# ELEMENTARY APPLICATIONS OF PROBABILITY THEORY

Second edition

Henry C. Tuckwell

Texts in Statistical Science

# Elementary Applications of Probability Theory

CHAPMAN & HALL STATISTICS TEXTBOOK SERIES

**Editors:**

**Dr Chris Chatfield**
Reader in Statistics
School of Mathematical Sciences
University of Bath, UK

**Professor Jim V. Zidek**
Department of Statistics
University of British Columbia, Canada

OTHER TITLES IN THE SERIES INCLUDE

**Practical Statistics for Medical Research**
D.G. Altman

**Interpreting Data**
A.J.B. Anderson

**Statistical Methods for SPC and TQM**
D. Bissell

**Statistics in Research and Development**
Second edition
R. Caulcutt

**The Analysis of Time Series**
Fourth edition
C. Chatfield

**Problem Solving – A Statistician's Guide**
C. Chatfield

**Statistics for Technology**
Third edition
C. Chatfield

**Introduction to Multivariate Analysis**
C. Chatfield and A.J. Collins

**Modelling Binary Data**
D. Collett

**Modelling Survival Data in Medical Research**
D. Collett

**Applied Statistics**
D.R. Cox and E.J. Snell

**Statistical Analysis of Reliability Data**
M.J. Crowder, A.C. Kimber, T.J. Sweeting and R.L. Smith

**An Introduction to Generalized Linear Models**
A.J. Dobson

**Multivariate Analysis of Variance and Repeated Measures**
D.J. Hand and C.C. Taylor

**The Theory of Linear Models**
B. Jorgensen

**Statistical Theory**
Fourth edition
B. Lindgren

**Essential Statistics**
Second edition
D.G. Rees

**Decision Analysis: A Bayesian Approach**
J.Q. Smith

**Applied Nonparametric Statistical Methods**
Second edition
P. Sprent

**Elementary Applications of Probability Theory**
H.C. Tuckwell

**Statistical Process Control: Theory and Practice**
Third edition
G.B. Wetherill and D.W. Brown

**Statistics in Engineering**
A practical approach
A.V. Metcalfe

*Full information on the complete range of Chapman & Hall statistics books is available from the publishers.*

# Elementary Applications of Probability Theory

## With an introduction to stochastic differential equations

---

## Second edition

Henry C. Tuckwell

Senior Research Fellow
Stochastic Analysis Group of the
Centre for Mathematics and its Applications
Australian National University
Australia

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

**Published by Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK**

Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Blackie Academic & Professional, Wester Cleddens Road, Bishopbriggs, Glasgow G64 2NZ, UK

Chapman & Hall GmbH, Pappelallee 3, 69469 Weinheim, Germany

Chapman & Hall USA, One Penn Plaza, 41st Floor, New York NY 10119, USA

Chapman & Hall Japan, ITP-Japan, Kyowa Building, 3F, 2-2-1 Hirakawacho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, Thomas Nelson Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition 1988
Second edition 1995



Typeset in 10/12 pt Times by Thomson Press (India) Ltd, New Delhi

ISBN 0 412 57620 1



A catalogue record for this book is available from the British Library

Library of Congress Catalog Card Number: 94-68995

To Silvia Dori

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# Contents

# Preface

For the second edition, two additional chapters, Chapters 11 and 12, have been written. The added material should make the book suitable for two consecutive courses in elementary and intermediate applications of probability.

The new material consists of an introduction to stochastic differential equations. It is hoped that this will be useful for applied mathematical modelling of the behaviour of many naturally occurring randomly fluctuating quantities. An attempt has been made to explain the material with a certain amount of rigour, but hopefully without so much detail that a practical understanding is impaired.

The stochastic differential equations in this book are first order equations with an additional noise term. This added term usually contains a Gaussian 'white noise' so that the resulting solution is called a *diffusion process.*

*Chapter 11* starts with a brief reminder of the nature of ordinary deterministic differential equations, followed by an explanation of the essential differences between deterministic and stochastic equations. These have been illustrated with data in neurophysiology and economics.

There follows a thorough discussion of the properties of the standard Wiener process which forms a cornerstone of the theory, and a section on white noise which is a useful concept, especially for modelling. The simplest stochastic differential equations, being those of the Wiener process with drift, are then introduced.

The analytical approach, by which is meant a study of such quantities as transition probabilities through the equations they satisfy, is introduced at the end of Chapter 11. This method of study is essential if one is to obtain accurate estimates for many properties of the random processes one is considering.

