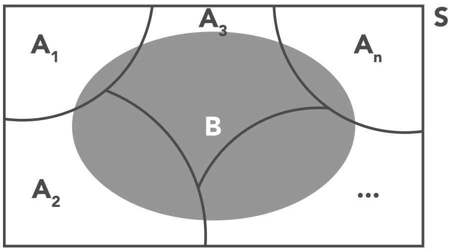


Fig. 4.3 Set $\mathcal{B}$ is shaded

$$\mathcal{S} = \{\mathcal{A}_1,\ \mathcal{A}_2,\ ...,\ \mathcal{A}_n\}$$

By Definition 3.23, the events are pairwise mutually exclusive and their union is $\mathcal{S}$.

From the Fig. 4.3, we see that $\mathcal{B}$ can be expressed as

$$\mathcal{B} = \bigcup_{i=1}^{n} (\mathcal{A}_i \cap \mathcal{B})$$

and since

$$\begin{aligned} \mathcal{A}_i \cap \mathcal{A}_j &= \emptyset, \quad \text{for } \ i \neq j \\ (\mathcal{A}_i \cap \mathcal{B}) \cap (\mathcal{A}_j \cap \mathcal{B}) &= \emptyset, \quad \text{for } \ i \neq j \quad \text{(by Theorem 1.2)} \end{aligned}$$

Therefore, by Corollary 4.3,

$$P(\mathcal{B}) = \sum_{i=1}^{n} P(\mathcal{A}_i \cap \mathcal{B}) \qquad (i)$$

Theorem 4.12 TOTAL PROBABILITY

Suppose $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ is a partition of sample space $\mathcal{S}$ and $\mathcal{B}$ an event defined on the same sample space $\mathcal{S}$ such that $P(\mathcal{B}) > 0$. Then

$$\begin{aligned} P(\mathcal{B}) &= \sum_{i=1}^{n} P(\mathcal{A}_i)P(\mathcal{B}|\mathcal{A}_i) \\ &= P(\mathcal{A}_1)P(\mathcal{B}|\mathcal{A}_1) + P(\mathcal{A}_2)P(\mathcal{B}|\mathcal{A}_2) + ... + P(\mathcal{A}_n)P(\mathcal{B}|\mathcal{A}_n) \end{aligned}$$

Proof
From Theorem 4.9,

$$P(\mathcal{A}_i \cap \mathcal{B}) = P(\mathcal{A}_i)P(\mathcal{B}|\mathcal{A}_i)$$

and from Theorem 4.11,

$$P(\mathcal{B}) = P(\mathcal{A}_i \cap \mathcal{B})$$

from which the result follows.

Example 4.14
A group of visitors to the University of Ghana consisted of 15 students from the University of Oxford and 20 students from the University of Ibadan. Among the students from the University of Oxford were 8 females and among the students from the University of Ibadan were 5 females. A student was selected (at random) to give a vote of thanks at the end of the visit. What is the probability that the student is a female?

Solution
Let

$$\begin{aligned} \mathcal{H}_1 &= \{\text{a student is from University of Oxford}\} \\ \mathcal{H}_2 &= \{\text{a student is from University of Ibadan}\} \\ F &= \{\text{a female student}\} \end{aligned}$$

Then either the female student came from University of Oxford and was a female or she came from University of Ibadan and was a female. This is the union of two disjoint events which are $\mathcal{H}_1 \cap \mathcal{F}$ and $\mathcal{H}_2 \cap \mathcal{F}$. Hence

$$\begin{aligned} P(\mathcal{F}) &= P(\mathcal{H}_1)P(\mathcal{F}|\mathcal{H}_1) + P(\mathcal{H}_2)P(\mathcal{F}|\mathcal{H}_2) \\ &= \left(\frac{15}{35}\right)\left(\frac{8}{15}\right) + \left(\frac{20}{35}\right)\left(\frac{5}{20}\right) \\ &= \frac{13}{35} \end{aligned}$$

Directly:
Number of female students from both Universities $\mathcal{A}_1$ and $\mathcal{A}_2$ is:

$$\begin{aligned} 8 \text{ females from } \mathcal{H}_1 + 5 \text{ females from } \mathcal{H}_2 &= 13 \text{ female students from both universities} \\ \text{Total number of students} &= 35 \end{aligned}$$

Hence,

$$P(\text{female}) = \frac{13}{35}$$

4.7.2 PROBABILITY TREE DIAGRAM

Many simple problems that can be solved with total probability can also be solved with the probability tree diagrams. In fact, the tree diagram can be extremely useful in more complex problems.

A probability tree diagram is a form of graphical display which combines the addition and multiplication laws. It enables us to

a) determine the appropriate sample space of an experiment;
b) visually see all the possible events;
c) follow all the "what ifs";
d) calculate total probability; and
e) calculate the probability with Bayes' Theorem, discussed in the sequel.

In the probability diagram, the first event is represented by a dot from where branches are drawn to represent all possible outcomes of the event. The first level of branching corresponds to the hypothesis (prior) and the second refers to the event outcome (condition). The probability of each outcome is written on its branch where the ones on the first level of the branches are the prior probabilities and those on the second level are conditional probabilities. In fact, the tree diagram will show more information than is required to answer the question being asked. The probability that any particular path of the tree occurs is, by the multiplication rule, the product of the probabilities written on its segments (that is, the path of the branch from the root to the end of the path). The probability of any outcome is found by adding the probabilities of all branches that are part of that event. Fig. 4.4 shows an example of a probability tree diagram.

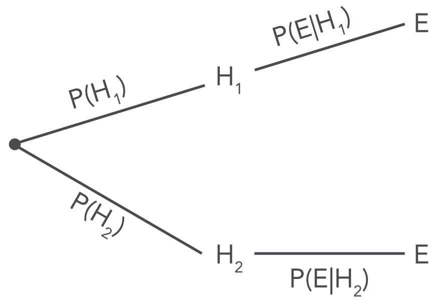


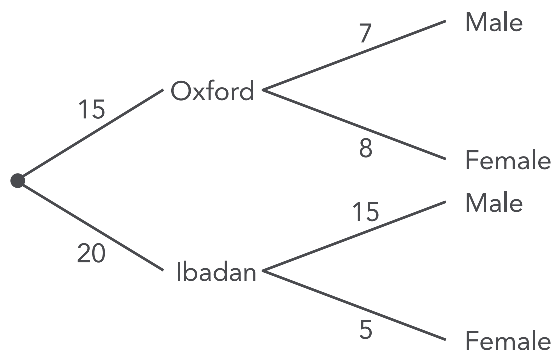
Fig. 4.4 Probability Tree Diagram for Computing Total Probability

A good understanding of the tree diagram will give us the ability to "reverse" the process or compute reverse probabilities discussed in the following section. Thomas Bayes created a formula for this revere process which has come to be called the Bayes' Theorem.

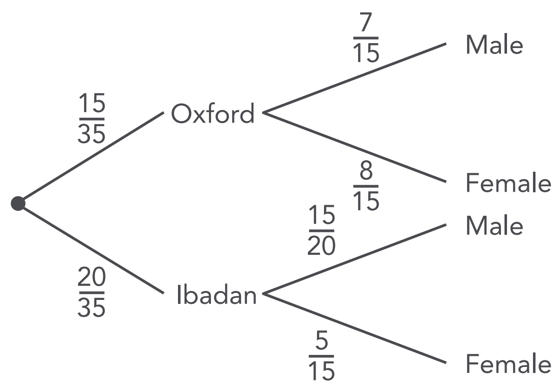
Example 4.15
Use the probability tree diagram to solve Example 4.14.

Solution
In this example, the hypothesis is that the student came from either the University of Oxford or the University of Ibadan. The second branching refers to the event outcome which is either a male or a female.

The following tree diagram can be constructed.



From this tree diagram, we obtain the probability tree diagram by dividing the values on the branch by their corresponding totals. The result is presented below:



The selected female student could come from either the University of Oxford or the University of Ibadan.

Now from the diagram, the probability of selecting a female student given that she came from the University of Oxford is to follow the University of Oxford's branch through to the female branch and multiply the probabilities. This gives:

$$P(\mathcal{F}|\mathcal{H}_1) = \left(\frac{15}{35}\right)\left(\frac{8}{15}\right) = \frac{8}{35}$$

Similarly, the probability of selecting a female student given that she came from the University of Ibadan is to follow the University of Ibadan's branch through to the female branch and multiply the probabilities. This gives:

$$P(\mathcal{F}|\mathcal{H}_2) = \left(\frac{20}{35}\right)\left(\frac{5}{20}\right) = \frac{5}{35}$$

Hence the probability of selecting a female student is the probability that she came either from the University of Oxford or from the University of Ibadan. That is:

$$P(\text{Female student}) = \frac{8}{35} + \frac{5}{35} = \frac{13}{35}$$

4.7.3 USING TABLES IN COMPUTING TOTAL PROBABILITY

The calculations of the total probability and the Bayes' Theorem are complicated enough to create an abundance of opportunities for errors and/or incorrect substitution of the involved probabilities. Fortunately, they can be presented in a tabular form.

Let

$P(\mathcal{H}_i)$ be the probability of hypothesis $\mathcal{H}_i$;
$P(\mathcal{B}|\mathcal{H}_i)$ be the probability of event $\mathcal{B}$ given hypothesis $\mathcal{H}_i$;

These probabilities are presented in Table 4.1.

Then, the probability of the event outcome $\mathcal{B}$ is the sum of column 4. This value is the total probability.

Hypothesis $\mathcal{H}_i$	$P(\mathcal{H}_i)$	$P(\mathcal{B}\|\mathcal{H}_i)$	$P(\mathcal{H}_i)P(\mathcal{B}\|\mathcal{H}_i)$
1	2	3	$4 = (2) \times (3)$
$\mathcal{H}_1$	$P(\mathcal{H}_1)$	$P(\mathcal{B}\|\mathcal{H}_1)$	$P(\mathcal{H}_1)P(\mathcal{B}\|\mathcal{H}_1)$
$\mathcal{H}_2$	$P(\mathcal{H}_2)$	$P(\mathcal{B}\|\mathcal{H}_2)$	$P(\mathcal{H}_2)P(\mathcal{B}\|\mathcal{H}_2)$
Total	1		$P(\mathcal{H}_1)P(\mathcal{B}\|\mathcal{H}_1) + P(\mathcal{H}_2)P(\mathcal{B}\|\mathcal{H}_2)$

TABLE 4.1 Computation of Total Probability in Tabular Form

Example 4.16

Present the probability problem in Example 4.13 in a tabular form and find the probability that a student selected is a female.

Solution

Let $\mathcal{M}$ and $\mathcal{F}$ be males and females respectively. The problem may be summarised as in the table below.

University	$\mathcal{M}$	$\mathcal{F}$	Total
Oxford	7	8	15
Ibadan	15	5	20

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be the hypotheses that the student selected is from the University of Oxford and the University of Ibadan, respectively. Then, the results can be summarised in the table below.

The probability of selecting a female is the probability of selecting a female from either the University of Oxford ($\frac{8}{35}$) or the University of Ibadan ($\frac{5}{35}$). The total probability is, therefore,

$$P(\text{Female}) = \frac{8}{35} + \frac{5}{35} = \frac{13}{35}$$

which is what is shown in the last row of the column of the table below.

University i	Hypothesis $\mathcal{H}_i$	$P(\mathcal{H}_i)$	$P(\mathcal{F}\vert\mathcal{H}_i)$	$P(\mathcal{H}_i) \times P(\mathcal{F}\vert\mathcal{H}_i)$
1	2	3	4	$5 = (3) \times (4)$
Oxford	$\mathcal{H}_1$	$15/35$	$8/15$	$15/35 \times 8/15 = 8/35$
Ibadan	$\mathcal{H}_2$	$20/35$	$5/20$	$20/35 \times 5/20 = 5/35$
Total		1	$47/60$	$13/35$

4.8 BAYES' THEOREMS

4.8.1 CONCEPT OF BAYES' THEOREM

In Example 4.14, a student was selected and we computed the probability that the student was a female. As noticed earlier, that female could come from either the University of Oxford or the University of Ibadan. Recollect that it was given as:

$$P(\text{Female}) = P(\text{Oxford})P(\text{Female}|\text{Oxford}) + P(\text{Ibadan})P(\text{Female}|\text{Ibadan})$$

Suppose that we now have an additional information that a female student was selected and we are interested in computing the probability that she was from Oxford, that is, $P(\text{Oxford}|\text{Female})$.

This process of reversing the order of the conditioning is called the *Bayes' Theorem* or *Bayes' rule* or *Bayes' law*, due to the English theologian and Mathematician, Reverend Thomas Bayes (1702–1761). The Bayes' Theorem describes the probability of an event, based on conditions that are related to the event.

Definition 4.3 BAYES' THEOREM

Suppose there is a hypothesis $\mathcal{H}$ and some observed evidence $\mathcal{E}$, then Bayes' Theorem is the relationship between the probability of the hypothesis after getting the evidence $P(\mathcal{H}|\mathcal{E})$ and the probability of the hypothesis before getting the evidence $P(\mathcal{H})$:

$$P(\mathcal{H}|\mathcal{E}) = P(\mathcal{H})\frac{P(\mathcal{E}|\mathcal{H})}{P(\mathcal{E})}$$

In Definition 4.3,

$P(\mathcal{H})$ is known as the ***prior probability*** and it is the (initial) probability of the event originally obtained before any new or additional evidence or information is collected;

$P(\mathcal{H}|\mathcal{E})$ is referred to as the ***posterior probability*** and it is the probability of the event that has been revised after new or additional evidence or information has been collected;

$\dfrac{P(\mathcal{E}|\mathcal{H})}{P(\mathcal{E})}$ is the factor that relates the posterior and the prior probabilities and is called the ***likelihood ratio***.

Thus, Bayes' theorem can be stated as:
the posterior probability equals the prior probability times the likelihood ratio.

Theorem 4.13 BAYES' THEOREM

Suppose $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_n$ are mutually exclusive and collectively exhaustive events in sample space $\mathcal{S}$ with prior probabilities $P(\mathcal{A}_i)$ such that $P(\mathcal{A}_i) \neq 0$ for $i = 1, 2, \cdots, n$. Let $\mathcal{B}$ be any event in $\mathcal{S}$ for which conditional probabilities of $\mathcal{B}$ given $\mathcal{A}_i$ are $P(\mathcal{B}|\mathcal{A}_i)$ with $P(\mathrm{B}) \neq 0$. Assume that $P(\mathcal{A}_i)$ and $P(\mathcal{B}|\mathcal{A}_i)$ for $i = 1, 2, \cdots, n$ are known, then the posterior probability of $\mathcal{A}_i$ given that $\mathcal{B}$ has occurred is

$$P(\mathcal{A}_i|\mathcal{B}) = \frac{P(\mathcal{A}_i)P(\mathcal{B}|\mathcal{A}_i)}{\sum_{i=1}^{n} P(\mathcal{A}_i)P(\mathcal{B}|\mathcal{A}_i)}$$

Proof
From the definition of conditional probability (Definition 4.2),

$$P(\mathcal{A}_i|\mathcal{B}) = \frac{P(\mathcal{A}_i \cap \mathcal{B})}{P(\mathcal{B})}$$

We now replace the numerator by its equivalent expression in Theorem 4.9 and the denominator by its equivalent expression in Theorem 4.12 to obtain:

$$P(\mathcal{A}_i|\mathcal{B}) = \frac{P(\mathcal{A}_i)P(\mathcal{B}|\mathcal{A}_i)}{\sum_{i=1}^{n} P(\mathcal{A}_i)P(\mathcal{B}|\mathcal{A}_i)}$$

In the Bayes' Theorem, $\mathcal{A}_i$ is a hypothesis (causes) and $\mathcal{B}$ an event based on this hypothesis (consequences). Hence,

$P(\mathcal{A}_i|\mathcal{B})$ = the probability of the hypothesis $\mathcal{A}_i$, given the occurrence of event $\mathcal{B}$;

$P(\mathcal{A}_i|\mathcal{B})$ = the probability of the event $\mathcal{B}$, given the occurrence of hypothesis $\mathcal{A}_i$.

Bayes' Theorem is applicable in situations where quantities of the form $P(\mathcal{B}|\mathcal{A}_i)$ and $P(\mathcal{A}_i)$ are known and we wish to determine $P(\mathcal{A}_i|\mathcal{B})$ The probabilities, $P(\mathcal{A}_i)$ are the ***prior probabilities*** of $\mathcal{A}_i$, $P(\mathcal{A}_i|\mathcal{B})$ are the ***posterior probabilities*** of $\mathcal{A}_i$, given that the event $\mathcal{B}$ has occurred and $P(\mathcal{B}|\mathcal{A}_i)$ are the ***likelihoods***.

The Bayes' Theorem enables us to find the probabilities of the various events $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ which could cause $\mathcal{A}$ to occur, given the consequence $\mathcal{B}$. For this reason, Bayes' Theorem is often referred to as a ***theorem on the probability of causes*** or ***Bayes' retrodiction formula***.

Note

The terms on the right-hand side of the Bayes' Theorem (Theorem 4.13) are all conditioned on the events $\mathcal{A}_i$ (hypothesis), while that on the left-hand side is conditioned on $\mathcal{B}$ (event outcome).

Example 4.17
Suppose, in Example 4.14, the student selected is a female. What is the probability that the student came from

a) University of Oxford?
b) University of Ibadan?

Solution
All the probabilities have been calculated in Example 4.14. We use the notations in that example.

a) The probability that a female student selected came from Oxford is

$$\begin{aligned} P(\mathcal{H}_1|\mathcal{F}) &= \frac{P(\mathcal{H}_1)P(\mathcal{F}|\mathcal{H}_1)}{P(\mathcal{H}_1)P(\mathcal{F}|\mathcal{H}_1) + P(\mathcal{H}_1)P(\mathcal{F}|\mathcal{H}_1)} \\ &= \frac{(15/35)(8/15)}{(15/35)(8/15) + (20/35)(5/20)} = \frac{8}{13} \end{aligned}$$

b) The probability that a female student selected came from Ibadan is

$$\begin{aligned} P(\mathcal{H}_2|\mathcal{F}) &= \frac{P(\mathcal{H}_2)P(\mathcal{F}|\mathcal{H}_2)}{P(\mathcal{H}_2)P(\mathcal{F}|\mathcal{H}_2) + P(\mathcal{H}_2)P(\mathcal{F}|\mathcal{H}_2)} \\ &= \frac{(20/35)(5/20)}{(15/35)(8/15) + (20/35)(5/20)} = \frac{5}{13} \end{aligned}$$

Note

If the only information concerning the group of visitors consisted of the fact that 13 were females and 22 were males, and we asked "What is the chance that a student selected at random to give the vote of thanks is a female?" we should reply, "The probability is $\frac{13}{35}$". However, if we increase our knowledge by means of the additional fact that the randomly selected student was found to be from University of Oxford, then our probability value rises from $\frac{13}{35}$ to $\frac{8}{13}$. On the other hand, if we increase our knowledge by means of the additional fact that the randomly selected student was found to be from University of Ibadan, then our probability value also rises from $\frac{13}{35}$ to $\frac{5}{13}$.

This illustrates the reason for the assertion that "If our probability is a measure of the importance of our state of ignorance it must change its value whenever we add new knowledge".

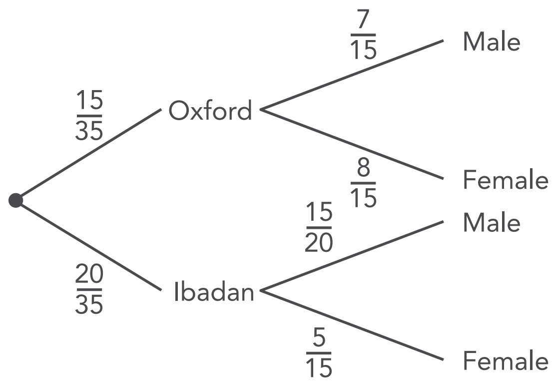
4.8.2 PROBABILITY TREE DIAGRAM FOR SOLVING PROBLEMS RELATED TO BAYES' THEOREM

The probability tree diagram so constructed in Fig. 4.4 can be used to compute posterior probabilities, that is, in solving problems relating to the Bayes' Theorem. The posterior probability of any outcome is found by multiplying the probabilities of all branches that are part of that event and dividing it by the total probability related to that event.

Example 4.18
Use the probability tree diagram to solve Example 4.17.

Solution

For convenience, we reproduce the diagram in Example 4.15 here.



a) From the diagram, knowing that a female has been selected, the probability that she came from University of Oxford is to

i) Follow the Oxford branch through to the female branch and multiply their probabilities:

$$\left(\frac{15}{35}\right)\left(\frac{8}{15}\right)=\frac{8}{35}$$

ii) Compute the total probability for female branches for both Universities:

$$P(\mathcal{F})=\left(\frac{15}{35}\right)\left(\frac{8}{15}\right)+\left(\frac{20}{35}\right)\left(\frac{5}{20}\right)=\frac{13}{35}$$

iii) Divide the result in (i) by that in (ii):

$$P(\mathcal{H}_1|\mathcal{F})=\frac{8/35}{13/35}=\frac{8}{13}$$

which is the same as the results in Example 4.17.

b) We ask the reader to find out that the probability that the female student came from the University of Ibadan.

4.8.3 USING TABLES IN SOLVING PROBLEMS ASSOCIATED WITH BAYES' THEOREM

A more elaborate computation of problems in Bayes' Theorem can be obtained by presenting the probabilities in a tabular form, similar to the one for total probability in Table 4.1. Here, Table 4.1 is extended to include the posterior probabilities (Table 4.2).

We advise readers to adopt this type of table in solving problems relating to Bayes' Theorem.

Hypothesis	Prior Probability	Likelihood	Joint Probability	Posterior Probability
$\mathcal{H}_i$	$P(\mathcal{H}_i)$	$P(\mathcal{A}\vert\mathcal{H}_i)$	$P(\mathcal{H}_i \cap \mathcal{A})$	$P(\mathcal{H}_i\vert\mathcal{A})$
1	2	3	$4 = (2) \times (3)$	$5{=}(4)/P(\mathcal{T})$
$\mathcal{H}_1$	$P(\mathcal{H}_1)$	$P(\mathcal{A}\vert\mathcal{H}_1)$	$P(\mathcal{H}_1)P(\mathcal{A}\vert\mathcal{H}_1)$	$P(\mathcal{H}_1)P(\mathcal{A}\vert\mathcal{H}_1)/P(\mathcal{T})$
$\mathcal{H}_2$	$P(\mathcal{H}_2)$	$P(\mathcal{A}\vert\mathcal{H}_2)$	$P(\mathcal{H}_2)P(\mathcal{A}\vert\mathcal{H}_2)$	$P(\mathcal{H}_2)P(\mathcal{A}\vert\mathcal{H}_2)/P(\mathcal{T})$
Total	1		$P(\mathcal{T})$	1

TABLE 4.2 Computation of Total Probability in Tabular Form[16]

Example 4.19
Rework Example 4.18 using Table 4.2.

