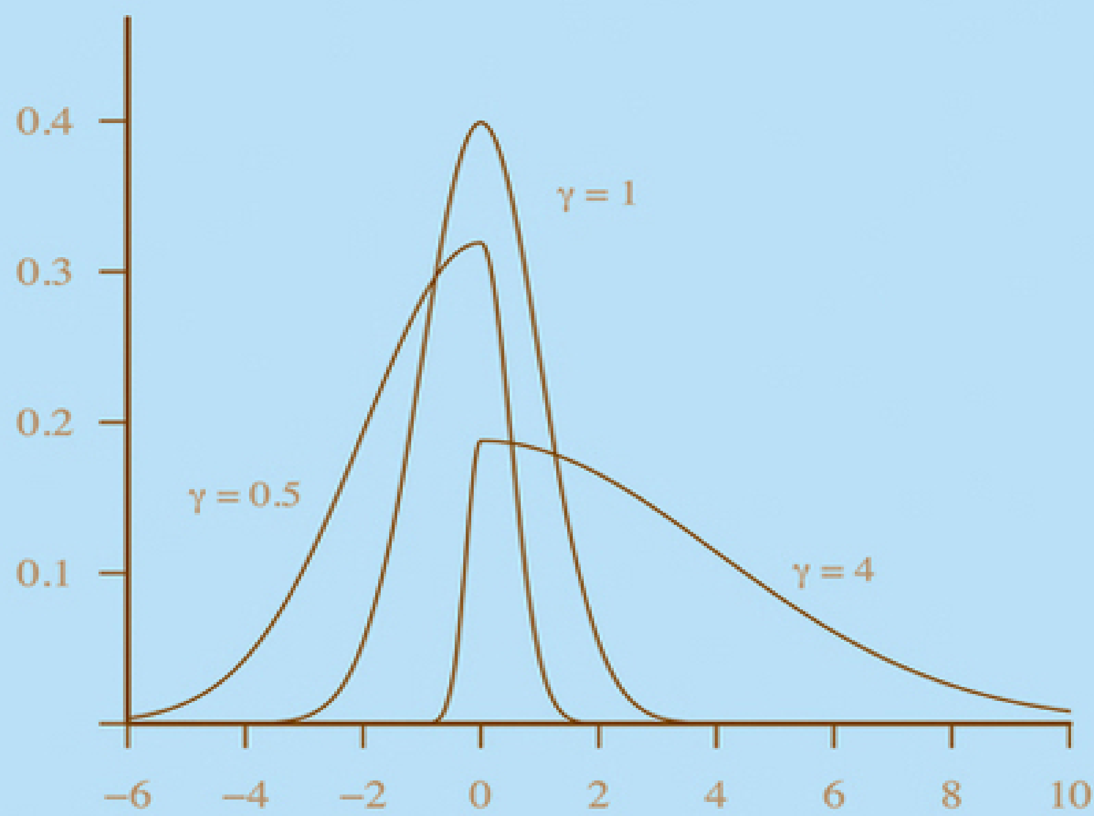
Statistical Distributions

FOURTH EDITION

Catherine Forbes, Merran Evans,
Nicholas Hastings, and Brian Peacock



WILEY

Contents

Statistical Distributions

Fourth Edition

Catherine Forbes
Monash University, Victoria, Australia

Merran Evans
Monash University, Victoria, Australia

Nicholas Hastings
Albany Interactive, Victoria, Australia

Brian Peacock
Brian Peacock Ergonomics, SIM University, Singapore

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

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

Published simultaneously in Canada.

Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002. Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Statistical distributions. – 4th ed. / Catherine Forbes... [et al.].

p. cm. – (Wiley series in probability and statistics)

Includes bibliographical references and index.

ISBN 978-0-470-39063-4 (pbk.)

1. Distribution (Probability theory) I. Forbes, Catherine.

QA273.6.E92 2010

519.2'4–dc22

2009052131

TO Jeremy and Elana Forbes
Caitlin and Eamon Evans
Tina Hastings
Eileen Peacock

Preface

This revised handbook provides a concise summary of the salient facts and formulas relating to 40 major probability distributions, together with associated diagrams that allow the shape and other general properties of each distribution to be readily appreciated.