*Chapter 12* starts with a continuation of the analytic method by introducing the equations of Kolmogorov (or Fokker-Planck). Steady state and time-dependent solutions of these equations are then found for problems of classical interest - those involving the Wiener and Ornstein-Uhlenbeck processes.

In section 12.5, Ito's stochastic integral and the corresponding stochastic differential equations are introduced. This is followed by a heuristic derivation

of the formula for change of variable. A brief treatment is given of Stratonovich's integral which leads to simpler rules of calculus.

Sections 12.6 and 12.7 concern certain aspects of modelling with stochastic differential equations. The connection between the analytical method and the *direct* method involving stochastic differential equations is given, followed by examples drawn from theoretical population biology, mathematical economics and theoretical neurobiology.

Finally, a brief discussion is undertaken of various practical considerations–solution of stochastic differential equations by simulation, numerical methods of solution of equations for transition probabilities, and estimation of parameters.

Both Chapters 11 and 12 end with sets of exercises, some of those in Chapter 12 being suitable for project work.

It is a pleasure to thank the following colleagues for helpful suggestions with respect to the new material: Professors Joseph Gani and Christopher C. Heyde of the Institute of Advanced Studies in Canberra; and Professor James A. Koziol of Scripps Research Institute, La Jolla.

Henry C. Tuckwell
*Versailles, June 1994*

# Preface to the first edition

This book concerns applications of probability theory. It has been written in the hope that the techniques presented will be useful for problems in diverse areas. A majority of the examples come from the biological sciences but the concepts and techniques employed are not limited to that field. To illustrate, birth and death processes (Chapter 9) have applications to chemical reactions, and branching processes (Chapter 10) have applications in physics but neither of these specific applications is developed in the text.

The book is based on an undergraduate course taught to students who have had one introductory course in probability and statistics. Hence it does not contain a lengthy introduction to probability and random variables, for which there are many excellent books. Prerequisites also include an elementary knowledge of calculus, including first-order differential equations, and linear algebra.

The basic plan of the book is as follows.

*Chapter 1:* a review of basic probability theory;
*Chapters 2–5*: random variables and their applications;
*Chapter 6*: sequences of random variables and concepts of convergence;
*Chapters 7–10*: theory and properties of basic random processes.

The outline is now given in more detail.

*Chapter 1* contains a brief review of some of the basic material which will be needed in later chapters; for example, the basic probability laws, conditional probability, change of variables, etc. It is intended that Chapter 1 be used as a reference rather than a basis for instruction. Students might be advised to study this chapter as the material is called upon.

*Chapter 2* illustrates the interplay between geometry and probability. It begins with an historically interesting problem and then addresses the problem of finding the density of the distance between two randomly chosen points. The second such case, when the points occur within a circle, is not easy but the result is useful.

*Chapter 3* begins with the properties of the hypergeometric distribution. An important application is developed, namely the estimation of animal populations by the capture–recapture method. The Poisson distribution is

then reviewed and one-dimensional Poisson point processes introduced together with some of their basic properties. There follows a generalization to two dimensions, which enables one to study spatial distributions of plants and to develop methods to estimate their population numbers. The chapter concludes with the compound Poisson distribution which is illustrated by application to a neurophysiological model.

*Chapter 4* introduces several of the basic concepts of reliability theory. The relevant properties of the standard failure time distributions are given. The interesting spare parts problem is next and the concluding sections discuss methods for determining the reliability of complex systems.

*Chapter 5* commences by explaining the usefulness of computer simulation. There follows an outline of the theory of random number generation using the linear congruential method and the probability integral transformation. The polar method for normal random variables is given. Finally, tests for the distribution and independence properties of random numbers are described.

*Chapter 6* deals with sequences of random variables. Some methods for studying convergence in distribution and convergence in probability are developed. In particular, characteristic functions and Chebyshev's inequality are the main tools invoked. The principal applications are to proving a central limit theorem and a weak law of large numbers. Several uses for the latter are detailed.

*Chapter 7* starts with the definition of random (stochastic) processes and introduces the important Markov property. The rest of the chapter is mainly concerned with the elementary properties of simple random walks. Included are the unrestricted process and that in the presence of absorbing barriers. For the latter the probability of absorption and the expected time of absorption are determined using the difference equation approach. The concluding section briefly introduces the Wiener process, so fundamental in advanced probability. The concept of martingale and its usefulness are discussed in the exercises.

*Chapter 8* is on Markov chains. However, the theory is motivated by examples in population genetics, so the Hardy-Weinberg principle is discussed first. Elementary general Markov chain theory is developed for absorbing Markov chains and those with stationary distributions.