Solution
The following table is an extension of the table in Example 4.15 by adding a column for the posterior probabilities:

University	Hypothesis	Prior	Likelihood	Joint	Posterior
i	$\mathcal{H}_i$	$P(\mathcal{H}_i)$	$P(\mathcal{F}\vert\mathcal{H}_i)$	$P(\mathcal{H}_i \cap \mathcal{F})$	$P(\mathcal{H}_i\vert\mathcal{F})$
(1)	(2)	(3)	(4)	$(5) = (3) \times (4)$	$(6) = (5)/P(\mathcal{T})$
Oxford	$\mathcal{H}_1$	$15/35$	$8/15$	$8/35$	$8/13$
Ibadan	$\mathcal{H}_2$	$20/35$	$5/20$	$5/35$	$5/13$
Total				$13/35$	1

From the last column of the table, the probability that a female student selected came from Oxford is $\frac{8}{13}$ which is the same as in Example 4.17.

4.9 STATISTICAL INDEPENDENCE

4.9.1 DEFINITION OF STATISTICAL INDEPENDENCE

We have discussed the concept of mutually exclusive events, that is, the situation when events $\mathcal{A}$ and $\mathcal{B}$ cannot occur together. We have noted earlier that if $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive, then $P(\mathcal{A}|\mathcal{B}) = 0$, since the given occurrence of $\mathcal{B}$ precludes the occurrence of $\mathcal{A}$. The other extreme situation is when $\mathcal{B} \subseteq \mathcal{A}$ in which case $P(\mathcal{A}|\mathcal{B}) = 1$. In both situations, we noted that if we know that $\mathcal{B}$ has occurred then we have very definite information about the probability of the occurrence of $\mathcal{A}$. However, there are many situations when the knowledge of the occurrence of some event $\mathcal{B}$ does not have any bearing whatsoever on the occurrence or non-occurrence of $\mathcal{A}$.

> **Definition 4.4** INDEPENDENT EVENTS
> **(Conceptual Definition)**
>
> Two events A and B are said to be independent if the occurrence of one does not affect the probability that the other occurs

Frequently, we will postulate that two events are independent or it will be clear from the nature of the experiment that the two events are independent.

Example 4.20
Throw two dice once. Is the probability of obtaining a six (or any other point) on the first die independent of obtaining a six (or any other point) on the second die?

Solution
Our intuition says the two dice should not exercise any control over each other; that is, the outcome on one die should be "independent" of the outcome on the other.

This shows that $\mathcal{A}$ and $\mathcal{B}$ are independent according to the condition, just as our intuition would indicate.

Note

One cannot prove mathematically whether two real-world dice behave independently or not. All we are illustrating is that if the sample space described in Example 4.20 is used to model a pair of dice, with the assumption that the outcomes are equally likely, then the theoretical dice of the model behave independently. In other cases we need to show mathematically that the events are independent.

Definition 4.5 STATISTICAL INDEPENDENCE
(Formal Definition)

Events $\mathcal{A}$ and $\mathcal{B}$ on the same sample space $\mathcal{S}$ are said to be (statistically) independent if the probability of the joint occurrence of $\mathcal{A}$ and $\mathcal{B}$ is equal to the product of their individual probabilities.

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$$

Note

If both $P(A) > 0$ and $P(B) > 0$ and if A and B are mutually exclusive, then the two events cannot be independent (that is, they are dependent). In such a case, the left-hand side of Definition 4.5 is zero, but the right-hand side is positive.

Example 4.21
A box contains ten numbered identical balls. A ball is picked at random with replacement from the box. Consider the events:

$$A = \{x | x \leq 4\}, \quad B = \{x | x \text{ is even}\}$$

Are the two events $\mathcal{A}$ and $\mathcal{B}$ statistically independent?

Solution

$$\begin{aligned}
\mathcal{S} &= \{1,2,3,4,5,6,7,8,9,10\} \\
\mathcal{A} &= \{1,2,3,4\} \\
\mathcal{B} &= \{2,4,6,8,10\} \\
\mathcal{A}\cap\mathcal{B} &= \{2,4\}
\end{aligned}$$

So

$$\begin{aligned}
P(\mathcal{A}\cap\mathcal{B}) &= \frac{2}{10} \\
P(\mathcal{A}) &= \frac{4}{10} \\
P(\mathcal{B}) &= \frac{5}{10} \\
P(\mathcal{A})P(\mathcal{B}) &= \frac{2}{5}\cdot\frac{5}{10} = \frac{2}{10} = P(\mathcal{A}\cap\mathcal{B})
\end{aligned}$$

Hence $\mathcal{A}$ and $\mathcal{B}$ are statistically independent.

For more than two events, verifying independence is slightly complicated.

Definition 4.6

For $n > 2$, events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ are said to be mutually independent if and only if the probability of the intersection of any $2, 3, \cdots, n$ of these events is equal to the product of their respective probabilities

From Chapter 2, we noted that

$$\sum_{r=0}^{n}\binom{n}{r} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$$

It follows, therefore, that there are $2^n - n - 1$ of such subcollections that should all be mutually independent for independence of the n events to be established.

Theorem 4.14

Three events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are independent if

a) they are pairwise independent, and
b) $P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) = P(\mathcal{A})P(\mathcal{B})P(\mathcal{C})$

That is, three events $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are independent if and only if all the four $(2^3 - 3 - 1 = 4)$ of the following conditions are satisfied:

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}), \quad P(\mathcal{A} \cap \mathcal{C}) = P(\mathcal{A})P(\mathcal{C}),$$
$$P(\mathcal{B} \cap \mathcal{C}) = P(\mathcal{B})P(\mathcal{C}), \quad P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) = P(\mathcal{A})P(\mathcal{B})P(\mathcal{C})$$

Example 4.22
Consider the sample space

$$\mathcal{S} = \{x | x \leq 10\}$$

and the events:

$$A = \{x|x \leq 4\}, \quad B = \{x|x \text{ is even}\}, \quad C = \left\{x|x = \frac{n(n+1)}{2}\right\}$$

Verify whether all the three events are independent.

Solution
$\mathcal{A} = \{1,2,3,4\}; \quad \mathcal{B} = \{2,4,6,8,10\} \quad \mathcal{C} = \{1,3,6,10\}$
$\mathcal{A} \cap \mathcal{B} = \{2,4\}, \quad \mathcal{A} \cap \mathcal{C} = \{1,3\}, \quad \mathcal{B} \cap \mathcal{C} = \{6,10\}, \quad \mathcal{A} \cap \mathcal{B} \cap \mathcal{C} = \emptyset$

Then

$$\begin{aligned} P(\mathcal{A} \cap \mathcal{B}) &= \frac{2}{10} \\ P(\mathcal{A} \cap \mathcal{C}) &= \frac{2}{10} \\ P(\mathcal{B} \cap \mathcal{C}) &= \frac{2}{10} \\ P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) &= 0 \end{aligned}$$

But

$$P(\mathcal{A})P(\mathcal{B})P(\mathcal{C}) = \frac{4}{10} \times \frac{5}{10} \times \frac{4}{10} \neq 0$$

Hence $\mathcal{A}$ and $\mathcal{B}$ are not independent.

Note

a) Three events can be pairwise independent even though they are not statistically independent. Example 4.22 illustrates this.
b) If $P(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) = P(\mathcal{A})P(\mathcal{B})P(\mathcal{C})$, it does not necessarily imply that the events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are pairwise independent

Corollary 4.8
Events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ are jointly independent (mutually independent) if and only if, for any subcollection of k events $(k \leq n)$, $\mathcal{A}_{i1}, \mathcal{A}_{i2}, \cdots, \mathcal{A}_{in}$,

$$P\left(\bigcup_{j=1}^{k}\right) = P\left(\prod_{j=1}^{k} \mathcal{A}_{ij}\right)$$

In general, $n > 2$ events can be pairwise independent but it does not imply necessarily that they are statistically independent.

Theorem 4.15

If events $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are independent then $\mathcal{C}$ and $\mathcal{A}\cap\mathcal{B}$ are independent

Proof
By Definition 4.5,

$$P(\mathcal{A}\cap\mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$$

and by Theorem 4.14,

$$P(\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}) = P(\mathcal{A})P(\mathcal{B})P(\mathcal{C}) \qquad (i)$$

Eq. (i) may be written as

$$P(\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}) = P(\mathcal{A}\cap\mathcal{B})P(\mathcal{C}) \qquad (ii)$$

Let

$$Z = P(\mathcal{A}\cap\mathcal{B})$$

Then Eq. (ii) can be written as

$$P(\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}) = P(Z)P(\mathcal{C})$$

so $P(Z) = (\mathcal{A}\cap\mathcal{B})$ and $P(\mathcal{C})$ are independent.

Note

a) If $\mathcal{A}$ is independent of $\mathcal{B}$ and $\mathcal{A}$ is independent of $\mathcal{C}$ then $\mathcal{B}$ is not necessarily independent of $\mathcal{C}$.
b) If $\mathcal{A}$ is independent of $\mathcal{B}$ and $\mathcal{A}$ is independent of $\mathcal{C}$, then $\mathcal{A}$ is not necessarily independent of $\mathcal{B}\cup\mathcal{C}$.

Independence as defined in Definition 4.5 means that independence would not be defined if $P(\mathcal{B}) = 0$ since $P(\mathcal{A}|\mathcal{B})$ is defined only when $P(\mathcal{B}) \neq 0$.

Another way of establishing independence is given in Theorem 4.16.

Theorem 4.16

Suppose $\mathcal{A}$ and $\mathcal{B}$ are events defined on the same sample space $\mathcal{S}$. If they are independent, then

$$P(\mathcal{A}|\mathcal{B}) = P(\mathcal{A})$$

Proof
From Definition 4.2,

$$\begin{aligned} P(\mathcal{A}|\mathcal{B}) &= \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} \\ &= \frac{P(\mathcal{A})P(\mathcal{B})}{P(\mathcal{B})} \quad \text{(by Definition 4.5)} \\ &= P(\mathcal{A}) \end{aligned}$$

Similarly

$$P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B})$$

Theorem 4.16 proves to be helpful because

a) it gives a better expression of the intuitive idea of independence as provided in Definition 4.4;
b) it provides an easy way to compute the probability of both $\mathcal{A}$ and $\mathcal{B}$ occurring if the individual probabilities of $\mathcal{A}$ and $\mathcal{B}$ are known and if it is also known that $\mathcal{A}$ and $\mathcal{B}$ are independent.

Example 4.23
Suppose $\mathcal{A}$ and $\mathcal{B}$ are independent event. If $P(\mathcal{A}) = \frac{2}{5}$ and $P(\mathcal{B}) = \frac{2}{3}$, calculate (a) $P(\mathcal{A}|\mathcal{B})$ and (b) $P(\mathcal{B}|\mathcal{A})$.

Solution
Since the two events are independent,

i) By Theorem 4.16,

$$P(\mathcal{A}|\mathcal{B}) = P(\mathcal{A}) = \frac{2}{5}$$

ii) By Theorem 4.16,

$$P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B}) = \frac{2}{3}$$

c) it is applicable even when $P(\mathcal{B}) = 0$ in determining $P(A|B)$, and it is for this one reason why virtually all works in the field of probability and statistics rather make use of Theorem 4.9 and Corollary 4.6[17].

Example 4.24
The Regent University College of Science and Technology wanted to employ staff to participate in the teaching of a newly introduced programme. The following table gives the proportion of the candidates who were selected for employment[18]. Are being a PhD holder and a male independent?

Qualification	Male	Female	Total
PhD	0.38	0.20	0.58
Masters	0.25	0.17	0.42
Total	0.63	0.37	1.00

Solution

Let
$\mathcal{A}$ be the event that a PhD holder is employed;
$\mathcal{B}$ be the event that a male is employed.

From the table

$$\begin{aligned} P(\mathcal{A}) &= 0.58 \\ P(\mathcal{B}) &= 0.63 \\ P(\mathcal{A} \cap \mathcal{B}) &= 0.38 \end{aligned}$$

Hence,

$$\begin{aligned} P(\mathcal{B}|\mathcal{A}) &= \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{A})} \\ &= \frac{0.38}{0.58} = 0.66 \neq P(\mathcal{A}) \\ P(\mathcal{A}|\mathcal{B}) &= \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})} \\ &= \frac{0.38}{0.0.63} = 0.60 \neq P(\mathcal{B}) \end{aligned}$$

Therefore, the events $\mathcal{A}$ and $\mathcal{B}$ are not independent.

Theorem 4.17

If two events $\mathcal{A}$ and $\mathcal{B}$ defined on the same sample space $\mathcal{S}$ are statistically independent, then

a) $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$ are independent;
b) $\mathcal{A}$ and $\overline{\mathcal{B}}$ are independent;
c) $\overline{\mathcal{A}}$ and $\mathcal{B}$ are independent.

Proof
We shall prove for the case of the pair $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$.

$$\begin{aligned}
P(\overline{\mathcal{A}} \cap \overline{\mathcal{B}}) &= P(\overline{\mathcal{A} \cup \mathcal{B}}) \\
&= 1 - P(\mathcal{A} \cup \mathcal{B}) \\
&= 1 - [P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})] \\
&= 1 - P(\mathcal{A}) - P(\mathcal{B}) + P(\mathcal{A})P(\mathcal{B}) \\
&= [1 - P(\mathcal{A})] - P(\mathcal{B})[1 - P(\mathcal{A})] \\
&= [1 - P(\mathcal{A})][1 - P(\mathcal{B})] \\
&= P(\overline{\mathcal{A}})P(\overline{\mathcal{B}})
\end{aligned}$$

This proof shows that if two events are independent, then their complements are also independent.

The other two cases of Theorem 4.15 say that if two events are independent, then each event is independent of the complement of the other. The reader is asked to prove them in Exercise 4.33.

Corollary 4.9
If events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ defined on the same sample space $\mathcal{S}$ are statistically independent, then $\overline{\mathcal{A}}_1, \overline{\mathcal{A}}_2, \cdots, \overline{\mathcal{A}}_n$ are independent events.

This corollary can be proved by induction.

4.9.2 PROBABILITY OF OCCURRENCE OF "AT LEAST ONE" EVENT

One of the importance of statistical independence in probability is associated with the calculation of the probability of the occurrence of *at least one* of the events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ when the probabilities of these events are given.

Theorem 4.18

Let $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ be independent events and $\overline{\mathcal{A}}_1, \overline{\mathcal{A}}_2, \cdots, \overline{\mathcal{A}}_n$ be their corresponding complementary events. Then

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n) = 1 - P(\overline{\mathcal{A}}_1)P(\overline{\mathcal{A}}_2)\cdots P(\overline{\mathcal{A}}_n)$$

Proof
By Corollary 4.9, the events $\overline{\mathcal{A}}_1, \overline{\mathcal{A}}_2, \cdots, \overline{\mathcal{A}}_n$ are independent.

Hence

$$P(\overline{\mathcal{A}}_1 \cap \overline{\mathcal{A}}_2 \cap \cdots \cap \overline{\mathcal{A}}_n) = P(\overline{\mathcal{A}}_1)P(\overline{\mathcal{A}}_2)\cdots P(\overline{\mathcal{A}}_n)$$

By Theorem 4.5

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cdots \cup \mathcal{A}_n) + P(\overline{\mathcal{A}}_1 \cap \overline{\mathcal{A}}_2 \cap \cdots \cap \overline{\mathcal{A}}_n) = 1$$

so that

$$\begin{aligned} P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n) &= 1 - P(\overline{\mathcal{A}}_1 \cap \overline{\mathcal{A}}_2 \cap \cdots \cap \overline{\mathcal{A}}_n) \\ &= 1 - P(\overline{\mathcal{A}}_1)P(\overline{\mathcal{A}}_2)\cdots P(\overline{\mathcal{A}}_n) \end{aligned}$$

Corollary 4.10
If all the events $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ have the same probability p then

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n) = 1 - (1-p)^n = 1 - q^n$$

where $q = 1 - p$

Example 4.25
Refer to Example 4.5. Use Theorem 4.18 and its corollary to calculate the probabilities.

Solution

a) Let $\mathcal{A}_i$ $(i = 1, 2, \cdots, 6)$ denote the event that the athlete i wins a Gold medal and $P(\mathcal{A}_i)$ its corresponding probability. Then the probability that the athletes i does not win a Gold medal is

$$q = P(\overline{\mathcal{A}}_1) = P(\overline{\mathcal{A}}_2) = \cdots = P(\overline{\mathcal{A}}_6) = 1 - \frac{2}{5} = \frac{3}{5}$$

By Corollary 4.10, the probability that at least one of the athletes wins a Gold medal is

$$P(\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_6) = 1 - q^6 = 1 - \left(\frac{3}{5}\right)^6 = 0.953344$$

b) $P(A_1 \cup A_2 \cup \ldots \cup A_n)$

$$\begin{aligned}
&= 1 - P(\overline{\mathcal{A}}_1)P(\overline{\mathcal{A}}_2)\cdots P(\overline{\mathcal{A}}_6) \\
&= 1 - \left(1 - \frac{1}{4}\right)\left(1 - \frac{2}{5}\right)\left(1 - \frac{1}{5}\right)\left(1 - \frac{2}{7}\right)\left(1 - \frac{1}{6}\right)\left(1 - \frac{4}{9}\right) \\
&= 1 - \left(\frac{3}{4}\right)\left(\frac{3}{5}\right)\left(\frac{4}{5}\right)\left(\frac{5}{7}\right)\left(\frac{5}{6}\right)\left(\frac{5}{9}\right) = \frac{37}{42}
\end{aligned}$$

4.9.3 MUTUALLY EXCLUSIVE EVENTS AND STATISTICAL INDEPENDENCE

Most beginners always confuse the concept of mutually exclusive events with statistical independence. In general, mutually exclusiveness has to do with simultaneous occurrence or non-occurrence of events while statistical independence has to do with the effect of the occurrence of one event on the occurrence or non-occurrence of another event.

What we should note is that:

a) Mutual exclusiveness is associates with the addition rule while statistical independence is associated with the multiplication rule.
b) Two mutually exclusive events are strongly dependent, that is, if two events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive then if $\mathcal{A}$ occurs, $\mathcal{B}$ cannot occur and vice versa, and this is, the very opposite of independent events. Independence means that the outcome of one event does not influence the outcome of the other.
c) For two mutually exclusive events $\mathcal{A}$ and $\mathcal{B}$ to be independent, at least one of them should have a probability of zero.
d) Unlike mutually exclusive events, independent events cannot be spotted in Venn diagrams.

We summarise the basic differences in Table 4.3.

	Mutually Exclusive Events	Independent Events
1	$P(\mathcal{A} \cap \mathcal{B}) = 0$	$(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$
2	$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$	$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A})P(\mathcal{B})$
3	$P(\mathcal{A}\|\mathcal{B}) = 0$	$P(\mathcal{A}\|\mathcal{B}) = P(\mathcal{A})$
4	$P(\mathcal{A}\|\overline{\mathcal{B}}) = \dfrac{P(\mathcal{A})}{1 - P(\mathcal{B})}$	$P(\mathcal{A}\|\overline{\mathcal{B}}) = P(\mathcal{A})$

Table 4.3 Mutually Exclusive and Independent Events

This chapter concludes discussions on the fundamental theory of probability. In the next two chapters we shall introduce the concept of a random variable and its basic numerical characterisation.

EXERCISES

4.1 Show that the probability that exactly one of the events $\mathcal{A}$ and $\mathcal{B}$ occurs equals

$$P(\mathcal{A}) + P(\mathcal{B}) - 2P(\mathcal{A} \cap \mathcal{B}).$$

4.2 If $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events and it is known that $P(\mathcal{A}) = 0.20$ while $P(\mathcal{B}) = 0.30$. Estimate

(a) $P(\overline{\mathcal{A}})$ (b) $P(\overline{\mathcal{B}})$ (c) $P(\mathcal{A} \cup \mathcal{B})$ (d) $P(\overline{\mathcal{A}} \cap \overline{\mathcal{B}})$

4.3 Suppose that in Exercise 4.2, the events $\mathcal{A}$ and $\mathcal{B}$ are not mutually exclusive, re-evaluate the probabilities, assuming $\mathcal{A} and \mathcal{B}$ are independent.

4.4 Three mutually exclusive events $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ together exhaust the sample space of a random experiment. If $P(\mathcal{C}) = 2P(\mathcal{A})$ and $P(\mathcal{B}) = P(\mathcal{A} \cup \mathcal{C})$ find $P(\mathcal{A})$, $P(\mathcal{B})$, $P(\mathcal{C})$.

4.5 In a sample space $\mathcal{S}$, three events $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ have the probabilities

$$P(\mathcal{A}) = P(\mathcal{B}) = \frac{1}{3}, \quad P(\mathcal{C}) = \frac{1}{4}, \quad P(\mathcal{A} \cap \mathcal{B}) = \frac{1}{6}, \quad P(\mathcal{A} \cap \mathcal{C}) = \frac{1}{8} \text{and } P(\mathcal{B} \cap \mathcal{C}) = 0$$

Find the probability $P(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C})$

4.6 If $P(\mathcal{A}) = p$, $P(\mathcal{B}) = r$, and $P(\mathcal{A} \cup \mathcal{B}) = pr$, find

(a) $P(\overline{\mathcal{A}} \cup \overline{\mathcal{B}})$ (b) $P(\mathcal{A}^c \cup \mathcal{B})$

4.7 If $\mathcal{A}$ and $\mathcal{B}$ are disjoint events and $P(\mathcal{A}) = 0.3$ and $P(\mathcal{B}) = 0.6$, calculate

(a) $P(\mathcal{A} \cup \mathcal{B})$ (b) $P(\overline{\mathcal{A}} \cup \overline{\mathcal{B}})$ (c) $P(\overline{\mathcal{A}} \cup \mathcal{B})$ (d) $P(\overline{\mathcal{A}} \cap \mathcal{B})$

4.8 $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are three events in a sample space such that

$$P(\mathcal{A}) = x, \quad P(\mathcal{B}) = y, \quad P(\mathcal{A} \cap \mathcal{B}) = z.$$

Express each of the following probabilities in terms of x, y, and z:

(a) $P(\overline{\mathcal{A}} \cup \overline{\mathcal{B}})$ (b) $P(\overline{\mathcal{A}} \cap \overline{\mathcal{B}})$
(c) $P(\overline{\mathcal{A}} \cup \mathcal{B})$ (d) $P(\overline{\mathcal{A}} \cap \mathcal{B})$.