In the introductory chapters the fundamental concepts of the subject are covered with clarity, and the rules governing the relationships between variates are described. Extensive use is made of the inverse distribution function and a definition establishes a variate as a generalized form of a random variable. A consistent and unambiguous system of nomenclature can thus be developed, with chapter summaries relating to individual distributions.

Students, teachers, and practitioners for whom statistics is either a primary or secondary discipline will find this book of great value, both for factual references and as a guide to the basic principles of the subject. It fulfills the need for rapid access to information that must otherwise be gleaned from many scattered sources.

The first version of this book, written by N. A. J. Hastings and J. B. Peacock, was published by Butterworths, London, 1975. The second edition, with a new author, M. A. Evans, was published by John Wiley & Sons in 1993, with a third edition by the same authors published by John Wiley & Sons in 2000. This fourth edition sees the addition of a new author, C. S. Forbes. Catherine Forbes holds a Ph.D. in Mathematical Statistics from The Ohio State University, USA, and is currently Senior Lecturer at Monash University, Victoria, Australia. Professor Merran Evans is currently Pro Vice-Chancellor, Planning and Quality at Monash University and obtained her Ph.D. in Econometrics from Monash University. Dr. Nicholas Hastings holds a Ph.D. in Operations Research from the University of Birmingham. Formerly

Mount Isa Mines Professor of Maintenance Engineering at Queensland University of Technology, Brisbane, Australia, he is currently Director and Consultant in physical asset management, Albany Interactive Pty Ltd. Dr. Brian Peacock has a background in ergonomics and industrial engineering which have provided a foundation for a long career in industry and academia, including 18 years in academia, 15 years with General Motors' vehicle design and manufacturing organizations, and 4 years as discipline coordinating scientist for the National Space Biomedical Institute/NASA. He is a licensed professional engineer, a licensed private pilot, a certified professional ergonomist, and a fellow of both the Ergonomics and Human Factors Society (UK) and the Human Factors and Ergonomics Society (USA). He recently retired as a professor in the Department of Safety Science at Embry Riddle Aeronautical University, where he taught classes in system safety and applied ergonomics.

The authors gratefully acknowledge the helpful suggestions and comments made by Harry Bartlett, Jim Conlan, Benoit Dulong, Alan Farley, Robert Kushler, Jerry W. Lewis, Allan T. Mense, Grant Reinman, and Dimitris Ververidis.

Catherine Forbes
Merran Evans
Nicholas Hastings
Brian Peacock

Chapter 1

Introduction

The number of puppies in a litter, the life of a light bulb, and the time to arrival of the next bus at a stop are all examples of random variables encountered in everyday life. Random variables have come to play an important role in nearly every field of study: in physics, chemistry, and engineering, and especially in the biological, social, and management sciences. Random variables are measured and analyzed in terms of their statistical and probabilistic properties, an underlying feature of which is the distribution function. Although the number of potential distribution models is very large, in practice a relatively small number have come to prominence, either because they have desirable mathematical characteristics or because they relate particularly well to some slice of reality or both.

This book gives a concise statement of leading facts relating to 40 distributions and includes diagrams so that shapes and other general properties may readily be appreciated. A consistent system of nomenclature is used throughout. We have found ourselves in need of just such a summary on frequent occasions—as students, as teachers, and as practitioners. This book has been prepared and revised in an attempt to fill the need for rapid access to information that must otherwise be gleaned from scattered and individually costly sources.

In choosing the material, we have been guided by a utilitarian outlook. For example, some distributions that are special cases of more general families are given extended treatment where this is felt to be justified by applications. A general discussion of families or systems of distributions was considered beyond the scope of this book. In choosing the appropriate symbols and parameters for the description of each distribution, and especially where different but interrelated sets of symbols are in use in different fields, we have tried to strike a balance between the various usages, the need for a consistent system of nomenclature within the book, and typographic simplicity. We have given some methods of parameter estimation where we felt it was appropriate to do so. References listed in the Bibliography are not the primary sources but should be regarded as the first “port of call”.