*Chapter 9* concerns birth and death processes, which are motivated by demographic considerations. The Poisson process is discussed as a birth process because of its fundamental role. There follow the properties of the Yule process, a simple death process and the simple birth and death process. The treatment of the latter only states rather than derives the equation satisfied by the probability generating function but this enables one to derive the satisfying result concerning the probability of extinction.

*Chapter 10* contains a brief introduction to the theory of branching processes, focusing on the standard Galton-Watson process. It is motivated

by the phenomenon of cell division. The mean and variance are derived and the probability of extinction determined.

It should be mentioned that references are sometimes not to the latest editions of books; for example, those of Hoel, Pielou, Strickberger and Watson.

In the author's view there is ample material for a one-quarter or one-semester course. In fact some material might have to be omitted in such a course. Alternatively, the material could be presented in two courses, with a division at Chapter 6, supplemented by further reading in specialist areas (e.g. ecology, genetics, reliability, psychology) and project work (e.g. simulation).

I thank the many Monash students who have taken the course in applied probability on which this book is based. In particular, Derryn Griffiths made many useful suggestions. It is also a pleasure to acknowledge the helpful criticisms of Dr James A. Koziol of Scripps Clinic and Research Foundation, La Jolla; and Drs Fima Klebaner and Geoffrey A. Watterson at Monash University. I am also grateful to Barbara Young for her excellent typing and to Jean Sheldon for her splendid artwork.

Henry C. Tuckwell
*Los Angeles, April 1987*

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

# 1

# A review of basic probability theory

This is a book about the applications of probability. It is hoped to convey that this subject is both a fascinating and important one. The examples are drawn mainly from the biological sciences but some originate in the engineering, physical, social and statistical sciences. Furthermore, the techniques are not limited to any one area.

The reader is assumed to be familiar with the elements of probability or to be studying it concomitantly. In this chapter we will briefly review some of this basic material. This will establish notation and provide a convenient reference place for some formulas and theorems which are needed later at various points.

## 1.1 PROBABILITY AND RANDOM VARIABLES

When an experiment is performed whose outcome is uncertain, the collection of possible **elementary outcomes** is called a **sample space**, often denoted by $\Omega$. Points in $\Omega$, denoted in the discrete case by $\omega_i$, $i = 1, 2, \ldots$ have an associated probability $P\{\omega_i\}$. This enables the probability of any subset $A$ of $\Omega$, called an **event**, to be ascertained by finding the total probability associated with all the points in the given subset:

$$P\{A\} = \sum_{\omega_i \in A} P\{\omega_i\}$$

We always have

$$0 \leqslant P\{A\} \leqslant 1,$$

and in particular $P\{\Omega\} = 1$ and $P\{\varnothing\} = 0$, where $\varnothing$ is the empty set relative to $\Omega$.

A **random variable** is a real-valued function defined on the elements of a sample space. Roughly speaking it is an observable which takes on numerical values with certain probabilities.

**Discrete random variables** take on finitely many or countably infinitely many values. Their probability laws are often called **probability mass functions**. The following discrete random variables are frequently encountered.

**Binomial**

A **binomial** random variable $X$ with parameters $n$ and $p$ has the probability law

$$\boxed{p_k = \Pr\{X = k\} = \binom{n}{k} p^k q^{n-k}} \tag{1.1}$$

$$\doteq b(k; n, p), \qquad k = 0, 1, 2, \ldots, n,$$

where $0 \leqslant p \leqslant 1$, $q = 1 - p$ and $n$ is a positive integer ($\doteq$ means we are defining a new symbol). The **binomial coefficients** are

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

being the number of ways of choosing $k$ items, without regard for order, from $n$ distinguishable items.

When $n = 1$, so we have

$$\Pr\{X = 1\} = p = 1 - \Pr\{X = 0\},$$

the random variable is called **Bernoulli**.

Note the following.

*Convention*

**Random variables are always designated by capital letters (e.g. $X$, $Y$) whereas symbols for the values they take on, as in $\Pr\{X = k\}$, are always designated by lowercase letters.**

The converse, however, is not true. Sometimes we use capital letters for non-random quantities.

**Poisson**

A **Poisson** random variable with parameter $\lambda > 0$ takes on non-negative integer values and has the probability law

$$\boxed{p_k = \Pr\{X = k\} = \frac{e^{-\lambda}\lambda^k}{k!}}, \qquad k = 0, 1, 2, \ldots. \tag{1.2}$$

For any random variable the total probability mass is unity. Hence if $p_k$ is given by either (1.1) or (1.2),

$$\boxed{\sum_k p_k = 1}$$

where summation is over the possible values $k$ as indicated.