4.9 A certain town of population size 100,000 has 3 newspapers, $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$. The proportion of the inhabitants that read these papers are as follows:

$\mathcal{A}$: 10%, $\mathcal{B}$: 30%, $\mathcal{C}$: 5%, $\mathcal{A}$ and $\mathcal{B}$: 8%,
$\mathcal{A}$ and $\mathcal{C}$: 2%, $\mathcal{B}$ and $\mathcal{C}$: 4% $\mathcal{A}$ and $\mathcal{B}$ and $\mathcal{C}$: 1%.

a) Find the number of people reading only one newspaper.
b) How many people read at least two newspapers.
c) If $\mathcal{A}$ and $\mathcal{C}$ are morning papers, and $\mathcal{B}$ is an evening paper, how many people read at least one morning plus an evening paper?
d) How many people read one morning paper and one evening paper?

4.10 The following data were given in a study of a group of 1,000 subscribers to a certain magazine. In reference to sex, marital status, and education, there were 312 males, 470 married persons, 525 college graduates, 86 married males, and 25 married male college graduates. Show that the numbers reported in the study must be incorrect.

4.11 If you hold 5 tickets to a lottery for which n tickets were sold and 3 prizes are to be given, what is the probability that you will win at least 1 prize?

4.12 Prove Corollary 4.1.

4.13 A man has n keys, of which one will open his door. If he tries the keys at random, what is the probability that he will open the door on his k^{th} try

a) if he discards those keys that do not work,
b) if he does not discard previously tried keys?

4.14 Roll a pair of fair dice once. Given that the two numbers that occur are not the same, what is the probability that the sum is 7 or that the sum is 4 or that the sum is 12.

4.15 Suppose a fair coin is tossed four times and on each toss a head comes up. What is the probability that it will come up head on the fifth toss as well? Comment on the result.

4.16 Two balls are selected at random without replacement from a box which contains 4 white and 8 black balls. Compute the probability that

(a) both balls are white (b) the second ball is white.

4.17 A box contains 8 bulbs, of which 5 are good and 3 are defective. If 3 bulbs are randomly taken from the box, what is the probability that all are good?

4.18 Box 1 contains 4 defective and 16 non-defective light-bulbs. Box 2 contains 1 defective and 1 non-defective light bulbs. Roll a fair die once. If you get a 1 or a 2, then we select a bulb at random from box 1. Otherwise we select a bulb from box 2. What is the probability that the selected bulb will be defective?

4.19 A box contains 8 red, 3 white and 9 blue balls. If 3 balls are drawn at random without replacement, determine the probability that

a) all 3 are red
b) the balls are drawn in the order red, white, blue.

4.20 Suppose that in Exercise 4.18 a bulb selected is defective. What is the probability that the bulb comes from box 2?

4.21 Two fair dice are thrown once. Find the probability that the sum is 5, given that one of the dice came up with a 3?

4.22 Refer to Exercise 4.21. Find the probability that the sum is 5, given that neither die came up with a 1.

4.23 A pair of dice, one blue and one white, is rolled once and the sum of the numbers that show is 7.

a) What is the probability that the blue die shows the number 3?
b) If it is known that the first number is 3 what is the probability that it was on the blue die?

4.24 A manufacturing company has two plants, 1 and 2. Plant 1 produces 40% of the company's output and plant 2 produces the other 60%. Of the output produced by plant 1, 95% are good and of that produced by plant 2, 10% are defective. If a product is randomly selected from the output of this company, what is the probability that the output will be good?

4.25 Three shots are fired at a target in succession. The probability of a hit in the first shot is $p_1 = 0.3$, in the second, $p_2 = 0.6$, and in the third $p_3 = 0.8$. The probability of destroying the target in the first hit $\lambda_1 = 0.4$, the second hit, $\lambda_2 = 0.7$, and in the third hit, $\lambda_3 = 1.0$. What is the probability that the target will be

a) destroyed
b) by three shots.

4.26 From past experience with the illness of his patients, a doctor has gathered the following information: 5% feel that they have cancer and do have cancer, 45% feel that they have cancer and don't have cancer, 10% do not feel that they have cancer and do have it, and finally 40% feel that they do not have cancer and really do not have it. A patient is randomly selected from this doctor's practice. What is the probability

a) that he has cancer, given that
   i) he feels that he has it,
   ii) he does not feel that he has it,

a) What is the probability that he feels he has cancer, given that
   i) he does not have it
   ii) he does have it.

4.27 Refer to Exercise 4.24. If a product selected at random was good, what is the probability that it comes from plant 1.

4.28 Suppose that you are a political prisoner in Ghana and are to be sent to one of two prisons Nsawam or James Fort. The probabilities of being sent to these two places are 0.7 and 0.3, respectively. Suppose it is known that 85% of the residents of Nsawam wear a white shirt or blouse, whereas in James Fort it is 80%. Late one night you are blindfolded and thrown on a truck. Two weeks later (you estimate) the truck stops, you are told you have arrived at your place of exile, and your blindfold is removed. The first person you see is not wearing a white shirt or blouse. What is the probability that the prison you are being sent to is Nsawam?

4.29 Of 100 patients in a hospital with a certain disease, ten are chosen to undergo a drug treatment that increases the percentage cured rate from 50 percent to 75 percent. If a doctor later encounters a cured patient, what is the probability that he received the drug treatment?

4.30 (Bertrand's Paradox)

Three boxes each contain two coins. In one box, $\mathcal{B}_1$, both coins are gold, in another, $\mathcal{B}_2$, both are silver, and in the third, $\mathcal{B}_3$, one is gold and the other is silver. A box is chosen at random and from it, a coin is chosen at random. If this coin is gold, what is the probability that it came from the box containing two gold coins?

4.31 Suppose it is known that 3% of cigarette smokers develop lung cancer, whereas only 0.5% of non-smokers develop lung cancer. Furthermore, suppose that 30% of adults smoke. If it is known that a randomly chosen adult had developed lung cancer, what is the probability that the person is a smoker?

4.32 Two hunters independently fired at a bird. The probability that the first hunter will kill the bird is $p_1 = 0.8$ and that of the second is $p_2 = 0.4$. Suppose the bird is killed by a single hit. What is the probability that it was killed by the first hunter.

4.33 Prove Theorem 4.15.

4.34 A box contains ten numbered identical balls. A ball is picked at random with replacement from the box. Consider the events:

$$\mathcal{A} = \{1, 2, 3, 4\}; \quad \mathcal{B} = \{1, 2, 3, 5, 8\}; \quad \mathcal{C} = \{3, 4, 5, 6, 7\}$$

Verify whether they are independent.

4.35 Refer to the experiment in Exercise 4.34. Consider the events:

$$\mathcal{A} = \{1, 2, 3, 4\}; \quad \mathcal{B} = \{2, 4, 6, 8, 10\} \quad \mathcal{C} = \{1, 3, 6, 9, 10\}$$

Verify whether they are independent.

4.36 Refer to the experiment in Exercise 4.34. Consider the events:

$$\mathcal{A} = \{1, 3, 5, 7\}; \quad \mathcal{B} = \{4, 5\} \quad \mathcal{C} = \{1, 2, 3, 4\}$$

Verify whether they are independent.

4.37 Suppose $\mathcal{A}$ and $\mathcal{B}$ are independent with $P(\mathcal{A}) = 0.54$ and $P(\mathcal{B}) = 0.4$. Find $P(\mathcal{A} \cup \mathcal{B}^c)$.

5 RANDOM VARIABLES

5.1 INTRODUCTION

The foundation of probability theory was set forth in the preceding chapters. The ideas of an experiment, a sample space corresponding to the experiment and events in a sample space were introduced. The axioms of a probability measure on these events were also postulated. In this chapter, the concept of a random variable and its probability and distribution functions are introduced. This is to unify the study of probabilistic situations which is achieved by mapping the original sample space to the real line for any experiment. This means we will only have to study the one sample space, the real line.

5.2 CONCEPT OF RANDOM VARIABLE

The random variable is one of the fundamental ideas of probability theory. A fully rigorous discussion of random variables requires a knowledge of measure theory but the content of this chapter is sufficient for the needs of introductory course in probability theory.

5.2.1 DEFINITION OF RANDOM VARIABLE AND ITS SCHEMATIC REPRESENTATION

We shall consider a random variable simply as a real-valued function defined on a sample space $\mathcal{S}$.

Definition 5.1 RANDOM VARIABLE

Let $\mathcal{S}$ be the sample space associated with some experiment $\mathcal{E}$. A random variable X is a function that assigns a real number $X(s)$ to each sample point $s \in \mathcal{S}$

Schematically, we may present this concept of a random variable as in Figure 5.1. A random variable (r.v.) is sometimes called a ***stochastic variable***, a ***random function*** or a ***stochastic function*** and usually denoted by a capital letter such as X, Y, S, T, Z and the corresponding lower case letters, $x, y, s, t, z,$ to denote particular values in its range.

Fig. 5.1 Concept of Random Variable

One of the essential things that the notion of random variables does is to provide us the power of abstraction and thus enabling us avoid dealing with unnecessarily complicated sample spaces. Suppose X is a random variable and x a real number. We define

$$\mathcal{A}_X = \{s \in \mathcal{S} | X(s) = x\}$$

where event $\mathcal{A}_X$ is the subset of $\mathcal{S}$ consisting of all sample points s which the random variable X assigns the value x. Clearly,

$$\begin{aligned} \mathcal{A}_X \cap \mathcal{A}_Y &= \emptyset, \quad \text{if } x \neq y \quad \text{and} \\ \bigcup_{X \in R} \mathcal{A}_X &= \mathcal{S} \end{aligned}$$

Thus, the collection of events $\mathcal{A}_X$ for all x defines an ***event space*** which we may find more convenient to work with rather than the original sample space, provided our only interest in performing the experiment is with the resulting experimental value of the random variable X.

Example 5.1
Let $\mathcal{E}$ be an experiment of tossing a coin twice and let us be interested in the number of heads $\mathcal{H}$ which come up. Then the sample space $\mathcal{S}$ associated with this experiment is

$$\mathcal{S} = \{\mathcal{HH}, \mathcal{HT}, \mathcal{TH}, \mathcal{TT}\}$$

A random variable X and a sample space $\mathcal{S}$ can be defined such that for $s \in \mathcal{S}$, $X(s)$ is the number of heads in the point $s \in \mathcal{S}$. Pictorially the random variable can be viewed as a mapping (see Fig. 5.2).

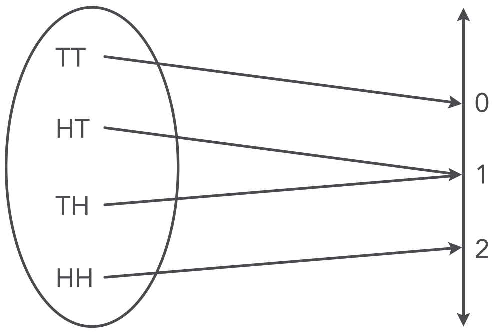


Fig. 5.2 Number of Heads with a Toss of two Coins

This example may be presented in a table, such as the one below:

Sample point	$\mathcal{HH}$	$\mathcal{HH}$	$\mathcal{TH}$	$\mathcal{TT}$
X	2	1	1	0

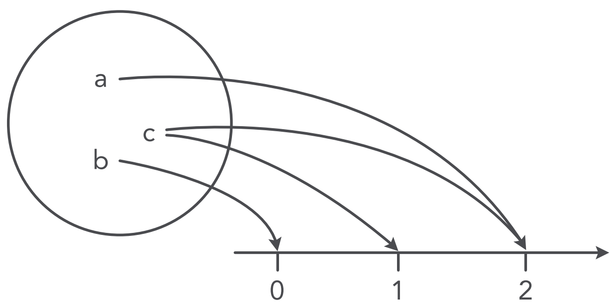
In this table the event $\{s : X(s) = 1\}$ is simply the set $\{\mathcal{HT}, \mathcal{TH}\}$. The notation $\{s : X(s) = 1\}$ is often shortened to $\{X = 1\}$. In general $\{X = x\}$ will be used to represent the event $\mathcal{A}_X$. Similarly $\{s : X(s) < 1\}$ is shortened to $\{X < 1\}$. And finally, the probabilities of the two events $(X = 1)$ and $(X < 1)$ are usually written as $P(X = 1)$ or simply $p(1)$ and $P(X < 1)$ respectively.

From Example 5.1, the original sample space comprised four sample points but the event space has three event points, namely, 0, 1, 2. In Example 5.4 we shall observe that when a coin is tossed three times, the original sample space contains eight sample points but the event space defined by X contains four event points. In general, for a sequence of n coin-tossing experiments, say, there are 2^n sample points in the original sample space whereas the event space defined by X will have $(n+1)$ event points.

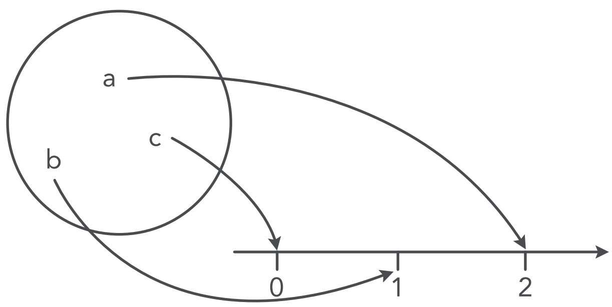
It is important to note that not every mapping from $\mathcal{S}$ to the real line will be considered as a random variable. For example, a function that assigns more than one real number to any element of $\mathcal{S}$ (a one to-many function), is unacceptable. Only functions that assign exactly one real number to each element of $\mathcal{S}$ are acceptable. Of course, the same real number may be assigned to many elements of $\mathcal{S}$ (that is, a many-to-one function is also satisfactory). The domain of a random variable X is the sample space $\mathcal{S}$ and the range space $\mathcal{R}_x$ is a subset of the real line.

Example 5.2
Refer to Example 5.1. Which of the following relations is a random variable?

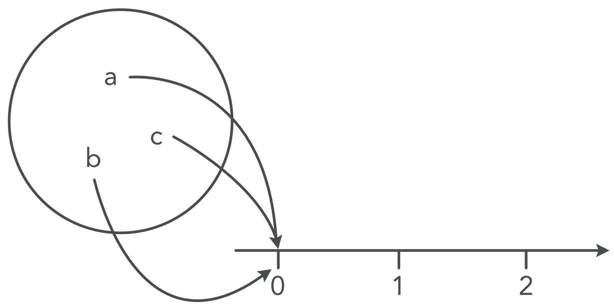
i) Suppose the sample point b corresponds to "no head", c to "one head" and also to "two heads", and a to "two heads". The mapping is given below:



ii) Suppose the sample point c corresponds to "no head", b to "one head" and a to "two heads":



iii) Suppose the sample point a corresponds to “no head”, b to “no head” and c to “no head”.



Solution
The function defined by (i) is not a random variable because the element c corresponds to both 1 and 2.
The function defined by (ii) is a random variable because it is a one-to-one function.
The function defined by (iii) is a random variable because it is a many-to-one function.

Just as there are often many sample spaces that can be of interest in an experiment, there can also be many different random variables of interest for the sample space.

Example 5.3
We may define the following other random variables for the sample space for Example 5.1:

a) The number of heads times the number of tails.
b) the number of heads plus the number of two tails.
c) The square of the number of tails minus half the number of heads.

5.2.2 TYPES OF RANDOM VARIABLES

We shall distinguish between two basic types of random variables: the *discrete* and the *continuous* random variables.

Definition 5.2 DISCRETE RANDOM VARIABLE

A random variable which takes on a finite or countably infinite number of values is called a discrete random variable

That is, the possible values of X may be listed as $x_1, x_2, \cdots, x_n, \cdots$. In the finite case the list terminates at n and in the countably infinite case, the list continues indefinitely.

> **Definition 5.3** CONTINUOUS RANDOM VARIABLE
>
> A random variable which takes on an uncountably infinite number of values is called a continuous random variable

A synonym for a continuous random variable is a ***non-discrete random variable***.

A random variable may also be partly discrete and partly continuous but such random variables will not be considered in this book.

5.3 PROBABILITY DISTRIBUTION OF RANDOM VARIABLE

The concept of probability distribution of a random variable is very crucial in probability theory. If we know the probability distribution of a random variable, we will be able to solve many problems related to that variable. In our discussion of the probability distribution, we shall do so separately for discrete and continuous cases.

5.3.1 PROBABILITY DISTRIBUTION OF DISCRETE RANDOM VARIABLE

For a complete characterisation of a discrete random variable *X*, it is necessary and sufficient to know

a) the list of all possible values $x_i, \ \ i = 1, 2, \cdots$ and
b) the corresponding probability $P(X = x_i)$ of each of these values.

When such information is available we say that the *distribution law* or the *probability distribution* or simply the *distribution* of the random variable X is known.

Definition 5.4 PROBABILITY DISTRIBUTION
(Discrete Case)

A probability distribution of a discrete random variable X is a sequence of the values $x_i \ (i = 1, 2, ...)$ of X, together with a probability assigned to each point $x_i, \ i = 1, 2, \cdots$

The values x_i may be either finite or countably infinite and in any order though for convenience it should be in increasing order of magnitude. The probability distribution of a discrete random variable is more often called a *probability function (p.f.) of a discrete random variable* or a *probability mass function (p.m.f.)* and denoted by $p(x_i)$ or $P(X = x_i)$. It is the probability that the random variable *X* assumes a specific value x_i. Note that for a specific value $x_i, \ p(x_i) = P(X = x_i)$

Representation of Probability Distribution of X

The probability distribution of a discrete random variable *X* can be represented by a *table*, a *graph* or a *formula*.

Tabular Form
It is convenient to specify the probability distribution of a random variable by means of a table having two rows: the upper row contains the possible values the random variable assumes and the lower row contains the corresponding probabilities of the values (see Table 5.1).

x_i	x_1	x_2	...	x_n
$p(x_i)$	$p(x_1)$	$p(x_2)$	...	$p(x_n)$

or

X	x_1	x_2	...	x_n
$P(X = x)$	$p(x_1)$	$p(x_2)$	...	$p(x_n)$

Table 5.1 Probability distribution of X

When the set of possible values of a random variable X is countably infinite then the probability distribution of the random variable X is given in the form of the following table:

x_i	x_1	x_2	...
$p(x_i)$	$p(x_1)$	$p(x_2)$	...

Graphical Form

The probability distribution may also be given graphically. The graph represents chance and not data.

Definition 5.5 PROBABILITY GRAPH

A graph of $p(x_i)$ against x_i is called a probability graph

To obtain a probability graph, vertical lines or bars are drawn above the possible values x_i of the random variable X on the horizontal axis. The height of each line or bar (rectangle) equals the probabilities of the corresponding values of the random variable (see Fig. 5.3).

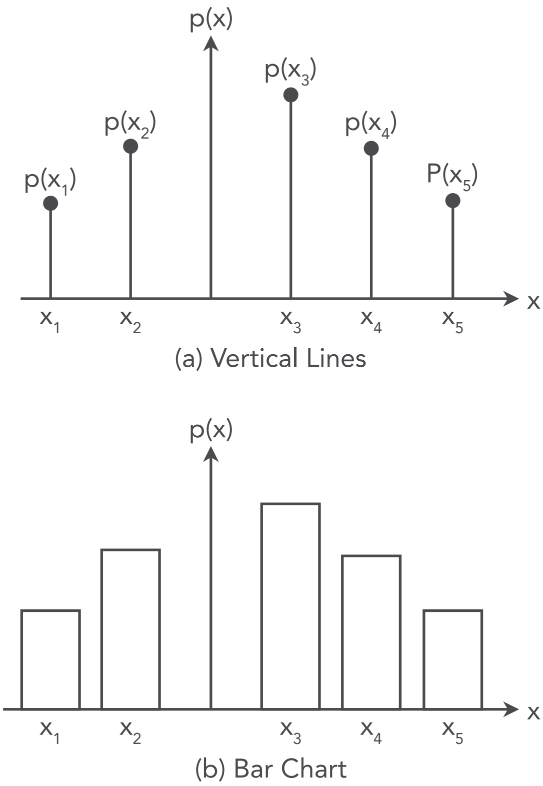


Fig. 5.3 Typical Graph of Probability Chart

Probability distributions of a discrete random variable may also be represented by a ***probability histogram*** (see Fig. 5.4).

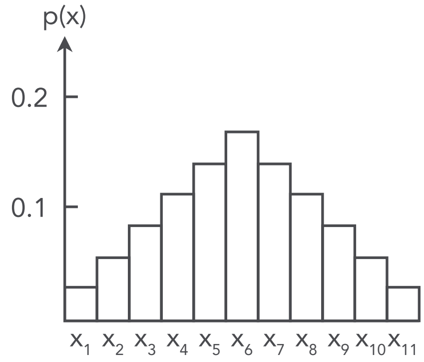


Fig. 5.4 Typical Probability Histogram

Similar to the probability bar chart, the height of each rectangle of a probability histogram is proportional to the probability that the random variable X takes on the value which corresponds to the midpoint of its base. The values $0, 1, 2, \cdots$ have respectively, the midpoints -0.5 and 0.5, 0.5, and 1.5, 1.5 and 2.5, $\cdots$ which are proportional to their corresponding probabilities. In using the midpoints we are in a sense " spreading" the values of the given discrete random variable over a continuous scale. We can, therefore, construct a polygon using these midpoints and then approximate the graph of a discrete random variable with a continuous curve.

If each of the rectangles of the histogram has a unit width then we say that the areas, rather than their heights, equals the corresponding probabilities. In the situation when the rectangles of the histogram do not all have unit width, we adjust the heights of the rectangles or modify the vertical scales.

Formula
Representing a discrete probability distribution by means of a formula is the most useful method. For example

$$P(X = x) = \frac{1}{n}, \qquad x = 1, 2, \cdots, n$$

defines a probability distribution of a random variable X.

Note

This is a special case of a discrete distribution known as the discrete uniform probability distribution.

We must point out immediately that not all functions of a discrete random variable qualify to be called probability mass functions. We shall now give a formal definition of a probability mass function.