In addition to listing the properties of individual variates we have considered relationships between variates. This area is often obscure to the nonspecialist. We have also made use of the inverse distribution function, a function that is widely tabulated and used but rarely explicitly defined. We have particularly sought to avoid the confusion that can result from using a single symbol to mean here a function, there a quantile, and elsewhere a variate.

Building on the three previous editions, this fourth edition documents recent extensions to many of these probability distributions, facilitating their use in more varied applications. Details regarding the connection between joint, marginal, and conditional probabilities have been included, as well as new chapters (Chapters 5 and 6) covering the concepts of statistical modeling and parameter inference. In addition, a new chapter (Chapter 38) detailing many of the existing standard queuing theory results is included. We hope the new material will encourage readers to explore new ways to work with statistical distributions.

Chapter 2

Terms and Symbols

2.1 Probability, Random Variable, Variate, and Number

Probabilistic Experiment

A probabilistic experiment is some occurrence such as the tossing of coins, rolling dice, or observation of rainfall on a particular day where a complex natural background leads to a chance outcome.

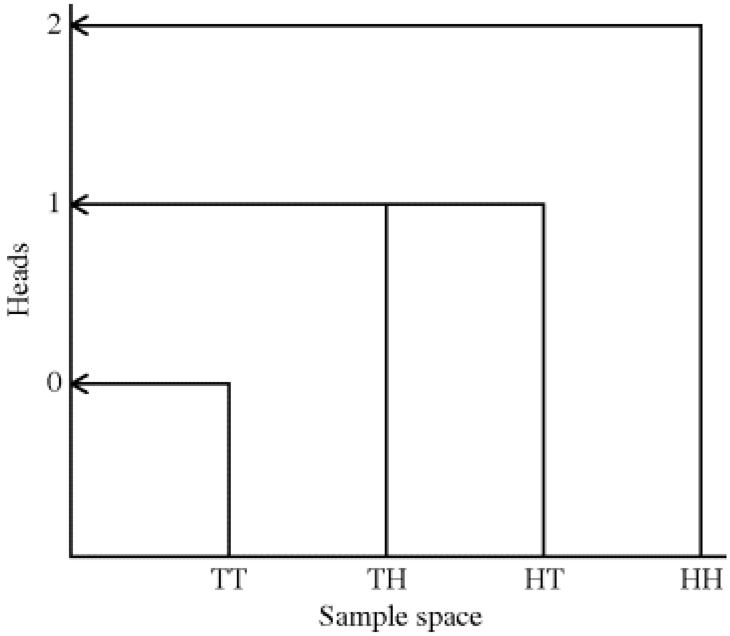
Sample Space

The set of possible outcomes of a probabilistic experiment is called the sample, event, or possibility space. For example, if two coins are tossed, the sample space is the set of possible results HH, HT, TH, and TT, where H indicates a head and T a tail.

Random Variable

A random variable is a function that maps events defined on a sample space into a set of values. Several different random variables may be defined in relation to a given experiment. Thus, in the case of tossing two coins the number of heads observed is one random variable, the number of tails is another, and the number of double heads is another. The random variable “number of heads” associates the number 0 with the event TT, the number 1 with the events TH and HT, and the number 2 with the event HH. Figure 2.1 illustrates this mapping.

Figure 2.1 The random variable “number of heads”.



Variate

In the discussion of statistical distributions it is convenient to work in terms of variates. A variate is a generalization of the idea of a random variable and has similar probabilistic properties but is defined without reference to a particular type of probabilistic experiment. A *variate* is the set of all random variables that obey a given probabilistic law. The number of heads and the number of tails observed in independent coin tossing experiments are elements of the same variate since the probabilistic factors governing the numerical part of their outcome are identical.

A *multivariate* is a vector or a set of elements, each of which is a variate. A *matrix variate* is a matrix or two-dimensional array of elements, each of which is a variate. In general, dependencies may exist between these elements.

Random Number

A *random number* associated with a given variate is a number generated at a realization of any random variable that is an element of that variate.