For any random variable $X$, the **distribution function** is

$$F(x) = \Pr\{X \leqslant x\}, \quad -\infty < x < \infty.$$

**Continuous** random variables take on a continuum of values. Usually the probability law of a continuous random variable can be expressed through its **probability density function**, $f(x)$, which is the derivative of the distribution function. Thus

$$\begin{aligned} f(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F(x) \\ &= \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr\{X \leqslant x + \Delta x\} - \Pr\{X \leqslant x\}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr\{x < X \leqslant x + \Delta x\}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr\{X \in (x, x + \Delta x]\}}{\Delta x} \end{aligned} \tag{1.3}$$

The last two expressions in (1.3) often provide a convenient prescription for calculating probability density functions. Often the latter is abbreviated to p.d.f. but we will usually just say 'density'.

If the interval $(x_1, x_2)$ is in the range of $X$ then the probability that $X$ takes values in this interval is obtained by integrating the probability density over $(x_1, x_2)$.

$$\boxed{\Pr\{x_1 < X < x_2\} = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x.}$$

The following continuous random variables are frequently encountered.

**Normal (or Gaussian)**

A random variable with density

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}, \qquad -\infty < x < \infty, \tag{1.4}$$

$$\text{where } -\infty < \mu < \infty \quad \text{and} \quad 0 < \sigma^2 < \infty,$$

is called **normal**. The quantities $\mu$ and $\sigma^2$ are the mean and variance (elaborated upon below) and such a random variable is often designated

$N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$ the random variable is called a **standard normal** random variable, for which the usual symbol is $Z$.

### Uniform

A random variable with constant density

$$\boxed{f(x) = \frac{1}{b-a}}, \qquad -\infty < a \leqslant x \leqslant b < \infty,$$

is said to be **uniformly distributed** on $(a, b)$ and is denoted $U(a, b)$. If $a = 0, b = 1$ the density is unity on the unit interval,

$$f(x) = 1, \qquad 0 \leqslant x \leqslant 1$$

and the random variable is designated $U(0, 1)$.

### Gamma

A random variable is said to have a **gamma density** (or gamma distribution) with parameters $\lambda$ and $\rho$ if

$$\boxed{f(x) = \frac{\lambda(\lambda x)^{\rho-1} e^{-\lambda x}}{\Gamma(\rho)}}, \qquad x \geqslant 0; \qquad \lambda, \rho > 0.$$

The quantity $\Gamma(\rho)$ is the **gamma function** defined as

$$\Gamma(\rho) = \int_0^\infty x^{\rho-1} e^{-x} \, dx, \qquad \rho > 0.$$

When $\rho = 1$ the gamma density is that of an **exponentially distributed random variable**

$$\boxed{f(x) = \lambda e^{-\lambda x}}, \qquad x > 0.$$

For continuous random variables the density must integrate to unity:

$$\boxed{\int f(x) \, dx = 1}$$

where the interval of integration is the whole range of values of $X$.

## 1.2 MEAN AND VARIANCE

Let $X$ be a discrete random variable with

$$\Pr\{X = x_k\} = p_k, \qquad k = 1, 2, \ldots.$$

The **mean**, **average** or **expectation** of $X$ is

$$\boxed{E(X) = \sum_k p_k x_k.}$$

For a binomial random variable $E(X) = np$ whereas a Poisson random variable has mean $E(X) = \lambda$.

For a continuous random variable with density $f(x)$,

$$\boxed{E(X) = \int x f(x)\,\mathrm{d}x.}$$

If $X$ is normal with density given by (1.4) then $E(X) = \mu$; a uniform $(a, b)$ random variable has mean $E(X) = \frac{1}{2}(a + b)$; and a gamma variate has mean $E(X) = \rho/\lambda$.

The ***n*th moment** of $X$ is the expected value of $X^n$:

$$E(X^n) = \begin{cases} \sum_k p_k x_k^n & \text{if } X \text{ is discrete,} \\ \int x^n f(x)\,\mathrm{d}x & \text{if } X \text{ is continuous.} \end{cases}$$

If $n = 2$ we obtain the **second moment** $E(X^2)$. The **variance**, which measures the degree of dispersion of the probability mass of a random variable about its mean, is

$$\begin{aligned} \mathrm{Var}(X) &= E[(X - E(X))^2] \\ &= E(X^2) - E^2(X). \end{aligned}$$

The variances of the above-mentioned random variables are:

binomial, $npq$; Poisson, $\lambda$; normal, $\sigma^2$; uniform, $\frac{1}{12}(b - a)^2$; gamma, $\rho/\lambda^2$.

The square root of the variance is called the **standard deviation**.