Definition 5.6 PROBABILITY MASS FUNCTION

Any function p defined on all possible values of the discrete random variable $X = x_i,\ i = 1, 2, \cdots$ is called a probability mass function if it satisfies the following properties:

Property 1: $p(x_i) \geq \quad 0, \quad i = 1, 2, \cdots$

Property 2: $\sum_i p(x_i) = \quad 1$

where the summation is over all possible values of the random variable X

This definition is obvious. A random variable *X* in one experiment takes *only one of its* n *possible values* with a corresponding probability which is non-negative. In other words, in an experiment, one of the only possible and pairwise mutually exclusive events:

$$X = x_1,\ X = x_2,\ \cdots,\ X = x_n$$

will happen by all means. Such events, as pointed out earlier, forms a partition of the sample space, and their union is a sure event. Consequently, the sum of the probabilities of these events:

$$p(x_1) + p(x_2) + \cdots + p(x_n) = 1$$

As a converse to Definition 5.6, we may state that any sequence of $p(x_i),\ i = 1, 2, \cdots$ satisfying these two properties defines some discrete random variable *X*.

Example 5.4
A fair coin is tossed three times. Let *X* represent the number of heads which come up.

a) Find the probability distribution corresponding to the random variable *X*.
b) Construct a probability graph.

Solution
a) The sample space is

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

The probability of each outcome is $\frac{1}{8}$, since all the outcomes are equally likely simple events.

With each sample point we can associate a number for X, as shown in the table below.

Sample point, s	$\mathcal{TTT}$	$\mathcal{TTH}$	$\mathcal{THT}$	$\mathcal{HTT}$	$\mathcal{HHT}$	$\mathcal{HTH}$	$\mathcal{THH}$	$\mathcal{HHH}$
Number of Heads, X	0	1	1	1	2	2	2	3
$P(X = x)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

The table above shows that the random variable can take the values $0, 1, 2, 3$. Our next task is to compute the probability distribution $p(x_i)$ of X.

$$\begin{aligned} P(X=0) &= P(\mathcal{TTT}) = \frac{1}{8} \\ P(X=1) &= P(\{\mathcal{TTH}\} \cup \{\mathcal{THT}\} \cup \{\mathcal{HTT}\}) \\ &= P(\mathcal{TTH}) + P(\mathcal{THT}) + P(\mathcal{HTT}) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8} \\
P(X=2) &= P(\{\mathcal{HHT}\}\cup\{\mathcal{HTH}\}\cup\{\mathcal{THH}\}) \\
&= P(\mathcal{TTH})+P(\mathcal{THT})+P(\mathcal{HTT}) \\
&= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8} \\
P(X=3) &= P(\mathcal{HHH})=\frac{1}{8}
\end{aligned}
$$

The probability distribution is thus tabulated as:

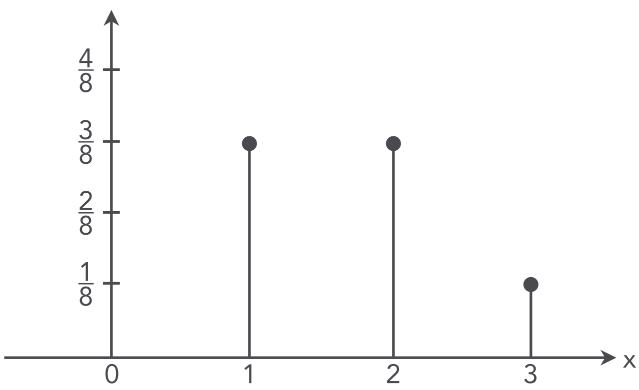
x_i	0	1	2	3
$p(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Note

This probability distribution satisfies the properties given in Definition 5.6, namely,

$$p(x_i)\geq 0 \quad \text{and} \quad \sum p(x_i)=1$$

b) The diagram which follows graphically describe the above distribution.



Example 5.5
A discrete function is given by

$$p(x)=\begin{cases}\frac{1}{21}(2x+3), & x=1,2,3\\[2ex] 0, & \text{elsewhere}\end{cases}$$

Verify that it is a probability function of some random variable.

Solution
Clearly, $p(x_i) \geq 0$
and

$$\frac{1}{21}\sum_{x=1}^{3}(2x+3) = \frac{1}{21}\left[(2+3)+(4+3)+(6+3)\right] = \frac{1}{21}(5+7+9) = 1$$

Hence the function is a p.m.f.

Example 5.6
A discrete random variable X has a p.m.f.

$$p(x) = \begin{cases} k(x-1), & x = 3,4,5 \\ \\ 0, & \text{elsewhere} \end{cases}$$

Find the constant k for which $p(x)$ is a probability function.

Solution
Since the function is a p.m.f.

$$\sum_{i=3}^{5} k(x-1) = 1$$

Now,

$$\sum_{3}^{5} k(x-1) = k\left[(3-1)+(4-1)+(5-1)\right] = 9k = 1$$

From which $k = \dfrac{1}{9}$.

5.3.2 PROBABILITY DISTRIBUTION OF CONTINUOUS RANDOM VARIABLE

In the previous section, we considered the probability distribution of a discrete random variable. For a continuous random variable, constructing such a table in which all possible values of the random variable would be listed is impossible. We can intuitively think of a continuous random variable as one that takes on values in some interval of the real line. The fundamental question we are asking is "what is the probability that the random variable assumes a value in some interval of numbers".

The probability distribution for a continuous random variable is more often called a ***probability density function*** (p.d.f.) or simply *density function* and is denoted by $f(x)$.

> **Definition 5.7** PROBABILITY DENSITY FUNCTION
>
> A function f defined on the real numbers is called a probability density function (p.d.f.) if it satisfies the following properties:
>
> *Property 1*: $f(x) \geq 0$ for all x
>
> *Property 2*: $\int_{-\infty}^{\infty} f(x)\,dx = 1$

The first property indicates that, unlike the p.m.f $p(x_i)$, the value of the p.d.f. $f(x)$ at the point x is not a probability and that it is perfectly acceptable if $f(x) > 1$ at the point x. It is when the function is integrated over some interval that a probability is obtained, hence the name “density”.

The second property is a mathematical statement of the fact that a real-valued random variable must certainly lie between -∞ and ∞. That is, the probability that a random variable X takes any value between -∞ and ∞ is a sure event.

If $f(x)$ is a p.d.f for a continuous random variable X then we can represent $y = f(x)$ graphically by a curve as in Figure 5.5. By Property 1, $f(x)$ is non-negative so the curve cannot fall below the x-axis. By Property 2 the total area under the curve (i.e bounded by the curve and the x-axis) must be 1.

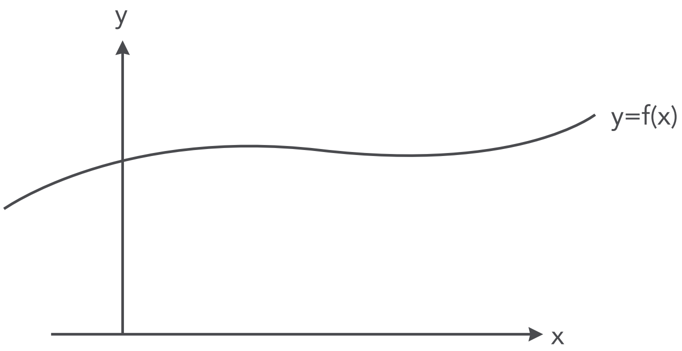


Fig. 5.5 Curve of Probability Density Function

The following definition gives the relationship between a random variable and its density function.

Definition 5.8

X is said to be a continuous random variable if there exists a function f, called the probability density function of X such that

a) $f(x) \geq 0$

b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

c) $P(a \leq X \leq b) = \int_a^b f(x)\,dx, \quad -\infty \leq a \leq b \leq \infty$

Geometrically, relation (c) in Definition 5.8 means the following: the probability that a random variable X takes the value in the interval (a, b) equals the area of the region defined by the curve of the probability distribution $y = f(x)$, the straight lines $x = a$ and $x = b$, and the x-axis (see shaded region in Figure 5.6).

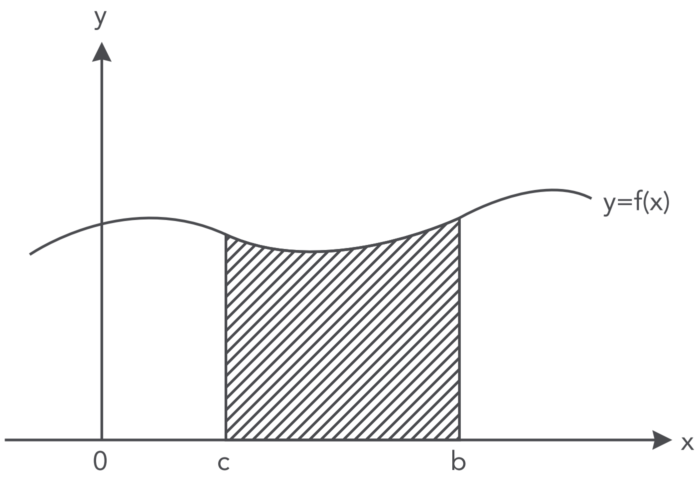


Fig. 5.6 Area of Probability Density Function

If all non-zero possible values of the random variable X lie in the interval (a, b) then they lie in the interval $(-\infty, \infty)$ as well. Thus

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Note

a) It is not actually necessary that $f(x)$ be continuous everywhere. It is only necessary that the derivative

$$\frac{d}{dx} f(x) = f'(x)$$

exists everywhere except at possibly a finite number of points in which case the function $f(x)$ is said to be piecewise continuous.

b) $f(x)$ needs not be less than unity for all values of X. It needs only be nonnegative piecewise continuous and have unit area; and any such function is the p.d.f. of some random variable.

Theorem 5.1

If X is a continuous random variable having p.d.f. $f(x)$, then for any number a,

$$P(X = a) = 0$$

Proof

$$\begin{aligned} P(X=a) &= P(a \le X \le a) \\ &= \int_a^a f(x)\,dx = 0 \end{aligned}$$

This theorem does not imply that the set $\{s|X(s)=a\}$ is empty but that the probability assigned to this set is zero. That is, the probability that a continuous random variable X takes one definite value, say a, is zero even though the probability density may not be zero. That is, if $x_0 \in X$, the fact that $P(X = x_0) = 0$ does not mean that x_0 cannot occur. For example, choose a point in the interval $(0,1)$ at random. Each point has a probability of zero of being chosen on any particular trial but on each trial some point is chosen.

Theorem 5.2

For any number a and b, with $a \le b$

$$\begin{aligned} P(a \le X \le b) &= P(a \le X < b) \\ &= P(a < X \le b) \\ &= P(a < X < b) \end{aligned}$$

Proof

$$\begin{aligned} P(a \leq X < b) &= P(X = a) + P(a < X < b) \\ &= 0 + P(a < X < b) \qquad \text{(by Theorem 5.1)} \\ &= P(a < X < b) \end{aligned}$$

The other parts may be proved in a similar way.

That is, the probability that X is in the interval from a to b is the same whether a itself is included or excluded or whether b also is included or excluded. This situation of course is quite different from that where a random variable is discrete.

Example 5.7

Let X be a continuous random variable such that

$$f(x) = \begin{cases} \dfrac{1}{8}x, & 0 < x < 4 \\ 0, & \text{elsewhere} \end{cases}$$

a) Show that $f(x)$ is a p.d.f.
b) Sketch the graph.

Solution

a) It is clear $f(x) \geq 0$ so Property 1 is satisfied. For the $f(x)$ to be a p.d.f., it must also satisfy Property 2, that is

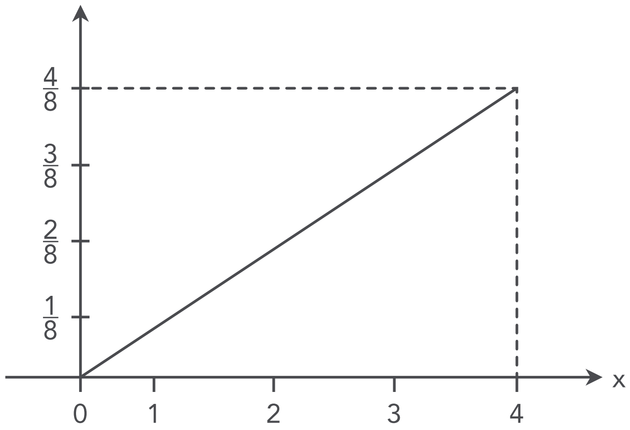
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Now,

$$\int_0^4 f(x)dx = \int_0^4 \frac{1}{8}x\,dx = 1$$

Hence $f(x)$ is a p.d.f.

b) The graph of the probability density of this example is presented below.



Example 5.8
A random variable X has the p.d.f.

$$f(x) = \begin{cases} ax, & 0 < x < 4 \\ 0, & \text{elsewhere} \end{cases}$$

where a is a constant.

a) Find the constant a;
b) Compute $P(2 < X < 3)$.

Solution

a) Since the function is a p.d.f.

$$\int_0^4 f(x)\,dx = \int_0^4 a\,x\,dx = 8a = 1$$

Hence $a = \frac{1}{8}$.

b) $P(2 < X < 3) = \int_2^3 \frac{1}{8}\,x\,dx = \frac{5}{16}$

5.4 CUMULATIVE DISTRIBUTION FUNCTION

5.4.1 DEFINITION OF CUMULATIVE DISTRIBUTION FUNCTION

The cumulative distribution function (c.d.f.) or simply the distribution function is the most universal characteristic of a random variable. It exists for all random variables whether they are discrete or continuous.

Definition 5.9 CUMULATIVE DISTRIBUTION FUNCTIONS

Let X be a random variable and x any real number. The cumulative distribution function of X is a function F defined as the probability that the random variable X takes a value less than or equal to x:

$$F(x) = P(X \leq x)$$

The function $F(x)$ in the above definition may be written as:

$$F(x) = P(-\infty < X \leq x)$$

5.4.2 CUMULATIVE DISTRIBUTION FUNCTION OF DISCRETE RANDOM VARIABLES

The next two theorems follow trivially from Definition 5.9.

Theorem 5.3 DISTRIBUTION FUNCTION OF DISCRETE RANDOM VARIABLE

Let X be a discrete random variable with probability $p(x_i)$, then the cumulative distribution function is given by

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

If X takes on only a finite number of values $x_1, x_2 \cdots, x_n$ then the distribution function $F(x)$ is given by

$$F(x) = \begin{cases} 0 & -\infty \ \leq x < x_1 \\ p(x_1) & x_1 \leq x < x_2 \\ p(x_1) + p(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ p(x_1) + p(x_2) + \cdots + p(x_n) = 1 & x_n \leq x < +\infty \end{cases}$$

Fig. 5.7 depicts the graph of $F(x)$ which is discontinuous at the possible values $x_1, x_2, \cdots, x_n$. At these points $F(x)$ is continuous from the right but discontinuous from the left. Because of the appearance of its graph, the cumulative distribution function for a discrete random variable is also called a ***staircase function*** or a ***step function***, having jump discontinuities at all possible values $X = x_i$ with a step at the x_i of height $p(x_i)$. The graph increases only through these jumps at $x_1, x_2, \cdots, x_n$. Everywhere else in the interval $[x_i, x_{i+1}]$, the cumulative distribution function $F(x_i)$ is constant.

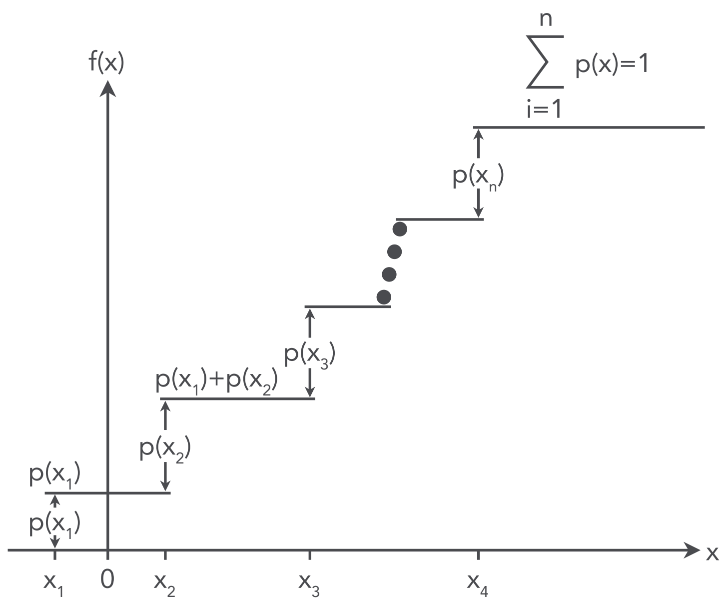


Fig. 5.7 Typical Distribution Function Graph (Discrete Case)

Example 5.9
Refer to Example 5.4.

a) Find the cumulative distribution function and,
b) Sketch its graph.

Solution

a) To obtain the cumulative distribution function by definition, we need the following steps:

Step 1

We find the probability distribution of the random variable X. The probability distribution of this example has been found already in Example 5.4. We reproduce the result here for convenience.

x_i	0	1	2	3
$p(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Step 2

We find the cumulative distribution function:

$$\begin{aligned} F(0) = P(X \leq 0) &= P(0 \leq X < 1) \\ &= P(X < 0) + P(X = 0) \\ &= 0 + p(0) = \frac{1}{8} \end{aligned}$$

$$\begin{aligned} F(1) = P(X \leq 1) &= P(0 \leq X < 2) \\ &= P(X < 0) + P(X = 0) + P(X = 1) \\ &= 0 + p(0) + p(1) \\ &= 0 + \frac{1}{8} + \frac{3}{8} = \frac{4}{8} \end{aligned}$$

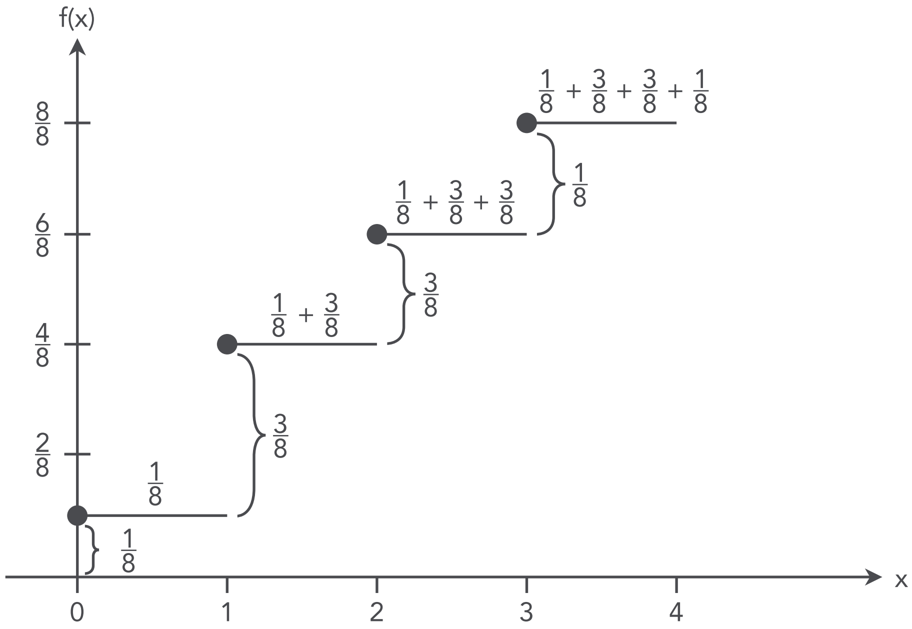
$$\begin{aligned} F(2) = P(X \leq 2) &= P(0 \leq X < 3) \\ &= P(X < 0) + P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0 + p(0) + p(1) + p(3) = 0 + \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8} \end{aligned}$$

$$
\begin{aligned}
F(3) = P(X \leq 3) &= P(0 \leq X \leq 3) \\
&= P(X < 0) + P(X = 0) + P(X = 1) + P(X = 2) \\
&\qquad + P(X = 3) \\
&= 0 + p(0) + p(1) + p(3) + p(4) \\
&= 0 + \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1
\end{aligned}
$$

Hence the cumulative distribution function is

$$
F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 4/8 & 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}
$$

b) The graph of $F(x)$ is shown in the figure below.



Finding the probability distribution from the distribution function is straightforward. If $F(x)$ is the distribution function of a discrete random variable X, we then find the points at which the distribution function jumps, and the jump sizes. The probability function has masses exactly at those jump points, with the probability masses being equal in magnitude to the respective jump sizes. It is for this reason that it is called probability mass function.

Example 5.10
Suppose we were given the distribution function of Example 5.7, find its probability distribution.

Solution
Note from the graph of this distribution function (see Example 5.7) that the magnitude or the height (that is, $p(x_i)$) of the jumps (steps) at $0, 1, 2, 3$, are $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$, respectively, hence

$$p(x) = \begin{cases} 1/8, & x = 0 \\ 3/8, & x = 1 \\ 3/8, & x = 2 \\ 1/8, & x = 3 \end{cases}$$

Note

We can obtain this result without the graph by finding the difference in the adjacent values of $F(x)$.

5.4.3 CUMULATIVE DISTRIBUTION OF CONTINUOUS RANDOM VARIABLES

Theorem 5.4 DISTRIBUTION FUNCTION OF CONTINUOUS VARIABLES

Let X be a continuous random variable with probability distribution function $f(x)$. Then the cumulative distribution function $F(x)$ is given by:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Proof

We know that

$$P(a < x < b) = \int_{a}^{b} f(x)\,dx$$

By Definition 5.9

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(-\infty < X \leq x) \\ &= \int_{-\infty}^{x} f(t)\,dt \end{aligned}$$

The typical graph of the cumulative distribution of the continuous random variable is shown in Fig. 5.8.

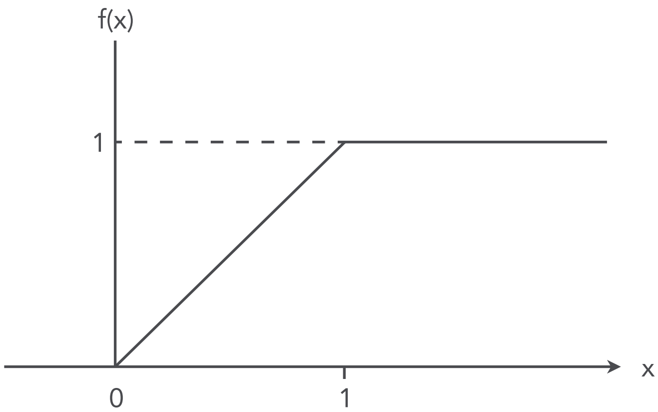


Fig. 5.8 Typical Distribution Function Graph (Continuous Case)

The graph of $F(x)$ is continuous. Its slope needs not be everywhere continuous, but where this is so, it is equal to the probability density function. That is to say,

$$F'(x) = f(x)$$

Example 5.11
The probability density function of X is given by

$$f(x) = \begin{cases} 0, & x < 0 \\ \dfrac{x}{2}, & 0 \le x \le 2 \\ 0, & x > 2 \end{cases}$$

Find the cumulative distribution function and sketch its graph.