2.2 Range, Quantile, Probability Statement, and Domain

Range

Let ***X*** denote a variate and let $\Re_X$ be the set of all (real number) values that the variate can take. The set $\Re_X$ is the *range* of ***X***. As an illustration (illustrations are in terms of random variables) consider the experiment of tossing two coins and noting the number of heads. The range of this random variable is the set $0, 1, 2$ heads, since the result may show zero, one, or two heads. (An alternative common usage of the term *range* refers to the largest minus the smallest of a set of variate values.)

Quantile

For a general variate ***X*** let *x* (a real number) denote a general element of the range $\Re_X$. We refer to *x* as the *quantile* of ***X***. In the coin tossing experiment referred to previously, $x \in 0, 1, 2$ heads; that is, *x* is a member of the set $0, 1, 2$ heads.

Probability Statement

Let $X = x$ mean "the value realized by the variate ***X*** is *x*." Let $\Pr[X \leq x]$ mean "the probability that the value realized by the variate ***X*** is less than or equal to *x*."

Probability Domain

Let α (a real number between 0 and 1) denote probability. Let $\Re_X^\alpha$ be the set of all values (of probability) that $\Pr[X \leq x]$ can take. For a continuous variate, $\Re_X^\alpha$ is the line segment [0, 1]; for a discrete variate it will be a subset of that segment. Thus $\Re_X^\alpha$ is the *probability domain* of the variate ***X***.

In examples we shall use the symbol ***X*** to denote a random variable. Let ***X*** be the number of heads observed when two coins are tossed. We then have

$$\Pr[X \leq 0] = \frac{1}{4}$$

$$\Pr[X \leq 1] = \frac{3}{4}$$

$$\Pr[X \leq 2] = 1$$

and hence $\Re_X^\alpha = \frac{1}{4}, \frac{3}{4}, 1$.

2.3 Distribution Function and Survival Function

Distribution Function

The *distribution function* *F* (or more specifically F_X) associated with a variate ***X*** maps from the range $\Re_X$ into the probability domain $\Re_X^\alpha$ or [0, 1] and is such that

(2.1) $F(x) = \Pr[X \leq x] = \alpha \qquad x \in \Re_X, \alpha \in \Re_X^\alpha.$

The function *F*(*x*) is nondecreasing in *x* and attains the value unity at the maximum of *x*. Figure 2.2 illustrates the distribution function for the number of heads in the experiment of tossing two coins. Figure 2.3 illustrates a general continuous distribution function and Figure 2.4 a general discrete distribution function.

Figure 2.2 The distribution function *F*: $x \rightarrow \alpha$ or $\alpha = F(x)$ for the random variable, “number of heads”.

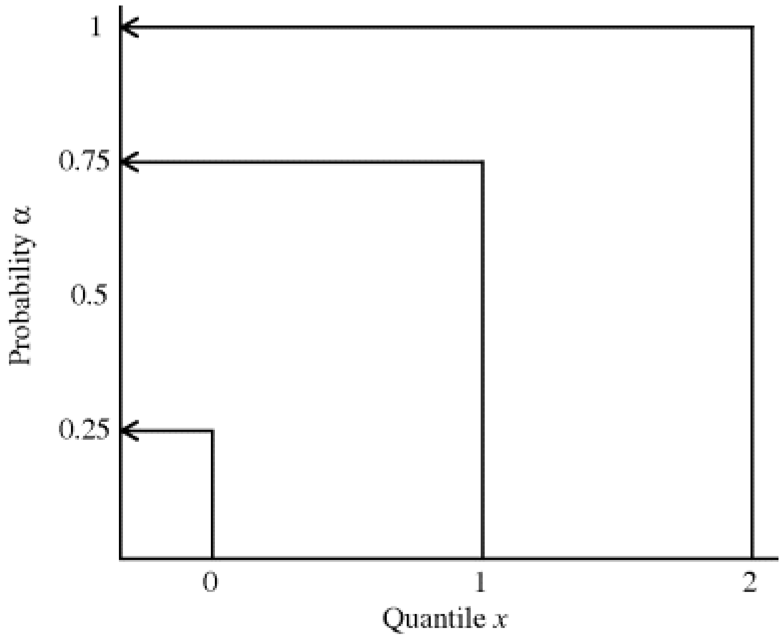


Figure 2.3 Distribution function and inverse distribution function for a continuous variate.