## 1.3 CONDITIONAL PROBABILITY AND INDEPENDENCE

Let $A$ and $B$ be two random events. The **conditional probability** of $A$ given $B$ is, provided $\Pr\{B\} \neq 0$,

$$\boxed{\Pr\{A \mid B\} = \frac{\Pr\{AB\}}{\Pr\{B\}}}$$

where $AB$ is the **intersection** of $A$ and $B$, being the event that both $A$ and $B$ occur (sometimes written $A \cap B$). Thus only the occurrences of $A$ which are simultaneous with those of $B$ are taken into account. Similarly, if $X, Y$ are random variables defined on the same sample space, taking on values $x_i, i = 1, 2, \ldots, y_j, j = 1, 2, \ldots$, then the conditional probability that $X = x_i$ given $Y = y_j$ is, if $\Pr\{Y = y_j\} \neq 0$,

$$\Pr\{X = x_i \mid Y = y_j\} = \frac{\Pr\{X = x_i, Y = y_j\}}{\Pr\{Y = y_j\}},$$

the comma between $X = x_i$ and $Y = y_j$ meaning 'and'.

The **conditional expectation** of $X$ given $Y = y_j$ is

$$E(X \mid Y = y_j) = \sum_i x_i \Pr\{X = x_i \mid Y = y_j\}.$$

The **expected value** of $XY$ is

$$E(XY) = \sum_{i,j} x_i y_j \Pr\{X = x_i, Y = y_j\},$$

and the **covariance** of $X, Y$ is

$$\begin{aligned} \operatorname{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

The covariance is a measure of the linear dependence of $X$ on $Y$.

If $X, Y$ are independent then the value of $Y$ should have no effect on the probability that $X$ takes on any of its values. Thus we define $X, Y$ as **independent** if

$$\Pr\{X = x_i \mid Y = y_j\} = \Pr\{X = x_i\}, \qquad \text{all } i, j.$$

Equivalently $X, Y$ are independent if

$$\Pr\{X = x_i, Y = y_j\} = \Pr\{X = x_i\} \Pr\{Y = y_j\},$$

with a similar formula for arbitrary independent events.

Hence for independent random variables

$$E(XY) = E(X)E(Y),$$

so their covariance is zero. Note, however, that $\operatorname{Cov}(X, Y) = 0$ does not always imply $X, Y$ are independent. The covariance is often normalized by defining the **correlation coefficient**

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

where $\sigma_X, \sigma_Y$ are the standard deviations of $X$, $Y$. $\rho_{XY}$ is bounded above and below by

$$\boxed{-1 \leqslant \rho_{XY} \leqslant 1}$$

Let $X_1, X_2, \ldots, X_n$ be **mutually independent** random variables. That is,

$$\Pr\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\}$$
$$= \Pr\{X_1 \in A_1\} \Pr\{X_2 \in A_2\} \ldots \Pr\{X_n \in A_n\},$$

for all appropriate sets $A_1, \ldots, A_n$. Then

$$\boxed{\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i)}$$

so that variances add in the case of independent random variables. We also note the formula

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y),$$

which holds if $X, Y$ are independent. If $X_1, X_2, \ldots, X_n$ are **independent identically distributed** (abbreviated to **i.i.d.**) random variables with $E(X_1) = \mu$, $\operatorname{Var}(X_1) = \sigma^2$, then

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \mu n; \qquad \operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2.$$

If $X$ is a random variable and $\{X_1, X_2, \ldots, X_n\}$ are i.i.d. with the distribution of $X$, then the collection $\{X_k\}$ is called a **random sample of size *n*** for $X$. Random samples play a key role in computer simulation (Chapter 5) and of course are fundamental in statistics.

## 1.4 LAW OF TOTAL PROBABILITY

Let $\Omega$ be a sample space for a random experiment and let $\{A_i, i = 1, 2, \ldots\}$ be a collection of nonempty subsets of $\Omega$ such that

(i) $A_i A_j = \varnothing, \qquad i \neq j;$
(ii) $\bigcup_i A_i = \Omega.$

(Here $\varnothing$ is the null set, the impossible event, being the complement of $\Omega$.) Condition (i) says that the $A_i$ represent **mutually exclusive** events. Condition (ii) states that when an experiment is performed, at least one of the $A_i$ must be observed. Under these conditions the sets or events $\{A_i, i = 1, 2, \ldots\}$ are said to form a **partition** or **decomposition** of the sample space.