Solution
If $x < 0$, then

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = 0$$

If $0 < x < 2$ then

$$F(x) \;=\; \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{x} f(t)\,dt$$

$$
\begin{aligned}
&= 0 + \int_0^x f(t)\,dt \\
&= \int_0^x \frac{t}{2}\,dt \\
&= \frac{x^2}{4}
\end{aligned}
$$

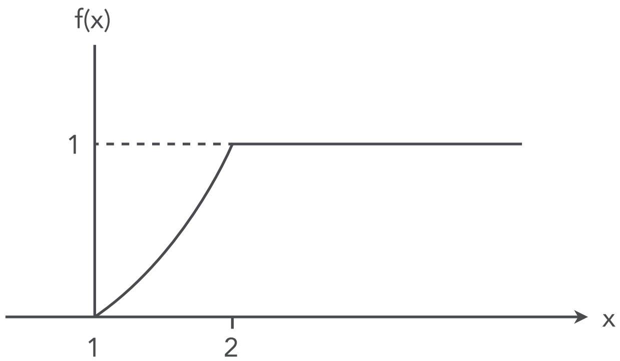
If $x > 2$, then

$$
\begin{aligned}
F(x) &= \int_{-\infty}^0 f(t)\,dt + \int_0^2 f(t)\,dt + \int_2^x f(t)\,dt \\
&= 0 + 1 + 0 = 1
\end{aligned}
$$

Thus,

$$
F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{x^2}{4}, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}
$$

The graph of $F(x)$ is shown below.



Note

The c.d.f. $F(x)$ for a continuous random variable will always be a continuous function but the p.d.f $f(x)$ may or may not be.

5.4.4 PROPERTIES OF CUMULATIVE DISTRIBUTION FUNCTION

Property 1

The function $F(x)$ is a probability, consequently,

$$0 \le F(x) \le 1$$

for all $x \in (-\infty, \infty)$.

Property 2

$F(x)$ is a nondecreasing function of x, that is, for two particular values of x_1 and x_2, if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.

Property 3

The probability that a random variable X takes the value within an interval (a, b) is equal to the increment of the distribution function in that interval:

$$P(a < X < b) = F(b) - F(a)$$

This means that all probabilities of interest can be computed once the cumulative function $F(x)$ is known.

Property 4

a) $F(+\infty) \quad = \quad \lim_{x \to +\infty} F(x) = 1$

b) $F(-\infty) \quad = \quad \lim_{x \to \ -\infty} F(x) = 0$

Property 5

$F(x)$ is always right continuous, that is

$$\lim_{x \to x_0 + 0} F(x) = F(x_0)$$

Note

It is not true that $F(x)$ is continuous from the left, that is,

$$\lim_{x \to x_0 - 0} F(x) \neq F(x_0)$$

for all points x_0. This is because we have defined $F(x) = P(X \leq x)$. If $F(x)$ had been defined as $P(X < x)$ (strict inequality), it would have been continuous on the left, but not on the right.

Any function satisfying all the five properties stated above is the c.d.f. of some random variable.

Example 5.12
A random variable X has c.d.f.

$$F(x) = \begin{cases} 0, & x \leq -5 \\ \dfrac{x}{6} + 1, & -5 < x \leq 0 \\ 1, & x > 0 \end{cases}$$

Find $P(-1 < X < 0)$.

Solution

$$P(-1 < X < 0) = F(0) - F(-1)$$

Now

$$\begin{aligned} F(0) &= \frac{0}{6} + 1 = 1 \\ F(-1) &= -\frac{1}{6} + 1 = \frac{5}{6} \end{aligned}$$

Hence

$$P(-1 < X < 0) = F(0) - F(-1) = 1 - \frac{5}{6} = \frac{1}{6}$$

5.4.5 RELATIONSHIP BETWEEN DISTRIBUTION AND DENSITY FUNCTIONS

From Distribution Function to Density Function

For a continuous random variable, if the cumulative distribution function is known, the computation of probability density function boils down to differentiation.

Theorem 5.5

Suppose X is a continuous random variable. If the cumulative distribution function $F(x)$ is known, then

$$f(x) = F'(x)$$

The reader will be asked to prove this theorem in Exercise 5.17.

Note

The derivative of $F(x)$ is $f(x)$ at all points where $f(x)$ is continuous, which is everywhere except at the point $x = x_0$ where $f(x)$ is discontinuous.

Example 5.13
A random variable X has a cumulative distribution function:

$$F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{x}{10}, & 0 \leq x < 10 \\ 1, & x \geq 10 \end{cases}$$

Find the probability density function of the random variable X.

Solution

$$f(x) = F'(x)$$

That is,

$$f(x) = F'(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{10}, & 0 \le x < 10 \\ 0, & x \ge 10 \end{cases}$$

or

$$f(x) = \begin{cases} \dfrac{1}{10}, & 0 < x < 10 \\ 0, & \text{elsewhere} \end{cases}$$

From Density Function to Distribution Function

For a continuous random variable if p.d.f. is known the computation of c.d.f. boils down to integration (see Theorem 5.4).

Example 5.14
A random variable X has the following probability density function

$$f(x) = \begin{cases} x+1, & -1 \le x < 0 \\ 1-x, & 0 \le x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find the cumulative distribution function.

Solution

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt$$

If $x < -1$, then $f(x) = 0$, consequently,

$$F(x) = 0$$

If $-1 \le x < 0$,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-1}^{x} (t+1)\, dt = \frac{(x+1)^2}{2}$$

If $0 \leq x < 1$ then

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,dt \\
&= \int_{-1}^{0} (t+1)\,dt + \int_{0}^{x} (1-t)\,dt \\
&= 1 - \frac{(1-x)^2}{2}
\end{aligned}
$$

If $x \geq 1$,

$$
\begin{aligned}
F(x) &= \int_{-1}^{0} (t+1)\,dt + \int_{0}^{1} (1-t)\,dt \\
&= 1
\end{aligned}
$$

Thus, the required cumulative distribution function is

$$F(x) = \begin{cases} 0, & x < -1 \\ \dfrac{(x+1)^2}{2}, & -1 \le x < 0 \\ 1 - \dfrac{(1-x)^2}{2}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

5.5 CONDITIONAL PROBABILITY OF RANDOM VARIABLE

Conditional probability of events was introduced and discussed in Chapter 4. The concept can easily be extended to random variables so we shall not go into the theory but illustrate with the following example.

Example 5.15
Consider a random variable X with probability density function

$$f(x) = \begin{cases} \dfrac{1}{30}, & 0 < x < 30 \\ 0, & \text{elsewhere} \end{cases}$$

Find

(a) $P(X > 25|X > 15)$ (b) $P(X < 20|X > 15)$
(c) $P(X > 15|X < 22)$ (d) $P(X < 13|X < 18)$

Solution

$$\begin{aligned} P(X > 25|X > 15) &= \frac{P(X > 25,\, X > 15)}{P(X > 15)} \\ &= \frac{P(X > 25)}{P(X > 15)} \\ &= \frac{\displaystyle\int_{25}^{30} \frac{1}{30}\, dx}{\displaystyle\int_{15}^{30} \frac{1}{30}\, dx} \\ &= \frac{5/30}{15/30} = \frac{1}{3} \end{aligned}$$

The reader will be asked to solve the remaining questions in Exercise 5.28.

The probability function and the probability distribution are complete characterisation of the probabilistic behaviour of a random variable. Nevertheless, there are other useful though relatively weak characterisation. These are the means and variances which give an indication of location and width or dispersion of the distributions. These are the topics of discussion in the next chapter.

EXERCISES

5.1 Two fair coins are tossed once. Let X represent the number of tails.
a) Find the probability distribution of X;
b) Construct a probability graph;
c) Find the distribution function of X and,
d) Sketch the graph.

5.2 A fair die is thrown once. Let X represent the number that show up.
a) Find the probability distribution of X;
b) Construct a probability graph;
c) Find the distribution function of X and,
d) Sketch the graph.

5.3 A discrete function is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/9 & 0 \leq x < 2 \\ 1/3 & 2 \leq x < 6 \\ 7/9 & 6 \leq x < 8 \\ 1 & x \geq 8 \end{cases}$$

a) Verify that it is a distribution function of some random variable;
b) Find its probability function;
c) Construct the graphs of $F(x)$ and the probability function $p(x)$.

5.4 Refer to Exercise 5.3. Compute

(a) $P(1 < X < 6)$ (b) $P(X \leq 5)$
(c) $P(4 < X \leq 8)$ (d) $P(0 \leq X < 3)$
(e) $P(X \geq 4)$ (f) $P(-2 \leq X \leq 7)$

5.5 A discrete function is given by

$$p(x) = \begin{cases} \dfrac{x-2}{k}, & x = 2, 3, 5 \\ 0, & \text{elsewhere} \end{cases}$$

a) Find the constant k for which $p(x)$ is a probability function;
b) Find the corresponding distribution function;
c) Construct the graphs of $p(x)$ and the c.d.f. $F(x)$.

5.6 Refer to Exercise 5.5. Compute

(a) $P(X \geq 3)$ (b) $P(X \leq 5)$
(c) $P(2 \leq X \leq 5)$ (d) $P(2 \leq X \leq 3)$
(e) $P(X \leq 1)$ (f) $P(-2 \leq X \leq 1)$

5.7 A discrete random variable X has the distribution given by

$$P(X = x) = \begin{cases} a\left(\frac{1}{3}\right)^{x-1}, & x = 1, 2 \\ 0, & \text{elsewhere} \end{cases}$$

a) Determine the value of a; (b) $P(X = 2)$.

5.8 A discrete function is given by

$$P(X=x)=\begin{cases}\frac{x}{2}-1, & x=1,2,3,4\\ 0, & \text{elsewhere}\end{cases}$$

Verify whether the function is a p.m.f.

5.9 A continuous random variable X is such that

$$f(x)=\frac{a}{1+x^2}, \qquad -\infty<x<\infty$$

Find the constant a, such that $f(x)$ is a probability density function.

5.10 Two fair dice are thrown once. Let X represent the sum of the results.
a) Find the probability distribution of X;
b) Construct a probability graph;
c) Find the distribution function of X and,
d) Sketch the graph.

5.11 A continuous random variable X is such that

$$f(x)=\begin{cases}e^{-ax}, & x\geq 0\\ 0, & x<0\end{cases}$$

where a is a positive constant.

a) Find the constant a such that $f(x)$ is a probability density function;
b) Sketch the graph of $f(x)$;
c) Find the cumulative function of X and sketch its graph.

5.12 Let X have the function given by

$$f(x)=\begin{cases}3(1-x)^2, & 0\leq x<1\\ 0, & \text{elsewhere}\end{cases}$$

a) Determine whether $f(x)$ is a probability distribution function
b) If $f(x)$ is a p.d.f. find the corresponding distribution function.

5.13 A random variable X has probability density function

$$f(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{2}x, & 0 \le x \le 2 \\ 0, & x > 2 \end{cases}$$

a) Find the cumulative density function of the random variable
b) Draw its graph.
c) Compute $P(X \le 2)$
d) Compute $P(X \le 2 | 1 \le X \le 3)$

5.14 Consider the function

$$f(x) = \begin{cases} A\,x^3(1-x), & 0 \le x < 1, \quad A > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Find the value of $\mathcal{A}$ for which $f(x)$ is a p.d.f.

5.15 Consider the function

$$f(x) = \begin{cases} A\,x^3(2x - x^3), & 0 \le x < \frac{5}{2}, \\ 0, & \text{elsewhere} \end{cases}$$

Find the value of $\mathcal{A}$ for which $f(x)$ is a p.d.f.

5.16 A continuous random variable X has p.d.f.

$$f(x) = \begin{cases} \dfrac{k}{x^2}, & \text{if} \quad 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Find the value of the constant k;

5.17 The cumulative distribution function of a continuous random variable X is

$$F(x) = \begin{cases} 1 - e^{-\alpha x}, & x \ge 0, \quad \alpha > 0 \\ 0, & x < 0. \end{cases}$$

Find

(a) the probability density function;
(b) $P(X > 3)$ (c) $P(X > 3)$ (d) $P(-2 < X \le 3)$

5.18 Suppose that the c.d.f. of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Find the p.d.f of X and sketch it

5.19 Prove Theorem 5.5.

5.20 Suppose X possesses the density function

$$f(x) = \begin{cases} kx, & 0 \leq x \leq 3 \\ 0, & \text{elsewhere.} \end{cases}$$

a) Find the value of k that makes $f(x)$ a probability density function
b) Find $F(x)$.
c) Graph $f(x)$ and $F(x)$.
d) Use $F(x)$ to find $P(1 < X \leq 3)$
e) Find $P(0.2 \leq X < 2.8)$

5.21 Let X be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} 3x^2, & \text{if} \quad 0 \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find

a) $F(x)$;
b) $P(0.12 \le X < 0.98)$; (c) $P(X > 0.5)$;
c) Graph both $f(x)$ and $F(x)$.

5.22 The length of time to failure (in hundreds of hours) for a certain transistor is a random variable X with distribution function given by

$$F(x) = \begin{cases} 1 - e^{-x^2}, & x \ge 0 \\ 0, & x < 0. \end{cases}$$

a) Show that $F(x)$ has the properties of a distribution function.
b) Find the probability density function $f(x)$.
c) Find the probability that the transistor operates for at least 200 hours.

5.23 A family on a University campus has a 150-gallon tank that is filled at the beginning of each week. The weekly demand of the family shows a relative frequency behaviour that increases steadily up to 100 gallons and then levels off between 100 and 150 gallons. If X denotes weekly demand in hundreds of gallons, the relative frequency of demand can be modelled by

$$f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 1, & 1 < x < 1.5 \\ 0, & \text{elsewhere} \end{cases}$$

a) Find $F(x)$
b) Find $P(0 \le X < 0.5)$
c) $P(0.5 < X \le 1.2)$
d) $P(X \ge 1 | X \le 1.4)$

5.24 As a measure of intelligence, mice are timed when going through a maize to reach a reward of food. The time (in seconds) required by any mouse is a random variable X with density function given by

$$f(x) = \begin{cases} \dfrac{b}{x^2}, & x \ge b \\ 0, & \text{elsewhere,} \end{cases}$$

where b is the minimum possible time needed to traverse the maize.

a) Show that $f(x)$ has the properties of a density function.
b) Find $F(x)$.
c) Find $P(X > b + c)$ for a positive constant c.

5.25 Let the distribution function of a random variable X be

$$F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{x}{8}, & 0 \le x < 2 \\ \dfrac{x^2}{16}, & 2 \le x < 4 \\ 1, & x \ge 4, \end{cases}$$

a) Find the density function of X.
b) Find $P(1 \le X < 3)$.
c) Find $P(X \ge 1.5)$
d) Find $P(X < 3)$

5.26 Let X have the density function given by

$$f(x) = \begin{cases} 0.5, & -1 \le x < 0 \\ 0.5 + cx, & 0 \le x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

a) Find c;
b) Find $F(x)$;
c) Sketch $F(x)$ and $f(x)$;
d) Use $F(x)$ in part (b) to find $F(-1)$, $F(0)$, and $F(1)$;
e) Find $P(0 \le X \le 0.5)$
f) $P(X > 0.5 | X > 0.1)$

5.27 The percentage of alcohol in a certain solvent is 100% where X may be regarded as a continuous random variable with probability density function

$$f(x) = \begin{cases} 4\,x^2(1-x), & 0 \le x < 1, \\ 0, & \text{elsewhere} \end{cases}$$

a) Obtain the distribution function of X;
b) Graph $F(x)$ and $f(x)$.
c) Use $F(x)$ in (b) to find $F(-1), F(0)$, and $F(1)$.

5.28 Refer to Example 5.22. Solve (b) to (d).

5.29 Refer to Exercise 5.14. Evaluate the probabilities

a) $P(X < \frac{1}{3})$
b) $P(\frac{1}{3} < X < \frac{2}{3})$
c) $P(X \geq \frac{2}{3})$

5.30 Refer to Exercise 5.27. Evaluate the probabilities

a) $P(X < \frac{1}{3})$
b) $P(\frac{1}{3} < X < \frac{2}{3})$
c) $P(X \geq \frac{2}{3})$

5.31 Refer to Exercise 5.22. Evaluate the probabilities

a) $P(X < \frac{1}{3})$
b) $P(\frac{1}{3} < X < \frac{2}{3})$
c) $P(X \geq \frac{2}{3})$

6 NUMERICAL CHARACTERISTICS OF RANDOM VARIABLES

6.1 INTRODUCTION

In the previous chapter, we considered the probability distributions of random variables. In a number of cases, we need to know much less about the random variable itself. We may merely want to have a general ideal about its behaviour. That is, we may want to give a general quantitative description of the random variable by obtaining a single value from its values and the corresponding probabilities of these values. Usually, the first task is to determine the location of the distribution, that is, a typical value to describe or summarise the entire set of values.

6.2 MEASURES OF LOCATION

Definition 6.1 MEASURE OF LOCATION

A measure of location is a single value located in a set of values which can be used to describe a distribution

A measure of location may be central or non-central. Synonyms of ***measures of central location*** are ***measures of central tendency*** or ***averages*** and synonyms of ***measures of non-central location*** are ***quantiles*** or ***fractiles.***

Two general approaches may be taken to compute a measure of central location, and these two approaches lead to two different classes of averages. The first approach is to define a mathematical operation which might yield a very useful summary value. This approach gives us what is known as a ***calculated average*** or ***mean'*** or ***mathematical expectation***. The second approach to compute a measure of central location is to describe a special place or location in the data and then try to find methods for giving a value for this location. Such averages are called "positionary averages". There are two such averages – the ***median*** and the ***mode***. These various measures of central location are discussed in Sections 6.3 to 6.5. The measures of "non-central" location in describing a distribution are called ***quantiles*** or ***fractiles***.

Density and distribution functions can be used to obtain simple descriptive measures of the location.

6.3 MODE

Definition 6.2 MODE

The mode or the modal value of a distribution is a value x_{mod} of the random variable at which the probability distribution function takes its maximum value

Mode of Distribution of Discrete Random Variable

For a discrete distribution, the mode is the value x_0 of the random variable which has the highest probability of occurrence.

Example 6.1
Toss a coin four times and define the random variable X as the number of heads. What is the mode of the distribution?

Solution
When a coin is tossed four times, the following are all the possible combinations:

$$\underbrace{\mathcal{HHHH}}_{4 \text{ heads}} \quad \underbrace{\mathcal{HHHT},\ \mathcal{HHTH},\ \mathcal{HTHH},\ \mathcal{THHH}}_{3 \text{ Heads}}$$
$$\underbrace{\mathcal{HHTT},\ \mathcal{HTHT},\ \mathcal{HTTH},\ \mathcal{THTH},\ \mathcal{THHT},\ \mathcal{TTHH}}_{2 \text{ heads}}$$
$$\underbrace{\mathcal{TTTH},\ \mathcal{TTHT},\ \mathcal{THTT},\ \mathcal{HTTT}}_{1 \text{ head}}, \quad \underbrace{\mathcal{TTTT}}_{\text{no head}}$$

There are 16 possible equally likely simple events. Hence

$$P(X=4)=\frac{1}{16}; \quad P(X=3)=\frac{4}{16}; \quad P(X=2)=\frac{6}{16};$$
$$P(X=1)=\frac{4}{16}, \quad P(X=0)=\frac{1}{16}$$

The value x_0 which has the highest probability of $\frac{6}{16}$ is 2 heads. Hence the mode is $X_{mod}=2$.

Mode of Distribution of Continuous Random Variable

For the case of a *continuous random variable,* the mode is obtained by differentiating the probability density function and equating the result to zero.

Example 6.2
The p.d.f. of a continuous random variable X is given by

$$f(x) = \begin{cases} 2xe^{-x^2}, & 0 \leq x < \infty \\ 0, & \text{otherwise} \end{cases}$$

Find the mode.

Solution
The derivative of $f(x)$ is

$$\begin{aligned} f'(x) &= 2e^{-x^2} - 4x^2e^{-x^2} \\ &= 2e^{-x^2}(1 - 2x^2) \end{aligned}$$

Equating $f'(x)$ to zero we obtain

$$2e^{-x^2}(1 - 2x^2) = 0$$

which yields the solution $x = \sqrt{\frac{1}{2}}$. Hence the mode is $\frac{1}{\sqrt{2}}$.

The mode may not exist for some distributions and there may also be more than one mode for a distribution. When a distribution has one mode it is said to be ***unimodal***, two modes is ***bimodal,*** three modes is ***trimodal***. In general if a distribution has many modes it is called ***multimodal***.

Example 6.3
Roll a die once. Let X be the number of spots that show up. Find the mode.

Solution
All the six possible numbers $1, 2, 3, 4, 5, 6$ have equal probabilities of $\frac{1}{6}$. Hence the mode does not exist.

6.4 MEDIAN

Definition 6.3 MEDIAN

The median is a value above and below which half the probability lies

Median of Distribution of Discrete Random Variable

Theorem 6.1

The median of the distribution of a discrete random variable (x_{med}) is a number x_0 such that

$$P(X \le x_0) \ge \frac{1}{2} \quad \text{and} \quad P(X \ge x_0) \ge \frac{1}{2}$$

Example 6.4
Refer to Example 6.1. Find the median for the number of heads.

Solution

$$\begin{aligned} P(X \le 2) &= \frac{11}{16} > \frac{1}{2} \\ P(X \ge 2) &= \frac{11}{16} > \frac{1}{2} \end{aligned}$$

Hence, the median is 2.

Note

a) From Theorem 6.1, the median is given in terms of the cumulative distribution and, hence, it can be readily determined from the graph of a cumulative distribution function.
b) The median may or may not exist for the case of a discrete random variable.
c) Theorem 6.1 may lead to two possible values, In such a case, we take the midpoint of the values as the median.

Example 6.5
Refer to Example 6.3. Find the median.

Solution
For $x_o = 3$

$$P(X \leq 3) = \frac{3}{6} = 0.5, \qquad P(X \geq 3) = \frac{4}{6} > 0.5$$

which satisfies Theorem 6.1. Therefore 3 may be considered a median of the distribution. However, for $x_o = 4$

$$P(X \leq 4) = \frac{4}{6} > 0.5, \qquad P(X \geq 4) = \frac{3}{6} = 0.5$$

Here too, Theorem 6.1 is satisfied which means that 4 may also be considered a median of the distribution.

As a matter of fact, any number in the range from 3 to 4 inclusive satisfies Theorem 6.1 and is a median. The midpoint of the values which is 3.5 is taken as the median.