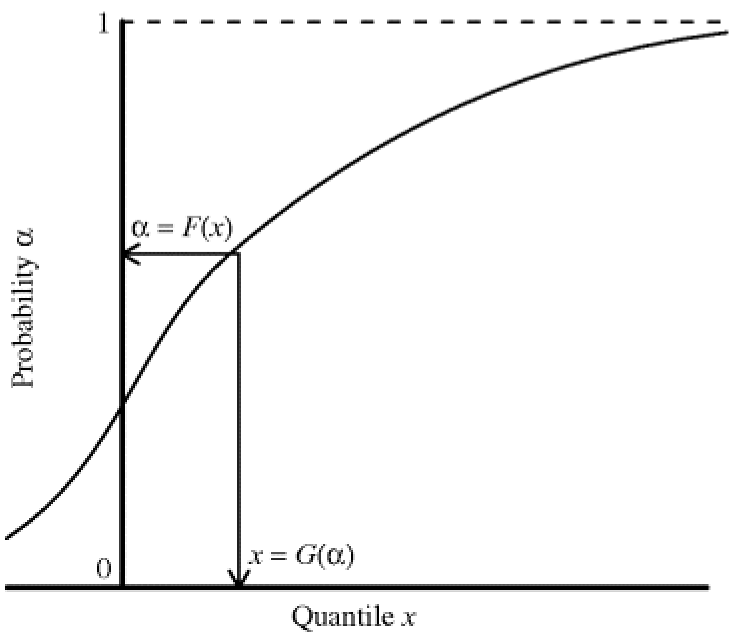
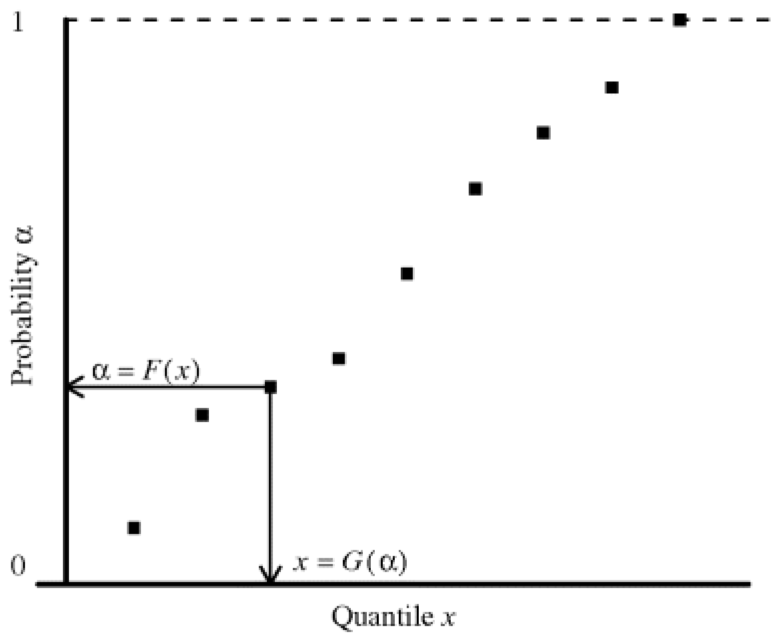


Figure 2.4 Distribution function and inverse distribution function for a discrete variate.



Survival Function

The *survival function* *S*(*x*) is such that

$$S(x) = \Pr[X > x] = 1 - F(x).$$

2.4 Inverse Distribution Function and Inverse Survival Function

For a distribution function *F*, mapping a quantile *x* into a probability α, the quantile function or inverse distribution function *G* performs the corresponding inverse mapping from α into *x*. Thus for $x \in \Re_X, \alpha \in \Re_X^\alpha$, the following statements hold:

(2.2) $\alpha = F(x)$

$$x = G(\alpha)$$

$$x = G(F(x))$$

(2.3) $\alpha = F(G(\alpha))$

$$\Pr[X \leq x] = F(x) = \alpha$$

(2.4) $\Pr[X \leq G(\alpha)] = F(x) = \alpha$