The **law or theorem of total probability** states that for any event (set) $B$,

$$\boxed{\Pr\{B\} = \sum_i \Pr\{B|A_i\} \Pr\{A_i\}}$$

A similar relation holds for expectations. By definition the expectation of $X$ conditioned on the event $A_i$ is

$$E(X|A_i) = \sum_k x_k \Pr\{X = x_k|A_i\},$$

where $\{x_k\}$ is the set of possible values of $X$. Thus

$$\begin{aligned} E(X) &= \sum_k x_k \Pr\{X = x_k\} \\ &= \sum_k x_k \sum_i \Pr\{X = x_k|A_i\} \Pr\{A_i\} \\ &= \sum_i \Pr\{A_i\} \sum_k x_k \Pr\{X = x_k|A_i\}. \end{aligned}$$

Thus

$$\boxed{E(X) = \sum_i E(X|A_i) \Pr\{A_i\}}$$

which we call the **law of total probability applied to expectations.**

We note also the fundamental relation for any two events $A$, $B$ in the same sample space:

$$\boxed{\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}}$$

where $A \cup B$ is the **union** of $A$ and $B$, consisting of those points which are in $A$ or in $B$ or in both $A$ and $B$.

## 1.5 CHANGE OF VARIABLES

Let $X$ be a continuous random variable with distribution function $F_X$ and density $f_X$. Let

$$y = g(x)$$

be a strictly increasing function of $x$ (see Fig. 1.1) with inverse function

$$x = h(y).$$

Then

$$Y = g(X)$$

is a random variable which we let have distribution function $F_Y$ and density $f_Y$.

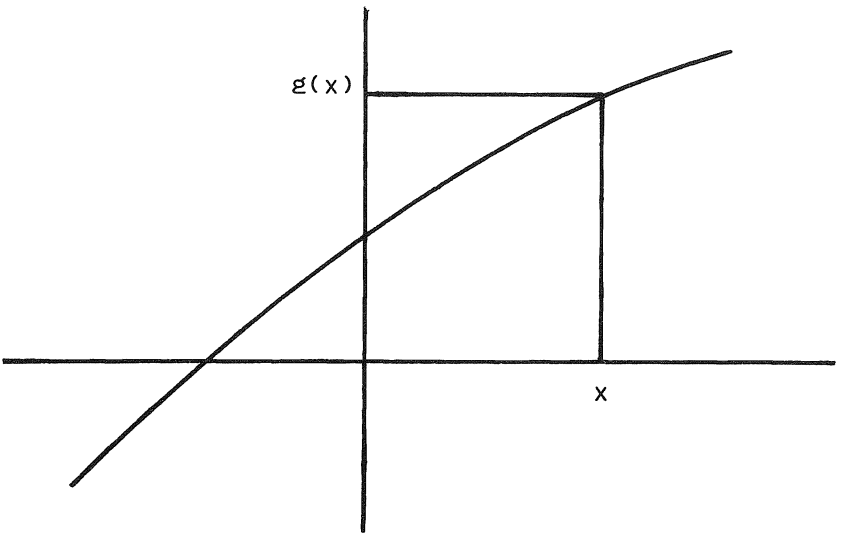


**Figure 1.1** $g(x)$ is a strictly increasing function of $x$.

It is easy to see that $X \leqslant x$ implies $Y \leqslant g(x)$. Hence we arrive at

$$\Pr\{X \leqslant x\} = \Pr\{Y \leqslant g(x)\}$$

By the definition of distribution functions this can be written

$$F_X(x) = F_Y(g(x)). \tag{1.5}$$

Therefore

$$F_Y(y) = F_X(h(y)).$$

On differentiating with respect to $y$ we obtain, assuming that $h$ is differentiable,

$$\frac{dF_Y}{dy} = \frac{dF_X(x)}{dx}\bigg|_{h(y)} \frac{dh}{dy}$$

or in terms of densities

$$f_Y(y) = f_X(h(y))\frac{dh}{dy}. \tag{1.6}$$

If $y$ is a strictly decreasing function of $x$ we obtain

$$\Pr\{X \leqslant x\} = \Pr\{Y \geqslant g(x)\}.$$

Working through the steps between (1.5) and (1.6) in this case gives

$$f_Y(y) = f_X(h(y))\left(-\frac{dh}{dy}\right). \tag{1.7}$$

Both formulas (1.6) and (1.7) are covered by the single formula

$$\boxed{f_Y(y) = f_X(h(y))\left|\frac{\mathrm{d}h}{\mathrm{d}y}\right|}$$

where $|\ \ |$ denotes **absolute value**. Cases where $g$ is neither strictly increasing nor strictly decreasing require special consideration.