Median of Distribution of Continuous Random Variable

For a continuous distribution, there is always in principle a single value x_0 such that $P(X \leq x_0) \geq \frac{1}{2}$, although it may be difficult to recover its numerical value.

Theorem 6.2

The median of the distribution of a continuous random variable, (x_{med}) is a number x_0 such that

$$\begin{aligned} P(X \leq x_0) &= F(x) = \frac{1}{2} \quad \text{and} \\ P(X \geq x_0) &= 1 - F(x) = \frac{1}{2} \end{aligned}$$

Proof

If $F(x_0) \geq \frac{1}{2}$ and $1 - F(x_0) \geq \frac{1}{2}$ then $F(x_0) \geq \frac{1}{2}$ and $F(x_0) \leq \frac{1}{2}$ which implies that $F(x_0) = \frac{1}{2}$.

Example 6.6
Refer to Example 6.2. Find the median.

Solution

$$\begin{aligned} P(X \leq x_0) \geq \frac{1}{2} &= F(x_0) \\ &= \int_0^{x_0} 2\,t\,e^{-t^2}\,dt \\ &= 1 - e^{-x_0^2} \end{aligned}$$

Now equating $F(x_0)$ to $\frac{1}{2}$ (by Theorem 6.2) we obtain

$$\begin{aligned} 1 - e^{-x_0^2} &= \frac{1}{2} \\ x_0 &= \sqrt{\ln 2} \end{aligned}$$

6.5 QUANTILES

The median is but one of a family of descriptive measures called ***quantiles*** or ***fractiles***.

Definition 6.4 QUANTILES

Let x_0 be a real number, then x_0 is a k quantile of the distribution of X if $P(X \leq x_0) \geq k$ and $P(X \geq x_0) \geq 1 - k$

Theorem 6.3

The k quantile of the distribution of $F(x)$ is that value x_k such that

$$F(x_k) = P(X \leq x_k) = k$$

The most familiar quantiles are when

a) $k = 0.5$, which corresponds, respectively, to the median;
b) $k = 0.25$; $k = 0.5$ and $k = 0.75$ which correspond, respectively, to the lower, middle (which is the median) and upper quartiles respectively;
c) $k = 0.1$; $k = 0.2$; $\cdots$ $k = 0.9$ which correspond, respectively, to the first, second, $\cdots$, ninth deciles;
d) $k = 0.01$; $k = 0.02; \cdots$, $k = 0.99$, which correspond respectively, to the first, second, $\cdots$, ninety-ninth percentiles.

Note

Like the median

a) quantiles can be readily obtained from the graph of a cumulative distribution function.
b) there may be more than one quantile for any particular value of k in the case of a discrete random variable.

Example 6.7
Refer to Example 6.2. Find the lower and upper quartiles.

Solution

a) From Example 15.20,

$$1 - e^{-x_p^2} = \frac{1}{4}$$

Hence the lower quartile is

$$x_p = \sqrt{\ln \frac{4}{3}}$$

from which

$$x_p = \sqrt{\ln \frac{4}{3}}$$

The upper quartile is

$$\begin{aligned} 1 - e^{-x_p^2} &= \frac{3}{4} \\ x_p &= \sqrt{\ln 4} \end{aligned}$$

We have just discussed the mode, the median and the quantiles of a distribution as simple descriptive measures of the centre. The other measure of the centre is the mean value known as mathematical expectation in probability theory. Some other single values of particular importance in Statistics are variance and moments of various orders. We shall discuss the moments in the next chapter.

6.6 MATHEMATICAL EXPECTATION

6.6.1 DEFINITION OF MATHEMATICAL EXPECTATION

The term “expectation” is used here in a special, statistical sense and not as we might understand it in ordinary everyday sense. We introduce the concept of mathematical expectation with an example.

Example 6.8

Suppose there are 1,000 students in a school. Suppose also that on the first day of re-opening, 600 of them can spend Gh¢ 5 each[19]; 200 students can spend Gh¢ 50 each; 150 students can spend Gh¢ 100 each; 40 students can spend Gh¢ 200 each; and 10 students can spend Gh¢ 500 each. What is the expected expenditure of students on a re-opening day?

Solution

The mean amount that could be spent on the re-opening day is found if the total sum

$$\text{Gh¢ } 41,000 \ (= 600 \times 5 + 200 \times 50 + 150 \times 100 + 40 \times 200 + 10 \times 500)$$

is divided by the number of students $(= 1,000)$. We write this as

$$5 \times \frac{600}{1,000} + 50 \times \frac{200}{1,000} + 100 \times \frac{150}{1,000}$$
$$+200 \times \frac{40}{1,000} + 500 \times \frac{10}{1,000} = 41,000$$

Now, let us define a random variable X to represent the amount a student spends on a re-opening day which has the values

$$\text{Gh¢ } 5, \quad \text{Gh¢ } 50, \quad \text{Gh¢ } 100 \quad 200 \quad \text{and} \quad \text{Gh¢ } 500$$

The corresponding probabilities of spending $P(X = x)$ are

$$\frac{600}{1,000} = 0.6, \quad \frac{200}{1,000} = 0.2, \quad \frac{150}{1,000} = 0.15, \quad \frac{40}{1,000} = 0.04 \quad \frac{10}{1,000} = 0.01$$

Therefore, the expected expenditure of students on a re-opening day is equal to the sum of the product of the value of the expenditure and their corresponding probabilities, namely,

$$(5 \times 0.6) + (50 \times 0.2) + (100 \times 0.15) + (200.04) + (500 \times 0.01) = 41,000$$

We shall now give a general definition of the mathematical expectation of a random variable.

Mathematical Expectation of Discrete Random Variable

Definition 6.5 MATHEMATICAL EXPECTATION OF DISCRETE VARIABLE

Let $x_1, x_2, \cdots$, be the range of a discrete random variable X which assumes the value of x_i with the probability $p(x_i), i = 1, 2 \cdots$, then the mathematical expectation of X is given by

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

provided that the finite series converges

If the infinite series diverges, the corresponding mathematical expectation does not exist.

Note

a) The mathematical expectation of a random variable X is often simply referred to as the *expectation* or the *expected value* of X. It is usually used for the mean of probability distribution as the average value in the long run.
b) The expectation of a random variable takes the unit of measurement of that random variable. For example, if the random variable assumes its values in kilometre, its expectation is also in kilometre.

Example 6.9
Given the following data, find $E(X)$.

x_i	0	1	2	3
$p(x_i)$	1/8	3/8	3/8	1/8

Solution

x_i	0	1	2	3	Total
$p(x_i)$	1/8	3/8	3/8	1/8	1
$x_i p(x_i)$	0	3/8	6/8	3/8	12/8

Hence,

$$E(X) = \sum_{i=1}^{4} x_i p(x_i) = 0 + \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \frac{3}{2}$$

Before we discuss the expectation of a continuous random variable, we shall explain how the arithmetic mean of a finite set of real numbers is an example of an “expectation”.

According to the relative frequency definition of probability (Definition 3.29),

$$p(x_i) = P(X = x_i) = \frac{f_i}{N}$$

where f_i is the absolute frequency of a given value in the population of size N.

Now, from Definition 6.1

$$E(X) = \sum_{i=1}^{n_i} x_i p(x_i) = \sum_{i=1}^{n_i} \left[x_i \left(\frac{f_i}{N} \right) \right] = \frac{1}{N} \sum_{i=1}^{n_i} f_i x_i$$

where $\sum n_i = N$.

This is the formula of the arithmetic mean for grouped data. When each value of X is equally likely and is separately listed, all $f's$ are equal to 1, and

$$E(X) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

which is the formula of the arithmetic mean for ungrouped data.

Mathematical Expectation of Continuous Random Variable

Definition 6.6 MATHEMATICAL EXPECTATION OF CONTINUOUS VARIABLE

Suppose X is a continuous random variable, then the mathematical expectation is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Similar to the discrete case, the integral may or may not converge and the expectation may or may not exist.

Example 6.10
A random variable X has the p.d.f

$$f(x) = \begin{cases} \frac{1}{8}x, & 0 \leq x \leq 4 \\ 0, & \text{elsewhere} \end{cases}$$

Find the expectation of X.

Solution

$$E(X) = \int_0^4 x\ \frac{1}{8}\ x dx = \frac{1}{8}\int_0^4 x^2 dx = \frac{1}{8}\left[\frac{x^3}{3}\right]_0^4 = \frac{8}{3}$$

6.6.2 PROPERTIES OF MATHEMATICAL EXPECTATION

In discussing the properties of mathematical expectation, we shall assume that the expectation itself exists. The expectation $E(X)$ is said to exist (that is, it is finite or $E(X) < \infty$) if X is a bounded random variable[20].

Property 1 Expectation of a Constant

Theorem 6.4

The expectation of a constant is the constant, that is, if c is a constant, then

$$E(c) = c$$

Proof
The constant value c may be considered as random which takes only one value c with probability 1. So its expectation is:

$$E(c) = c \times 1 = c$$

Property 2 **Expectation of Product of Constant and Variable**

Theorem 6.5

If c is a constant and X is a random variable, then

$$E(cX) = cE(X)$$

The proof of this theorem is left to the reader, see Exercise 6.3. The theorem means that if we change the units of measurement of X, then the expectation of X changes in the same way as X.

Example 6.11
For the data in Example 6.9, find the expectation for $2X$ and verify that it is equal to $2E(X)$.

Solution

$2x_i$	0	2	4	6	Total
$p(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

$$\begin{aligned} E(2X) &= \sum_{i=1}^{4} 2x_i p(x_i) \\ &= 0\left(\frac{1}{8}\right) + 2\left(\frac{3}{8}\right) + 4\left(\frac{3}{8}\right) + 6\left(\frac{1}{8}\right) \\ &= 0 + \frac{6}{8} + \frac{12}{8} + \frac{6}{8} = 3 \end{aligned}$$

But from Example 6.9,

$$E(X) = \frac{3}{2}$$

so that

$$2E(X) = 2\left(\frac{3}{2}\right) = 3$$

Hence

$$E(2X) = 2E(X)$$

Property 3 **Expectation of Linear Functions**

Theorem 6.6

If a and b are constant, and X is a random variable, then

$$E(aX + b) = aE(X) + b)$$

Proof

Suppose X is a discrete random variable. Then

$$
\begin{aligned}
E(aX+b) &= \sum_x (ax+b)p(x) \\
&= \sum_x a\,x\,p(x) + \sum_x bp(x) \\
&= a\sum_x x\,p(x) + b\sum_x p(x) \\
&= aE(X)+b \quad (\text{since by Definition 5.6, } \sum_x p(x) = 1)
\end{aligned}
$$

If X is continuous

$$
\begin{aligned}
E(aX+b) &= \int_{-\infty}^{\infty} (ax+b)f(x)dx \\
&= a\int_{-\infty}^{\infty} xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx \\
&= aE(X)+b \quad (\text{since by Definition 5.7, } \int_{-\infty}^{\infty} f(x)dx = 1)
\end{aligned}
$$

Example 6.12
Suppose the distribution of X is given by the following table:

x	-2	4
$p(x)$	0.4	0.6

If $a = 3$ and $b = 2$, verify that Theorem 6.6 is valid.

Solution
If $a = 3$ and $b = 2$, then we are required to show that

$$E(3X + 2) = 3E(X) + 2$$

Now,

x	$3(-2) + 2$	3(4) + 2
$p(x)$	0.4	0.6

x	-4	14
$p(x)$	0.4	0.6

$$E(3X + 2) = (-0.4)(0.4) + 14(0.6) = 6.8$$

But

$$E(X) = (-2)(0.4) + 4(0.6) = 0.8 + 2.4 = 1.6$$

and

$$3E(X) + 2 = 3(1.6) + 2 = 6.8$$

Hence Theorem 6.6 is valid.

Theorem 6.6 states that the expectation of a linear function of a random variable is that same linear function of the expectation. This is not generally true for all functions unless a linear function is involved. It can be illustrated that

a) $E(X^2) \neq [E(X)]^2$
b) $E(\ln X) \neq \ln E(X)$
c) $E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)}$

Example 6.13
Given the following table:

x	-1	1
$p(x)$	$\frac{1}{2}$	$\frac{1}{2}$

Verify whether or not $E(X^2) = [E(X)]^2$.

Solution

$$E(X) = (-1)\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = 0$$

Now,

$$E(X^2) = (-1)^2\left(\frac{1}{2}\right) + (1)^2\left(\frac{1}{2}\right) = 1$$

Thus,

$$E(X^2) = 1 \neq [E(X)]^2 = 0$$

***Property 4* Nonnegative of Expectation**

Theorem 6.7

Suppose X is a random variable. If $X \geq 0$, then $E(X) \geq 0$

Property 5 **Modulus Inequality of Expectation**

Theorem 6.8

If X is a random variable, then

$$|E(X)| \leq E(|X|)$$

Property 6 **Lyapunov Inequality of Expectation**

Theorem 6.9

Suppose X is a random variable. If $0 < a < b < \infty$, then

$$[E(|X|^a)]^{\frac{1}{a}} \leq [E(|X|^b)]^{\frac{1}{b}}$$

***Property 7* Expectation of Deviation**

Theorem 6.10

The expectation of the deviation of the random variable X from its expectation is zero:

$$E(X - \mu) = 0$$

where $\mu = E(X)$

Proof

From Theorem 6.4 and 6.6

$$\begin{aligned} E[X - E(X)] &= E(X - \mu) \\ &= E(X) - E(\mu) = \mu - \mu = 0 \end{aligned}$$

6.7 VARIANCE

6.7.1 DEFINITION OF VARIANCE

The expected value of a random variable X is its average value which can be viewed as an indication of the central value of the density or frequency function. The expected value is, therefore, sometimes referred to as the ***location parameter***. To describe the behaviour of a random variable adequately we also need to have some idea of how the values are dispersed about the mean. One such measure of dispersion is the variance of the random variable, denoted by $Var(X)$ or σ^2.

Definition 6.7 VARIANCE

The variance of a random variable X is the expectation of the square of the deviation of the random variable from its expected value:

$$\sigma^2 = Var(X) = E[X - E(X)]^2$$

or

$$\sigma^2 = Var(X) = E(X - \mu)^2$$

The unit of measurement of the variance is the unit of the random variable squared so the frequently considered measure of dispersion is the standard deviation which has the same unit of measurement as the random variable itself.

Definition 6.8 STANDARD DEVIATION

The positive square root of the variance, denoted as σ, is called the standard deviation:

$$\sigma = \sqrt{Var(X)}$$

Even though the standard deviation is the one that is used to measure how dispersed the probability distribution is about its mean, of how spread out on the average are the values of the random variable about its expectation, the dispersion is usually defined in terms of variance.

Theorem 6.11

If the expectation of the random variable X exists, then

$$Var(X) = E(X^2) - [E(X)]^2$$

or

$$Var(X) = E(X^2) - \mu^2$$

where $\mu = E(X)$

Proof
Using Definition 6.3 and Theorem 6.6, we have[21]

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) + E(-2\mu X) + E(\mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Variance of Discrete Random Variables

Definition 6.9 VARIANCE
(**Discrete Variable**)

Let X be a discrete random variable which takes on values $x_1, x_2, \cdots$ with respective probabilities $p(x_1), p(x_2), \cdots$ Then

$$Var(X) = \sum_i (x_i - \mu)^2 p(x_i)$$

where $\mu = E(X)$

Example 6.14
For the data in Example 6.9, find

a) $Var(X)$;
b) the standard deviation of X.

Solution

From Example 6.9, $E(X) = \mu = \frac{3}{2}$. Now,

x	0	1	2	3	Total
$p(x)$	1/8	3/8	3/8	1/8	1.0
$x - 3/2$	$-3/2$	$-1/2$	1/2	3/2	0
$(x - 3/2)^2$	2.25	0.25	0.25	2.25	5.0
$(x - 3/2)^2 p(x)$	0.28125	0.09375	0.09375	0.28125	0.75

From the table, $Var(X) = 0.75$.

Hence,

$$\sigma = \sqrt{0.75} = 0.866$$

As we have realized, it is cumbersome to use the formula in Definition 6.9 to calculate the variance. For computational purposes, it is advisable to use Theorem 6.8 which leads to the following theorem.

Theorem 6.12

If the expectation of a discrete random variable exists, then

$$Var(X) = \sum_{i=1}^{n} x_i^2 p(x_i) - \left[\sum_{i=1}^{n} x_i p(x_i)\right]^2$$

Proof

$$\begin{aligned}
Var(X) &= \sum_i (x_i - \mu)^2 p(x_i) \quad \text{(from Definition 6.5)} \\
&= \sum_i [x_i^2 p(x_i) - 2\mu x_i p(x_i) + \mu^2 p(x_i)] \\
&= \sum_i x_i^2 p(x_i) - 2\mu \sum_i x_i p(x_i) + \mu^2 \sum_i p(x_i) \\
&= \sum_i x_i^2 p(x_i) - 2\mu^2 + \mu^2 \quad \text{(since } \sum_i p(x_i) = 1)
\end{aligned}$$

from which the result follows.

Example 6.15
Work out Example 6.14 using Theorem 6.12.

Solution

x	$p(x)$	$xp(x)$	$x^2p(x)$
0	1/8	0	0
1	3/8	3/8	3/8
2	3/8	6/8	12/8
3	1/8	3/8	9/8
Total	1	3/2	3

Hence

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} x_i^2 p(x_i) - \left[\sum_{i=1}^{n} x_i p(x_i)\right]^2 \\ &= 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4} \end{aligned}$$

Variance of Continuous Random Variables

Definition 6.10 VARIANCE
(**Continuous Variable**)

Suppose X is a continuous random variable, then

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Example 6.16
Refer to Example 6.10. Find $Var(X)$.

Solution
From Example 6.10, $E(X) = \frac{8}{3}$.

Now

$$E(X^2) = \int_0^4 x^2 \frac{1}{8} x^2 \, dx = \frac{1}{8}\left[\frac{x^4}{4}\right]_0^4 = 8$$

Hence

$$Var(X) = E[X^2] - [E(X)]^2 = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}$$

6.7.2 PROPERTIES OF VARIANCE

Property 1 **Variance of a Constant**

> **Theorem 6.13**
>
> Suppose $c > 0$ is a constant, then
>
> $$Var(c) = 0$$

Proof

$$\begin{aligned} Var(c) &= E[c - E(c)]^2 \\ &= E(c - c)^2 \quad \text{(by Theorem 6.1)} \\ &= 0 \end{aligned}$$

That is, a constant has no variability or dispersion.

Theorem 6.14

If a is a constant and X is a random variable, then

$$Var(aX) = a^2 Var(X)$$

Proof
Applying Definition 6.3 and Theorem 6.2

$$\begin{aligned} Var(aX) &= E[aX - E(aX)]^2 \\ &= E[a\{X - E(X)\}]^2 \\ &= E\{a^2[X - E(X)]^2\} \\ &= a^2 E[X - E(X)]^2 \end{aligned}$$

But by Definition 6.3, the expression $E[X - E(X)]^2$ is the variance of a random variable X. Hence

$$Var(a\,X) = a^2 Var(X)$$

Example 6.17
For the data in Example 6.2, calculate $Var(2X)$ and check whether Theorem 6.11 is satisfied.

Solution
From Example 6.7, $Var(X) = 0.75$.
We shall now calculate $Var(2X)$ from the following table.

x	$2x$	$(2x)^2$	$p(x)$	$(2x)^2 p(x)$
	0	0	1/8	0
	2	4	3/8	12/8
	4	16	3/8	48/8
	6	36	1/8	36/8
Total				$96/8 = 12$

From Example 6.4, $E(2X) = 3$,
so that

$$[E(2X)]^2 = 9$$

And from the table above,

$$E[(2X)^2] = \sum_{x=1}^{n} (2x)^2 p(x) = 12$$

so that

$$Var(2X) = E[(2X)^2] - [E(2X)]^2 = 12 - 9 = 3$$

Hence, Theorem 6.11 is satisfied. Thus,

$$Var(2X) = 2^2 Var(X) = 4 \times 0.75 = 3$$

Property 3 **Variance of Linear Function**

Theorem 6.15

If a and b are constants, then for any random variable X,

$$Var(aX + b) = a^2 Var(X)$$

Proof

Using the definition of variance (Definition 6.3) and Theorems 6.2 and 6.3

$$\begin{aligned} Var(aX+b) &= E([aX + b - E(aX+b)]^2) \\ &= E([aX + b - aE(X) - b]^2) \\ &= E[a^2\{X - E(X)\}^2] \\ &= a^2 E[\{X - E(X)\}^2] \\ &= a^2 Var(X) \end{aligned}$$

Example 6.18

Refer to Example 6.12. Verify whether Theorem 6.12 is valid or not.

Solution

From Example 6.12, $a = 3,\ b = 2$. Let $Y = aX + b$.

$$Var(Y) = E[(aX+b)^2] - [E(aX+b)]^2$$

We find $E[(3X+2)^2]$

$(3x+2)^2$	$[3(-2)+2]^2$	$[3(4)+2]^2$
$p(x)$	0.4	0.6

$(3x+2)^2$	16	196
$p(x)$	0.4	0.6

$$E\left[(3X+2)^2\right] = 16(0.4) + 196(0.6) = 124$$

But from Example 6.5,

$$E[(3X+2)] = 6.8$$

Hence

$$Var(3X+2) = 124 - (6.8)^2 = 77.76$$

But by Theorem 6.14

$$Var(3X+2) = 3^2 Var(X)$$

Now,

$$Var(X) = E(X^2) - [E(X)]^2$$

x^2	$(-2)^2$	16
$p(x)$	0.4	0.6

$$E(X^2) = 4(0.4) + 16(0.6) = 1.6 + 9.6 = 11.2$$

From Example 6.12, $E(X) = 1.6$, so that

$$Var(X) = 11.2 - (1.6)^2 = 8.64$$

Hence

$$Var(3X+2) = 9(8.64) = 77.76$$

Theorem 6.15 is, therefore, valid.

EXERCISES

6.1 Refer to Exercise 5.16. Find
a) the mode;
b) the median;
c) the lower and upper quartiles.

6.2 Refer to Exercise 5.14. Find
(a) the mode; (c) the median; (d) the 8^{th} decile.

6.3 Refer to Exercise 5.21. Find
(a) the mode; (c) the median; (d) the 8^{th} decile.

6.4 Refer to Exercise 5.27. Without the use of a graph, find
a) the mode,
b) the median,
c) the lower and upper quartiles,
d) the eighth decile.