## 1.6 TWO-DIMENSIONAL RANDOM VARIABLES

Let $X, Y$ be random variables defined on the same sample space. Then their **joint distribution function** is

$$F_{XY}(x, y) = \Pr\{X \leqslant x, Y \leqslant y\}.$$

The mixed partial derivative of $F_{XY}$, if it exists, is the **joint density** of $X$ and $Y$:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}}{\partial x \partial y}.$$

As a rough guide we have, for small enough $\Delta x, \Delta y$,

$$f_{XY}(x, y)\Delta x \Delta y \simeq \Pr\{X \in (x, x + \Delta x], Y \in (y, y + \Delta y]\}.$$

If $X, Y$ are independent then their joint distribution function and joint density function factor into those of the individual random variables:

$$F_{XY}(x, y) = F_X(x)F_Y(y),$$
$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

In particular, if $X, Y$ are **independent standard normal random variables**,

$$f_{XY}(x, y) = \left(\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-x^2}{2}\right\}\right)\left(\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-y^2}{2}\right\}\right), \qquad -\infty < x, y < \infty,$$

which can be written

$$\boxed{f_{XY}(x, y) = \frac{1}{2\pi}\exp\{-\tfrac{1}{2}(x^2 + y^2)\}} \tag{1.8}$$

In fact if the joint density $X, Y$ is as given by (1.8) we may conclude that $X, Y$ are independent standard normal random variables.

### Change of variables

Let $U, V$ be random variables with joint density $f_{UV}(u, v)$. Suppose that the one–one mappings

$$X = G_1(U, V)$$
$$Y = G_2(U, V)$$

transform $U, V$ to the pair of random variables $X, Y$. Let the inverse transformations be

$$U = H_1(X, Y)$$
$$V = H_2(X, Y).$$

Then the joint density of $X, Y$ is given by

$$\boxed{f_{XY}(x, y) = f_{UV}(H_1(x, y), H_2(x, y))|J(x, y)|}$$

where $J(x, y)$ is the **Jacobian** of the inverse transformation given by

$$J(x, y) = \begin{vmatrix} \dfrac{\partial H_1}{\partial x} & \dfrac{\partial H_1}{\partial y} \\ \dfrac{\partial H_2}{\partial x} & \dfrac{\partial H_2}{\partial y} \end{vmatrix}$$

and $|J|$ is its absolute value. A proof of this result is given in Blake (1979).

## 1.7 HYPOTHESIS TESTING - THE $\chi^2$ GOODNESS OF FIT TEST

In testing the validity of a **stochastic** (also called **random, probabilistic**) **model** it is often necessary to perform statistical tests on data. The basic idea is to consider a random variable which can be observed when the random experiment of interest is performed. Such a random variable is called a **test statistic**. If under a given hypothesis values of a test statistic occur (when the experiment is performed) which are considered unlikely, one is inclined to reject that hypothesis.

### $\chi^2$ random variables

Apart from a test for independence developed in Section 5.6, the only statistical test which is used in this book is called the $\chi^2$ **goodness of fit test**. We first define $\chi^2$ random variables and then see how these are useful in testing hypotheses about probability distributions.

**Definition** $X_n$ **is a** $\chi^2$**-random variable with** $n$ **degrees of freedom if its density is**

$$\boxed{f_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}}, \qquad x > 0; \quad n = 1, 2, \ldots \tag{1.9}$$

The mean and variance of such a random variable are given by

$$E(X_n) = n,$$
$$\mathrm{Var}(X_n) = 2n.$$

Also, it may be shown that the density (1.9) is that of a sum of squares of $n$ independent standard normal random variables $Z_i, i = 1, \ldots, n$:

$$X_n = \sum_{i=1}^{n} Z_i^2.$$

### The $\chi^2$ statistic

Suppose that when a random experiment is performed, observations may fall into any of $n$ distinct categories. Assuming the truth of a particular hypothesis, $H_0$, let the probability be $p_i$ that any observation falls in category $i$. If there are $N$ observations altogether, the expected number, under $H_0$, that fall in category $i$ is $Np_i$. We may compare this with the number, $N_i$, of observations that actually do fall in category $i$ ($N_i$ is random, $N$ is not). To obtain an overall view of how well the observed data fits the model ($H_0$) we compute the sum of the $n$ squares of the deviations of the $N_i$ from the $Np_i$, each term in the sum being divided by the **expected number** $Np_i$. Thus the goodness of fit test statistic is the random variable

$$D_n = \sum_{i=1}^{n} \frac{(N_i - Np_i)^2}{Np_i}.$$

When $N$ is large the random variable $D_n$ has approximately the same probability distribution as $X$, a $\chi^2$-random variable whose number of degrees of freedom is determined as described below. We therefore put

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_i - Np_i)^2}{Np_i}, \tag{1.10}$$

where $n_i$ is the observed value of $N_i$, and call (1.10) the value of the **$\chi^2$-statistic**.