6.5 The probability distribution of a random variable X is represented in the table below.

X	-2	-1	0	0	4
$P(X=x)$	0.1	0.2	0.15	0.25	0.3

Find

(a) $E(X)$ (b) $E(6X)$ (c) $E(\frac{1}{2}X)$ (d) $E(X^2)$
(e) $Var(X)$ (f) $Var(6X)$ (g) $Var(\frac{1}{2}X)$ (h) $Var(X^2)$

6.6 Given the probability distribution of X in the table below:

X	1	2	3	4	5
$P(X=x)$	0.1	0.3	0.2	0.1	0.3

Find

(a) $E(X)$ (b) $E(6X)$ (c) $E(\frac{1}{2}X)$ (d) $E(X^2)$
(e) $Var(X)$ (f) $Var(6X)$ (g) $Var(\frac{1}{2}X)$ (h) $Var(X^2)$

6.7 Prove Theorem 6.2.

6.8 Refer to Exercise 5.5. Find (a) $E(X)$ (b) $Var(X)$.

6.9 Refer to Exercise 5.2. Find (a) $E(X)$ (b) $Var(X)$.

6.10 Refer to Exercise 5.9. Find (a) $E(X)$ (b) $Var(X)$.

6.11 If

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases}$$

determine $E(X^k)$

6.12 Refer to Exercise 5.27. Find (a) $E(X)$ (b) $Var(X)$.

6.13 Refer to Exercise 5.14. Find (a) $E(X)$ (b) $Var(X)$.

7 MOMENTS AND MOMENT-GENERATING FUNCTIONS

7.1 INTRODUCTION

In Chapter 6, we discussed the expected value, μ, and the standard deviation, σ, as useful numerical descriptive measures for respectively, locating the centre and describing the spread of the probability distribution $p(x)$. They however, do not provide a unique characterization of the distribution because a number of different distributions may have the same means and standard deviations.

In this chapter, we shall consider a set of numerical descriptive measures that under rather general conditions, uniquely determine the general probability distribution $p(x)$. These measures can be defined in terms of "moments" of a probability distribution of a random variable.

7.2 TYPES OF MOMENTS

7.2.1 DEFINITION OF MOMENTS

Definition 7.1 MOMENTS

Moments of a random variable X are the expectation of different k powers $(k = 1, 2, \cdots)$ of the random variable when this expectation exists

Strictly speaking, moments are associated with the distribution of X rather than with X itself. Thus, though we may be speaking of the moment of X, we mean the moment of the distribution of X.

There are, basically, three types of moments, namely,

a) Moment about the origin,
b) Moment about the mean, and
c) Moment about a point.

There are also what we call a *factorial moment* and an *absolute moment*. We should note that, unlike the moments about the origin, about the mean and about a point, the factorial moment in particular, may have theoretical application in advanced statistical theory but is of less practical use.

7.2.2 MOMENT ABOUT THE ORIGIN

The most basic moment of a random variable X is the one in relation to the origin of the coordinate.

Definition 7.2 MOMENT ABOUT ORIGIN

The k^{th} moments of a random variable X about the origin is defined to be the expectation $E(X^k)$

The k^{th} moment about the origin is also called the ***ordinary moment*** and denoted by μ_k'. In our discussion, we shall always assume that the moment exists.

Theorem 7.1

The moments μ_k' is said to exist if

$$E(|X|^k) < \infty$$

The moment about the origin of the order zero always exists and equals 1.

Definition 7.3 MOMENT ABOUT ORIGIN
(Discrete Case)

The k^{th} moment about the origin of the distribution of the discrete random variable X whose probability mass function is $p(x)$ is given by

$$\mu_k' = E(X^k) = \sum_{i=1}^{\infty} x_i^k p(x_i)$$

Definition 7.4 MOMENT ABOUT ORIGIN
(Continuous Case)

The k^{th} moment about the origin of the distribution of the continuous random variable X whose probability density function is $f(x)$ is given by

$$\mu_k' = E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx$$

When $k = 1$

$$\mu'_1 = \sum x_i p(x_i) = E(X) \text{ (discrete case)}$$

$$\mu'_1 = \int_{-\infty}^{\infty} x f(x) dx = E(X) \text{ (continuous case)}$$

That is, the first moment about the origin is the mean of the distribution. Since the moment, μ'_1, occurs so often in Probability and Statistics, it is given the symbol μ; that is,

$$E(X) = \mu'_1 = \mu$$

7.2.3 MOMENT ABOUT THE MEAN

Before giving the definition of the moment about the mean, we shall introduce the concept of *centred random variable*. Consider a random variable X with an expectation μ.

Definition 7.5 CENTRED RANDOM VARIABLE

A centred random variable X^* corresponding to a random variable X is the deviation of the random variable X from its expectation:

$$X^* = X - \mu$$

Theorem 7.2

The expectation of the centred random variable X^* is equal to zero:

$$E(X^*) = E(X - \mu) = 0$$

Proof
See proof of Theorem 6.7.

The centred random variable is obviously equivalent to moving the origin of the co-ordinate to the mean (central point) along the horizontal axis. The moment of the centred random variable is called the ***central moment*** or the ***moment about the mean***.

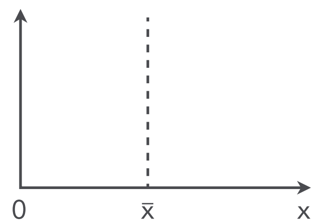


Fig. 7.1 Centred Random Variable X^*

Definition 7.6 CENTRAL MOMENT

The k^{th} central moment of a random variable X is defined as

$$E[(X - \mu)^k]$$

Definition 7.7 CENTRAL MOMENT
(Discrete Case)

The k^{th} central moment of a discrete random variable X whose probability mass function is $p(x_i)$ is given by

$$\mu_k = \sum_{i=1}^{\infty} (x-\mu)^k p(x_i)$$

Definition 6.14 CENTRAL MOMENT
(Continuous Case)

The k^{th} central moment of a continuous random variable X whose probability density function is $f(x)$ is given by

$$\mu_k = \int_{-\infty}^{\infty} (x-\mu)^k f(x) dx$$

Uses of Central Moments

The k^{th} central moment plays a major role in statistics. We shall derive the first three.

Theorem 7.3

The first central moment of a ransom variable X is zero

$$\mu_1 = E[(X-\mu)] = 0$$

Proof
See the proof of Theorem 6.7.

Theorem 7.4

The second central moment of a random variable X is given by

$$\mu_2 = E[(X-\mu)^2] = \mu_2' - \mu^2$$

where
$\mu_2' = E(X^2)$ is the second moment about the origin;
$\mu = E(X)$ is the first moment about the origin which is the mean

The second central moment, μ_2, is just the variance, $Var(X)$.

Proof

From Theorem 6.8

$$E[(X-\mu)^2] = E(X^2) - \mu^2 = \mu_2' - \mu^2$$

Theorem 7.5

The third central moment of a random variable X is given by

$$\mu_3 = E[(X-\mu)^3] = \mu_3' - 3\mu_2'\mu + 2\mu^3$$

Proof

$$\begin{aligned}
\mu_3 &= E[(X-\mu)^3] \\
&= E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] \\
&= E[X^3] - 3\mu E[X^2] + 3\mu^2 E[X] - \mu^3 \quad \text{(Using Theorems 6.1 and 6.3)} \\
&= \mu_3' - 3\mu\mu_2' + 3\mu^2\mu - \mu^3 = \mu_3' - 3\mu\mu_2' + 2\mu^3
\end{aligned}$$

Expressions for $\mu_4, \mu_5, etc.$ may be obtained similarly.

Central moments have an advantage over the other moments: the first central moment, as has been seen earlier (Theorem 6.14), is always equal to zero and the second central moment is the minimum value of the second moment of a random variable about any arbitrary point (see Theorem 6.18 in the sequel).

7.2.4 MOMENT ABOUT ANY POINT

In general, moments may be considered not only in relation to the origin of the coordinate (moment about origin) or in relation to the expectation (central moments) but also in relation to any arbitrary point a.

Definition 7.8 MOMENT ABOUT ANY POINT

The k^{th} moment of X about an arbitrary point a is

$$\mu_k^* = E[(X-a)^k]$$

Theorem 7.6

The second moment about an arbitrary point a,
$\mu_2^* = E[(X-a)^2]$, has the minimum value if and only if $a = \mu$

Proof

$$\begin{aligned}
\mu_2^* &= E[(X-a)^2] \\
&= E[(X-\mu+\mu-a)^2] \\
&= E[(X-\mu)^2 + 2(X-\mu)(\mu-a) + (\mu-a)^2] \\
&= E[(X-\mu)^2] + 2(\mu-a)E(X-\mu) + E((\mu-a)^2) \\
&= \mu_2 + (\mu-a)^2 \quad (\text{since by Theorem 6.14}, \;\; E(X-\mu)=0)
\end{aligned}$$

Obviously, this value reaches its minimum when $\mu = a$.

7.2.5 FACTORIAL AND ABSOLUTE MOMENTS

Definition 7.9 FACTORIAL MOMENT

For a positive integer k, the factorial moment of X of order k is given by

$$m_f = E[X(X-1)\cdots(X-k+1)]$$

Definition 7.10 ABSOLUTE MOMENT

An absolute moment of a random variable X of order k is defined as

$$m_a = E(|X|^k)$$

Theorem 7.7

If an absolute moment of a random variable X of order k exists, then all moments (ordinary, central, factorial and absolute) of order $r < k$ exist

7.3 USES OF MOMENTS

Moments have many uses in statistics.

Approximation of Distributions

Moments are used to approximate the probability distribution of a random variable (usually an estimator). Under some fairly general conditions, it can be used to show that two random variables X and Y have identical probability distributions.

Variance

The second central moment, μ_2, is frequently used in practice. It is denoted by the special symbol σ^2 and as noted earlier, is called the variance of the distribution which is used to determine the degree of concentration of the distribution about the mean μ. For central moments of higher than second order, interpretations in terms of shape may not be accurate. However, the third and the fourth moments are sometimes used in statistics.

Skewness

The third central moment, μ_3, can be used to determine the symmetry of a distribution.

Definition 7.11 SYMMETRIC DISTRIBUTION

A distribution of a random variable X is said to be symmetric if its frequency function $f(x)$ is such that, for all real x

$$f(\mu + x) = f(\mu - x)$$

where $\mu = E(X)$

It can be shown that for a symmetrical distribution, the odd number central moments vanish. This suggests that the odd number central moments of a distribution will measure its departure from symmetry, that is, its asymmetry.

The simplest of the odd central moment is the third central moment. Since all deviations are cubed, negative and positive deviations will tend to cancel each other, giving $\mu_3 = 0$, if the distribution is symmetrical about μ.

Theorem 7.8

If a distribution of a random variable X is symmetric, then

$$\mu_3 = E[(X - E(X)]^3 = 0$$

If the distribution is skewed to the right, then $\mu_3 > 0$. On the other hand, given a left (negatively) skewed distribution, we would have $\mu_3 < 0$. We should note that μ_3 alone is a rather poor measure of skewness since the size of the μ_3 is influenced by the units used to measure the values of X.

To make this measure dimensionless and thereby allowing us to compare the symmetry of two or more distributions whose units of measurement are different, we may form a relative measure by dividing μ_3 (the third moment about the mean) by the cube of the standard deviation (that is, skewness of the distribution relative to its degree of spread). The ratio which is often denoted a_3, is known as ***coefficient of skewness*** and is used to measure lack of symmetry.

Definition 7.12 SKEWNESS

The coefficient of skewness is given by

$$a_3 = \frac{\mu_3}{\sigma^3}$$

It can be shown that $|a_3| \leq 1$. The table below gives interpretation of the value of a_3.

$a_3 = 0$	Distribution is symmetric
$0 \leq \|a_3\| < 0.5$	Distribution is nearly symmetric
$0.5 \leq \|a_3\| < 1$	Distribution is moderately skewed
$a_3 \pm 1$	Distribution is highly skewed

The table, however, cannot be taken as a rule of thumb. It must be noted, however, that although, $a_3 = 0$ for symmetrical distributions, the converse is not true; it is possible for these quantities to vanish for non-symmetrical distributions. For this reason, the use of a_3 as a measure of lack of symmetry is limited.

Kurtosis

The fourth central moment $\mu_4 = E[(X-\mu)^4]$ is always non-negative and it can be used to measure the degree of peakedness (sharpness of the spike) of a unimodal p.d.f. The relative measure of *peakedness* is known as ***coefficient of kurtosis***.

Definition 7.13 KURTOSIS

The coefficient of kurtosis is given by

$$a_4 = \frac{\mu_4}{\sigma^4} - 3$$

Coefficient of kurtosis is used to measure the "peakedness" of a distribution of how "heavy" the tails of a distribution are. However, the interpretation of a_4 presents even more difficulties than those involved in the interpretation of coefficients of skewness.

Moment higher than μ_4 are only of theoretical interest and are usually difficult to interpret.

7.4 MOMENT-GENERATING FUNCTIONS

Even though the direct computation of the theoretical moments from Definition 7.1 may be easy, it is convenient for later theory to be able to calculate such moments directly by another method. It involves what is known as the *moment-generating function* (m.g.f) for a random variable which figuratively speaking packages all the moments of a random variable in a single power series (i.e. Maclaurin series) in the variable t. With the moment-generating function it is easier to obtain the moment of higher powers of probability distributions.

7.4.1 DEFINITION OF MOMENT-GENERATING FUNCTION

Definition 7.14 MOMENT-GENERATING FUNCTION

The moment-generating function for a random variable X, denoted as $M_X(t)$, is defined as

$$M_X(t) = E(e^{tX})$$

Definition 7.15 MOMENT-GENERATING FUNCTION
(Discrete Case)

Let X be a discrete random variable with probability distribution $p(x_i) = P(X = x_i),\ i = 1, 2, \cdots$. The function $M_X(t)$ is defined by

$$M_X(t) = \sum_{i=1}^{\infty} e^{tx_i} p(x_i)$$

Example 7.1
A discrete random variable X has a probability mass function

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Find the moment-generating function.

Solution

$$
\begin{aligned}
M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}
\end{aligned}
$$

Using the series expansion of

$$
e^y = \sum_{n=0}^{\infty} \frac{y^n}{n!}
$$

with $y = \lambda\, e^t$, we have

$$
\begin{aligned}
M_X(t) &= e^{-\lambda} e^{\lambda e^t} \\
&= e^{\lambda(e^t - 1)}
\end{aligned}
$$

Definition 7.16 MOMENT-GENERATING FUNCTION
(Continuous Case)

If X is a continuous random variable with probability density function $f(x)$, then the moment-generating function is defined as

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Example 7.2

A continuous random variable X has a probability density function

$$f(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

Find the moment-generating function.

Solution

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \int_0^{\infty} e^{tx} 2e^{-2x} dx \\ &= \int_0^{\infty} 2e^{-2x+tx} dx \\ &= 2\int_0^{\infty} e^{-(2-t)x} dx \\ &= 2\left[\frac{e^{-(2-t)x}}{2-t}\right]_0^{\infty} \\ &= \frac{2}{2-1}, \quad t < 2 \end{aligned}$$

Theorem 7.9

A moment-generating function for X exists if there exists a positive constant a such that $M_X(t)$ is finite for $|t| \leq a$

The moment-generating function of a random variable X exists only when the series is finite (in the discrete case) or the improper integral has a definite value (in the continuous case). This may not always be true because even if the moments are all finite and have definite values, the generating function may not converge for any value of t other than 0. It is for this reason that more advance book on probability theory tend to use what is called ***characteristics function*** (instead of the moment-generating function) which always exists for all random variables[22]. This is defined as in Definition 7.16 but with t replaced by it where $i = \sqrt{-1}$. This is also so in all the proofs of corresponding theorems. Advanced readers will recognize the moment-generating function as the Laplace transform of the function f, and the characteristics function as the Fourier transform.

In this text, whenever we make use of the moment-generating function, we shall always assume it exists.

7.4.2 APPLICATION OF MOMENT-GENERATING FUNCTIONS

The moment-generating function has two important applications. It is applied in

a) Finding any of the moments for X. If we can find $E(e^{tX})$, then we can find any of the moments for X.
b) Proving that a random variable possesses a particular probability distribution $p(x)$. If $M_X(t)$ exists for a probability distribution $p(x)$, it is unique (see Theorem 7.16).

Generating Moments from a Function

Let us recall the Maclaurin series expansion of the function e^x:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$$

It is known that this series converges for all values of x.

Thus

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots + \frac{(tx)^n}{n!} + \cdots$$

Now,

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E\left(1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots + \frac{(tX)^n}{n!} + \cdots\right) \end{aligned}$$

For a finite sum, the expected value of the sum equals the sum of the expected values (shown in my subsequent books). However, here we are dealing with an infinite series and hence cannot, immediately, apply such a result. It turns out, however, that under fairly general conditions, this operation is still valid. We shall assume that the required conditions are satisfied and proceed accordingly.

$$M_X(t) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \cdots + \frac{t^n E(X^n)}{n!} + \cdots \qquad (i)$$

because t is a constant so far as taking expectations of X is concerned.

We see that the coefficient of $\frac{t^k}{k!}$ in (i) is the k^{th} moment about the origin. For example, for

$$\begin{aligned}
k &= 1 \text{ the coefficient of } t \text{ is } E(X),\\
k &= 2 \text{ the coefficient of } \frac{t^2}{2!} \text{ is } E(X^2),\\
k &= 3 \text{ the coefficient of } \frac{t^3}{3!} \text{ is } E(X^3),\\
&\vdots\\
k &= n \text{ the coefficient of } \frac{t^n}{n!} \text{ is } E(X^n).
\end{aligned}$$

Theorem 7.10

If $M_X(t)$ exists, then for any positive integer k,

$$\left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0} = M_X^k(0) = \mu'_k$$

where $\frac{d^k M_X(t)}{dt^k} = M_X^k(t)$ is the k^{th} derivative of $M_X(t)$ with respect to t (since M_X is a function of the real variable t)

Proof

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \cdots + \frac{t^n E(X^n)}{n!} + \cdots \end{aligned}$$

Then

$$M'_X(t) = E(X) + tE(X^2) + \frac{t^2 E(X^3)}{2!} + \cdots + \frac{t^{n-1} E(X^n)}{(n-1)!} + \cdots$$

Setting $t = 0$,

$$M'_X(t) = E(X)$$

which is the expected value of the random variable X.
Thus, the first derivative of the moment-generating function evaluated at $t = 0$ yields the expected value of the random variable X.

If we compute the second derivative of $M_X(t)$, we obtain

$$M''_X(t) = E(X^2) + tE(X^3) + \cdots + \frac{t^{n-2} E(x^n)}{(n-2)!} + \cdots$$

Setting $t = 0$,

$$M''_X(0) = E(X^2)$$

Continuing in this manner, we obtain (assuming $M^n(0)$ exists),

$$M_X^{(n)}(0) = E(X^n)$$

Thus, from a knowledge of the function, $M_X(t)$, the moments may be "generated". Hence, the name "moment-generating function".

Example 7.3
For the density function of Example 7.2, find

a) the mean;
b) the variance.

Solution
From Example 7.2, the moment-generating function was found to be

$$\begin{aligned} M_X(t) &= \frac{2}{2-t} \\ &= \left(1-\frac{t}{2}\right)^{-1} \\ &= 1+\frac{t}{2}+\frac{t^2}{4}+\frac{t^3}{8}+\cdots \qquad (i) \end{aligned}$$

a) $M'_X(t) = \frac{1}{2}+\frac{2t}{4}+\frac{3t^2}{8}+\cdots,$
so that

$$E(X) = M'_X(0) = \frac{1}{2}$$

b) $M''_X(t) = \frac{2}{4}+\frac{6t}{8}+\cdots,$
so that

$$E(X^2) = M''_X(0) = \frac{2}{4} = \frac{1}{2}$$

Hence,

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

In fact, we could have obtained $E(X)$ and $E(X^2)$ from (i). Recall that the coefficient of $\frac{t^k}{k!}$ is the k^{th} moment about the origin. Now for

$k = 1$, the coefficient of t which is $E(X)$ is $\frac{1}{2}$;

$k = 2$, the coefficient of $\frac{t^2}{2!}$ which is $E(X^2)$ is $\frac{1}{2}$ since

$\frac{t^2}{4} = \frac{t^2}{2!}\left(\frac{1}{2}\right)$

7.4.3 PROPERTIES OF MOMENT-GENERATING FUNCTIONS

The moment-generating function has a number of useful properties. They are stated in theorems and we are encouraged to understand their import very well.

Property 1

Theorem 7.11

The moment-generating function always exists at $t = 0$ and equals 1:

$$M_X(0) = 1$$

Property 2

Theorem 7.12

Suppose that the random variable X has a moment-generating function $M_X(t)$. Let $Y = \alpha X$, (α is a constant), then $M_Y(t)$, the moments generation function of the random variable Y, is given by

$$M_Y(t) = M_X(\alpha t)$$

Proof

$$M_Y(t) = E(e^{tY}) = E[e^{(\alpha t)X}] = M_X(\alpha t)$$

That is, to find the moment-generating function of $Y = \alpha X$, evaluate the moment generating function of X at αt (instead of at t).

Property 3

Theorem 7.13

Suppose that the random variable X has moment-generating function $M_X(t)$. Let $Y = \alpha X + \beta$ (α, β are constants). Then $M_Y(t)$, the moment-generating function of the random variable Y, is given by

$$M_Y(t) = e^{\beta t} M_X(\alpha t)$$

Proof

$$M_Y(t) = E(e^{tY}) = E[e^{t(\alpha X+\beta)}] = E[e^{atX} e^{\beta t}] = e^{\beta t} E(e^{\alpha t X}) = e^{\beta t} M_X(\alpha t)$$

Example 7.3

The moment-generating function of a probability distribution is given by

$$M_X(t) = \frac{1}{1 - \frac{t}{\mu}}$$

Find the moment-generating function of Y, where $Y = 2x - 3$

Solution

$$\begin{aligned} M_Y(t) &= e^{-3t}M_X(2t) \qquad \text{(by Theorem 6.24)} \\ &= \frac{e^{-3t}}{1-\dfrac{2t}{\mu}} \qquad \text{(by Theorem 6.23)} \end{aligned}$$

Property 4

Theorem 7.14

Suppose that the random variable X has moment-generating function $M_X(t)$. Let $W = \alpha(X + \beta)$. Then $M_W(t)$, the moment-generating function of the random variable W, is given by

$$M_W(t) = e^{\alpha\beta t}M_X(\alpha t)$$

Proof
Replaying β by $\alpha\beta$ in Theorem 7.12, the result follows.

A special case of Theorem 7.13 is when $\beta = -\mu$ and $\alpha = \dfrac{1}{\sigma}$.

Property 5

Theorem 7.15

Suppose that the random variable X has moment-generating function $M_X(t)$. Let $\beta = -\mu$ and $\alpha = \dfrac{1}{\sigma}$, then

$$M_{\frac{x-\mu}{\sigma}}(t) = e^{-\frac{\mu t}{\sigma}}M_X\left(\frac{t}{\sigma}\right)$$

This is of particular importance when dealing with standardized random variables.

***Property 6* Uniqueness Property**

Theorem 7.16

Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$, respectively. If

$$M_X(t) = M_Y(t)$$

for all values of t, then

$$F_X(x) = F_Y(x), \text{ for all values of } x$$

or equivalently, X and Y have the same probability distribution

The proof of this theorem is too difficult to be given here.

Note

It is very important to understand exactly what the theorem says.

a) The theorem says that if two random variables have the same moment-generating function, then they have the same probability distribution. That is, the moment-generating function *uniquely* determines the probability distribution of the random variable. What this implies is that it is impossible for random variables with different probability distributions to have the same moment-generating functions.
b) What the theorem does not imply is that if two distributions have the same moments, then they are identical at all points. This is because in some cases, even though the moments exist, the moment-generating function does not, because the limit

$$\lim_{n\to\infty}\sum_{i=0}^{n}\frac{t^i m_i}{i!}$$

may not exist. An example of this situation is the lognormal distribution.

Property 7 **Sum[23] of Independent Random Variables[24]**

Theorem 7.17

Suppose that X and Y are independent random variables. Let $Z = X + Y$. Let $M_X(t)$, $M_Y(t)$ and $M_Z(t)$ be the moment-generating functions of the random variables X, Y, and Z, respectively. Then

$$M_Z(t) = M_X(t)M_Y(t)$$

Proof

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) \\
&= E(e^{t(X+Y)}) \\
&= E(e^{tX}e^{tY}) \\
&= E(e^{tX})E(e^{tY}) \text{ by independence of X and Y} \\
&= M_X(t)M_Y(t)
\end{aligned}$$

Note

This theorem may be generalized as follows.

Corollary 7.1
If $X_1, X_2, \cdots, X_n$ are independent random variables with moment-generating functions $M_{X_i}(t),\ i = 1, 2, \cdots, n$ then $M_Z(t)$, the moment-generating function of

$$Z = X_1 + X_2 + \cdots + X_n$$

is given by

$$M_Z(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_{n(t)}}$$

That is, the moment-generating function of the sum of independent random variables is equal to the product of their moment-generating functions.

From Chapter 5 to this chapter, we have been discussing in general terms the probability distributions of discrete and continuous random variables. In Volume II, we shall discuss special probability distributions. Under discrete probability distributions we shall consider the Bernoulli, Binomial, Geometric, Negative Binomial, Poisson, Hypergeometric and Multinomial distributions in the order. Under continuous distribution, the Uniform, Exponential, Gamma, Beta and Normal distributions will be considered, respectively.

EXERCISES

7.1 The random variable X has the probability density function

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

a) Find the moment-generating function of *X*.
b) From the moment-generating function, determine the mean and variance of *X*.

7.2 The time in hours between arrivals T of customers in a store, has probability density function

$$f_T(t) = \begin{cases} 10exp(-10t), & t \ge 0 \\ 0, & 0 < t \end{cases}$$

Determine the first four moments of T from its moment-generating function.

7.3 The random variable *X* has the probability density function

$$f(x) = \begin{cases} xe^{-x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

Find the moment-generating function of *X*.

7.4 The probability density function of a ransom variable X is given by

$$f(x) = \begin{cases} \frac{1}{4}, & 2 \leq x \leq 4 \\ \frac{1}{2}, & 4 < x \leq 5 \\ 0, & \text{elsewhere} \end{cases}$$

Find the mean and variance of X from its moment-generating function.

7.5 Given the following density function

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-a)}, & x \geq a \\ 0, & x < a \end{cases}$$

a) Find the moment-generating function of X.
b) Using the moment-generating function, find $E(X)$ and $Var(X)$.

7.6 If $Y = X - \mu$, find the moment-generating function of the random variable Y.

7.7 Refer to Exercise 5.14. Using the moment-generating function, find (a) $E(X)$ (b) $Var(X)$.

7.8 The p.d.f of X is given by

$$f(x) = \begin{cases} a + bx^2, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

If $E(X) = \dfrac{3}{5}$, find

a) a and b,
b) the variance of X.

7.9 If X and Y are independent random variables with respective probability mass functions

$$p(x) = \frac{e^{-\mu p}(\mu P)^x}{x!}, \quad x = 0, 1, 2 \cdots$$

and

$$p(y) = \frac{e^{-\mu p}(\mu p)^y}{y!}, \quad y = 0, 1, 2, \cdots$$

What is the moment-generating function of $X + Y$.

7.10 Refer to Exercise 5.27.

a) Find the moment-generating function of X.
b) Using the moment-generating function, find $E(X)$ and $Var(X)$.

7.11 If X and Y are independent random variables with respective probability mass functions

$$p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x}, \quad x 0, 1, 2, \cdots$$

and

$$p(y) = \left(\begin{array}{c} n \\ y \end{array} \right) p^y q^{n-y}, \quad y = 0, 1, 2 \cdots$$

What is the moment-generating function of $X + Y$

ANSWERS TO ODD-NUMBERED EXERCISES

Chapter 1

1.1 (a) $\{1,3,5,7,9\}$ (b) $\{x|x$ is a positive odd number less than10$\}$ 1.3 No 1.5 (c) 1.7 (a), (b), (c) 1.9 $\emptyset$ 1.11 (a) finite (c) infinite (e) finite (f) infinite (g) infinite
1.13 (a) $\{1,2,3,4,5,6,7\}$ (c) $\{1,2,3,4,5,6,7,8,10\}$ (e) $\{7,8,9,10\}$ (g) $\{1,2,3,7,8,9,10\}$ (i) $\{1,2,3,7\}$ (k) $\{4,5,6,7,8,9,10\}$ 1.21 $\{\{a,b,c\},\{a,b\},\{a,c\},\{b,c\},\{a\},\{b\},\{c\},\emptyset\}$
1.23 $\{\{1\},\{2\},\{3\}\}$; $\{\{1,2\},\{3\}\}$; $\{\{1,3\},\{2\}\}$; $\{\{1\},\{2,3\}\}$; $\{1,2,3\}$

Chapter 2

2.1 175,760,000 2.3 (a) 4 (c) 336 2.5 720 2.9 15120 2.15 5040 2.17 3360 2.19 (a) 120 (b) 72 2.21 (a) 35 (c) 6 (e) 28 (g) 230300 (i) 230300 2.23 (a) 1140 (b) 12 2.25 210 2.27 576 2.29 286
2.31

$$\begin{array}{c} 1 \\ 1 \quad 1 \\ 1 \quad 2 \quad 1 \\ 1 \quad 3 \quad 3 \quad 1 \\ 1 \quad 4 \quad 6 \quad 4 \quad 1 \\ 1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\ 1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \end{array}$$

2.33 (a) $x^4+4x^3y+6x^2y^2+4xy^3+y^4$ 2.35 (a) 35 (c) -1760 (e) 1120 2.37 $35x^4\ y^9$
2.39 1.9494

Chapter 3

3.1 $A=\{(2,6),(3,5),(5,3),(6,2)\}$ $D=\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$ $E=\{(3,3),(3,6),(6,3),(6,6)\}$ 3.3 (a) $A\cap(\overline{B\cup C})$ (c) $A\cup B\cup C$
(e) $[A\cap B\cap\bar{C}]\cup[A\cap\bar{B}\cap C]\cup[\bar{A}\cap B\cap C]$ (g) $A\cap B\cap C$ (i) $(A\cup B)\cap\bar{C}$
3.5 $E\cap F=\{(1,2),(1,4)(1,6)(2,1)(4,1),(6,1)\}$ $F\cap G=\{(1,4),(4,1)\}$
3.7 $\{(R,G),(R,B),(G,R),(G,B),(B,R),(B,G)\}$ 3.9 0.75 3.11 (a) 0.0044 (b) 0.03516 3.13 (a) (a) 4.933×10^{-7} (b) 0.0199 3.15 0.1512 3.17 (a) 0.1074 (b) 0.0619 3.19 (a) $\frac{7}{15}$
3.21 (a) $\frac{6}{10}$, $\frac{3}{10}$, $\frac{1}{10}$ 3.25 (a) $\frac{2}{5}$ (b) $\frac{9}{10}$ 3.27 $\frac{(n-k)!}{n!}$ 3.29 $\frac{1}{2}$

Chapter 4

4.3 (a) 0.8 (b) 0.7 (c) 0.44 (d) 0.56 4.5 $\frac{5}{8}$ 4.7 (a) 0.9 (b) 1 (c) 0.7 (d) 0.6 4.9 (a) 20,000 (b) 12,000 (c) 11,000 (d) 10,000 4.11 $\frac{\binom{n}{5}-\binom{n-3}{5}}{\binom{n}{5}}$ 4.13 (a) $\frac{1}{n(n-k)}$ (b) $\frac{(n-1)^{k-1}}{n^k}$ 4.15 $\frac{1}{2}$ 4.17 $\frac{5}{28}$ 4.19 (a) $\frac{14}{285}$ (b) $\frac{3}{95}$ 4.21 $\frac{2}{11}$ 4.23 (a) $\frac{1}{6}$ (b) $\frac{1}{6}$ 4.25 0.6044 4.27 0.413 4.29 0.1429 4.31 0.72 4.35 not independent

Chapter 5

5.1

X	0	1	2
$P(X = x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

5.3 (b) $f(x) = \begin{cases} \frac{1}{9}, & 0 \le x < 2 \\ \frac{2}{9}, & 2 \le x < 6 \\ \frac{2}{3}, & 6 \le x < 8 \end{cases}$ 5.5 (a) 4 (b) $f(x) = \begin{cases} 0, & \text{elsewhere} \\ \frac{x-1}{4}, & x = 2, 3, 5 \\ 1, & x > 5 \end{cases}$ 5.7 (a) $a = \frac{3}{4}$ (b) $\frac{1}{4}$ 5.9 $\frac{1}{\pi}$ 5.11 (a) 1 (b) $1 - e^{-x}$ 5.13 (a) $F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^2, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}$ (c) 1 (d) 1

5.15 -0.02 5.17 (a) $f(x) = \begin{cases} \alpha e^{-\alpha x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ (b) $1 - e^{-2\alpha}$ (c) $e^{-3\alpha}$ (d) $1 - e^{-3\alpha}$

5.21 (a) $F(x) = \begin{cases} 0, & x < 0 \\ x^3, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$ (b) 0.94 (c) 0.88

5.23 (a) $f(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{2}, & 0 \leq x \leq 1 \\ x - \frac{1}{2}, & 1 < x < 1.5 \\ 1, & x \geq 1.5 \end{cases}$ (b) 0.125 (c) 0.575 (d) $\frac{4}{9}$

5.25 (a) $f(x) = \begin{cases} \frac{1}{8}, & 0 < x < 2 \\ \frac{x}{8}, & 2 \leq x \leq 4 \\ 0, & \text{elsewhere} \end{cases}$ (b) $\frac{7}{16}$ (c) $\frac{13}{16}$ (d) $\frac{9}{16}$

5.27 (a) $f(x) = \begin{cases} 0, & x < 0 \\ 4(\frac{1}{3}x^3 - \frac{1}{4}x^4), & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$

(c) $F(-1) = 0,\ F(0) = 0,\ F(1) = \frac{1}{3}$ 5.29 (a) 0.0453 (c) 0.5390 5.31 (b) 0.2535

Chapter 6

6.1 (a) 1 (c) $\frac{8}{7}$, $\frac{8}{5}$ 6.5 (a) 1.05 (c) 0.525 (e) 4.5475 (g) 1.1369 6.9 (a) 7 (b) 5.83 6.11 $\frac{k!}{\lambda^k}\ \lambda > 0$ 6.13 $E(X) = \frac{2}{3}$ (b) $Var(X) = \frac{2}{63}$

Chapter 7

7.1 (a) $E(X) = \frac{2}{3},\ Var(X) = \frac{1}{18}$ 7.3 (a) $\frac{1}{(1-t)^2},\ |t| < 1$ 7.5 (a) $\frac{\lambda}{\lambda - t}e^{a\lambda},\ |t| < \lambda$ (b) $\frac{1}{\lambda}e^{a\lambda}$ 7.7 $E(X) = \frac{2}{3}$ (b) $Var(X) = \frac{2}{63}$ 7.9 $e^{2\mu p}(1 - e^t)$ 7.11 $(pe^t + q)^{2n}$

BIBLIOGRAPHY

Applebaum, D., (1996), *Probability and Information: An Integrated Approach* , Cambridge University Press, Cambridge.

Bajpal, A.C., Calus, L.M. and Fairley, J.A., (1978), *Statistical Methods for Engineers and Scientists,* John Wiley and Sons, New York.

Barlow, R. (1989), *Statistics, A Guide to the Use of Statistical Methods in the Physical Sciences,* John Wiley and Sons, New York.

Bartoszyński, R and Niewiadomska-Bugay, M (1996), *Probability and Statistical Inference,* John Wiley and Sons, New York.

Beaumont, G.P. (1986), *Probability and Random Variables,* Ellis Norwood Ltd, Chichester.

Berger, M.A. (1993) *An Introduction to Probability and Stochastic Processes,* Springer-Verlag, New York.

Billinsley, P. (1979) *Probability and Measure*, John Wiley and Sons, New York.

Birnbaum, Z.W. (1962), *Introduction to Probability and Mathematical Statistics,* Harper and Brothers Publishers, New York.

Blake, Ian F.B. (1979), *An Introduction to Applied Probability,* John Wiley and Sons, New York.

Bradley, J., (1988), *Introduction to Discrete Mathematics, Addison-Wesley Publishing Company,* Reading.

Breiman L., (1972), *Society for Industrial and Applied Maths,* Siam, Philadelphia.

Brémaud, Pierre (1994), *An Introduction to Probabilistic Modeling,* Springer – Verlag, New York.

Chow, Y.S., Teicher, H. (1988), *Probability Theory*, Springer-Verlag, New York.

David F.N. (1951), *Probability Theory for Statistical Methods*, Cambridge University Press, Cambridge.

Drake, A.W. (1967), *Fundamentals of Applied Probability Theory,* McGraw-Hill, New York.

Dudewicz, E.J. and Mishra, S.N. (1988), *Modern Mathematical Statistics,* John Wiley and Sons, New York.

Feller, W. (1970) *An Introduction to Probability Theory and its Applications* Vol. I, Second Edition, John Wiley and Sons, New York.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications* Vol. II, Second Edition, John Wiley and Sons, New York.

Freund, J.E. and Walpole, R.E. (1971), *Mathematical Statistics,* Prentice – Hall, Inc, Englewood Cliffs, New Jersey.

Fristedt, B. Gray, L. (1997), *Modern Approach to Probability Theory*, Birhhäuser, Boston.

Galambos, J. (1984), *Introductory Probability Theory,* Marcel Dekker, Inc., New York and Basel.

Gnedenko, B.V. and Khinchin, A.Ya. (1962), *An Elementary Introduction to the Theory of Probability,* Dover Publications, Inc., New York.

Goldberg, S. (1960), *Probability, An Introduction*, Prentice-Hall, Inc., Englewood Cliffe, New Jersey.

Grimmeth, G. and Weslsh, D. (1986), *Probability, An Introduction*, Clarendon Press, Oxford.

Guttman, I., Wilks, S.S and Hunter, J.S. (1982), *Introduction Engineering Statistics,* John Wiley and Sons, New York.

Haln, G.J., and Shapiro, S.S. (1967) *Statistical Models in Engineering*, John WIley and Sons.

Hoel, P.G. (1984), *Mathematical Statistics*, Fifth ed., John Wiley and Sons, New York.

Hoel, P.G., Port, S.C, and Stone, J.S. (1971), *Introduction to Probability Theory*, Houghton Mifflin.

Hogg, R.V. and Cragg, A.T. (1978), *Introduction to Mathematical Statistics*, Fourth Edition, Macmillan, New York.

Johnson, N.L. and Kotz, S.(1969) *Distributions in Statistics: Discrete Distributions* , John Wiley and Sons, New York.

Johnson, N.L. and Kotz, S.(1970) *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons, New York.

Johnson, N.L. and Kotz, S.(1970) *Distributions in Statistics: Continuous Univariate Distributions 1 & 2*, John Wiley and Sons, New York.

Kendal, M.G. and Stuart, A. (1963) *The Advanced Theory of Statistics,*, Vol. 1, 2nd ed., Charles Griffin & Company Ltd., London.

Kendal, M.G. and Stuart, A. (1976) *The Advanced Theory of Statistics,*, Vol. 3, 4th ed., Charles Griffin & Company Ltd., London.

Kendal, M.G. and Stuart, A. (1979) *The Advanced Theory of Statistics,*, Vol. 2, 4th ed., Charles Griffin & Company Ltd., London.

Kolmogorov, A.N. (1956), *Foundations of the Theory of Probability,* Chelsea Publishing Company, New York.

Lai, Chung Kai (1975), *Elementary Probability: Theory with Stochastic Processes,* Springer – Verlag, New York.

Lamperti, J. (1966), *Probability: A Survey of the Mathematical Theory*, W.A. Benjamin, INC., New York.

Lindgren, B.W. (1968) *Statistical Theory,* McMillan Company, New York.

Lupton R., (1993), *Statistics in Theory and Practice*, Princeton, Princeton University Press.

Mendenhall, W. and Scheaffer, R.L. (1990), *Mathematical Statistics With Applications,* PWS-KENT Publishing Company, Mass.

Meyer, D. (1970), *Probability*, W. A. Benjamin, Inc. New York.

Meyer, P.L. (1970), Introductory Probability and Its Applications, 2nd Edition, Addison Wesley Publishing Company.

Miller, I. and Freund, J.E. (1987), *Probability and Statistics for Engineers,* Third Edition, Prentice-Hall, Englewood Cliffs, New Delhi.

Mood, A.M., Graybill, F.A. and Boes, D.C (1974), *Introduction to the Theorey of Statistics*, Third Edition, McGraw-Hill, New York.

Page, L.B. (1989), *Probability for Engineering with Applications to Reliability,* Computer Science Press.

Prohorov, Yu.V. and Rozanov, Yu.A. (1969) *Probability Theory,* Springer-Verlag, New York.

Robinson, E.A. (1985), *Probability Theory and Applications* International Human Resources Dev. Corporation.

Ross, S. (1984), *A First Course in Probability*, 2nd Edition, Macmillan Publishing Company.

Spiegel, M.R. (1980), *Probability and Statistics,* McGraw-Hill Book Company, New York.

Stirzaker, D.(1994), *Elementaty Probability*, Cambridge University Press, Cambridge.

Stoyanov. J, et. al. (1989), *Exercise Manual in Probability Theory,* Kluwer Academic Publishers, London.

Studies in the History of Statistics and Probability, Vol. 1, edited by Pearson, E.S., and Kendal, M. (1970), Charles Griffin & Company Ltd., London.

Studies in the History of Statistics and Probability, Vol. 1, edited by Pearson, E.S., and Kendal, M. (1977), Charles Griffin & Company Ltd., London.

Trivedi, K.S. (1988), *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, Prentice-Hall of India, New Delhi.

Tucker, H.G. (1963), *An Introduction to Probability and Mathematical Statistics*, Academic Press, New York.

Wayne, W.D. (1991), *Biostatistics: A Foundation for Analysis in the Health Sciences,* Fifth Edition, John Wiley and Sons, Singapore.

Wilks, S.S. (1962), *Mathematical Statistics,* John Wiley and Sons, New York.

ENDNOTES

1. A terminating decimal is a decimal that ends. It is a decimal with a finite number of digits. For example, if we divide 1 by 8, we get 0.125.
2. A repeating (or recurring) decimal is a number whose decimal part eventually has the same sequence of digits repeating indefinitely. For example, ⅓=0.333.
3. The name power set is motivated by the fact that if set $\mathcal{A}$ is finite and has n distinct elements, then its power set $\mathcal{P}(\mathcal{A})$ contains exactly 2^n elements.
4. See Examples 2.15.
5. Objects may be persons, town, numbers, or anything of interest.
6. With the factorial function on any simple scientific calculator, this approximation is, with time, becoming more and more obsolete. However, it is highly used in advanced probability theory.
7. In an experiment a coin is *tossed* or *flipped* while a die is *rolled* or *thrown*.
8. The experiment of throwing two dice once is equivalent to the experiment of throwing a die twice. Both experiments leads to the same sample space.
9. In general, any events defined on non-overlapping sets of trials are independent.
10. It is believed that the first mathematician to calculate a theoretical probability correctly was Girolama Cardano, an Italian who lived from 1501 to 1576.
11. A finite probability space is obtained by assigning to each point $e_i \in S = \{e_1, e_2, \cdots, e_n\}$ a real number p_i, called the probability of e_i, which satisfies the following properties (a) $p_i \geq 0$ (b) $\sum_i^n p_i = 1$.
12. A probability function is a real-valued function defined over the events in a sample space with three fundamental properties given as axioms in Definition 3.31. A synonym of a probability function is a probability measure.
13. Axioms are basic assumption that one takes to develop ideas.
14. We can interchange $\cup$ with +
15. To determine the formula for the multiplication rule of probability for several events, we have to condition each event on the occurrence of all of the preceding events
16. In the table, recall that
    $P(\mathcal{H}_i \cap \mathcal{F}) = P(\mathcal{H}_i)P(\mathcal{F}|\mathcal{H}_i)$
    Note also that
    $P(T) = \sum_i P(\mathcal{F} \cap \mathcal{H}_i) = P(\mathcal{H}_1)P(\mathcal{A}|\mathcal{H}_1) + P(\mathcal{H}_2)P(\mathcal{A}|\mathcal{H}_2)$
17. Recollect from Definition 4.2 that $P(\mathcal{A}|\mathcal{B})$ is defined only when $P(B) > 0$ or $P(\mathcal{B}|\mathcal{A})$ is defined only when $P(A) > 0$ But in some cases we would like to determine the independence of events when their probabilities are zero.
18. The probabilities which appear in the margins of the table are called marginal probabilities.
19. ¢ is the currency of Ghana called the cedi
20. A random variable *X* is said to be bounded if $|X| \leq M < \infty$. This implies that $P(|X| \leq M) = 1$
21. And also the theorem that the expected value of a finite sum of random variables is the sum of the expected values of the random variables. This will be proved later in Volume III.

22. The characteristics function always exists for all random variables because it is the integral of a bounded function on a space of finite measure.
23. Sum of random variables will be treated in another book.
24. Two random variables X and Y are said to be independent if $F(x, y) = F_X(x)F_Y(y)$ (to be shown in another book).

INDEX