If there is close agreement between the observed values ($n_i$) and those predicted under $H_0$ ($Np_i$), then the values of $(N_i - Np_i)^2$ and hence $D_n$ will be small. Large observed values of the $\chi^2$-statistic therefore make us inclined to think that $H_0$ is false.

Critical values of $X_n$, denoted by $\chi^2_{n.\alpha}$, are defined as follows:

$$\Pr\{X_n > \chi^2_{n.\alpha}\} = \alpha$$

If the value of $\chi^2$ obtained in an experiment is less than the critical value, it is argued that the differences between the values of $N_i$ and $Np_i$ are not large enough to warrant rejecting $H_0$. On the other hand, if $\chi^2$ exceeds the critical value, $H_0$ is considered unlikely and is rejected. Often we put $\alpha = .05$, which

means that 5% of the time, values of $\chi^2$ greater than the critical value occur even when $H_0$ is true. That is, there is a 5% chance that we will (incorrectly) reject $H_0$ when it is true.

In applying the above $\chi^2$ goodness of fit test, the number of degrees of freedom is given by the number $n$, of 'cells', minus the number of linear relations between the $N_i$. (There is at least one, $\sum N_i = N$.) The number of degrees of freedom is reduced further by one for each estimated parameter needed to describe the distribution under $H_0$.

It is recommended that the expected numbers of observations in each category should not be less than 5, but this requirement can often be relaxed. A table of critical values of $\chi^2$ is given in the Appendix, p. 219.

For a detailed account of hypothesis testing and introductory statistics generally, see for example Walpole and Myers (1985), Hogg and Craig (1978) and Mendenhall, Scheaffer and Wackerly (1981). For full accounts of basic probability theory see also Chung (1979) and Feller (1968). Two recent books on applications of probability at an undergraduate level are those of Ross (1985) and Taylor and Karlin (1984).

## 1.8 NOTATION

*Little o*

A quantity which depends on $\Delta x$ but vanishes more quickly than $\Delta x$ as $\Delta x \to 0$ is said to be 'little o of $\Delta x$', written $o(\Delta x)$. Thus for example $(\Delta x)^2$ is $o(\Delta x)$ because $(\Delta x)^2$ vanishes more quickly than $\Delta x$. In general, if

$$\lim_{\Delta x \to 0} \frac{g(\Delta x)}{\Delta x} = 0,$$

we write

$$g(\Delta x) = o(\Delta x).$$

The little o notation is very useful to abbreviate expressions in which terms will not contribute after a limiting operation is taken. To illustrate, consider the Taylor expansion of $e^{\Delta x}$:

$$e^{\Delta x} = 1 + \Delta x + \frac{(\Delta x)^2}{2!} + \frac{(\Delta x)^3}{3!} + \cdots$$

$$= 1 + \Delta x + o(\Delta x).$$

We then have

$$\frac{d}{dx} e^x \Big|_{x=0} = \lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{1 + \Delta x + o(\Delta x) - 1}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta x} + \frac{o(\Delta x)}{\Delta x}$$

$$= 1.$$

*Equal by definition*

As seen already, when we write, for example,

$$q \doteq (1 - p)$$

we are defining the symbol $q$ to be equal to $1 - p$. This is not to be confused with approximately equal to, which is indicated by $\simeq$.

*Unit step function*

The **unit** (or **Heaviside**) step function located at $x_0$ is

$$H(x - x_0) = \begin{cases} 0, & x < x_0, \\ 1, & x \geqslant x_0. \end{cases}$$

Thus $H(x - x_0)$ has a jump of $+1$ at $x_0$ and it is **right-continuous**.

*i.i.d.*

As seen already, the letters i.i.d. stand for independent and identically distributed.

*Probability*

Usually the probability of an event $A$ is written

$$\Pr\{A\}$$

but occasionally we just write

$$P\{A\}.$$

## REFERENCES

Blake, I.F. (1979). *An Introduction to Applied Probability.* Wiley, New York.
Chung, K.L. (1979). *Elementary Probability Theory.* Springer-Verlag, New York.
Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Wiley, New York.
Hogg, R.V. and Craig, A.T. (1978). *Introduction to Mathematical Statistics.* Macmillan, New York.
Mendenhall, W., Scheaffer, R.L. and Wackerly, D.D. (1981). *Mathematical Statistics with Applications.* Duxbury, Boston.