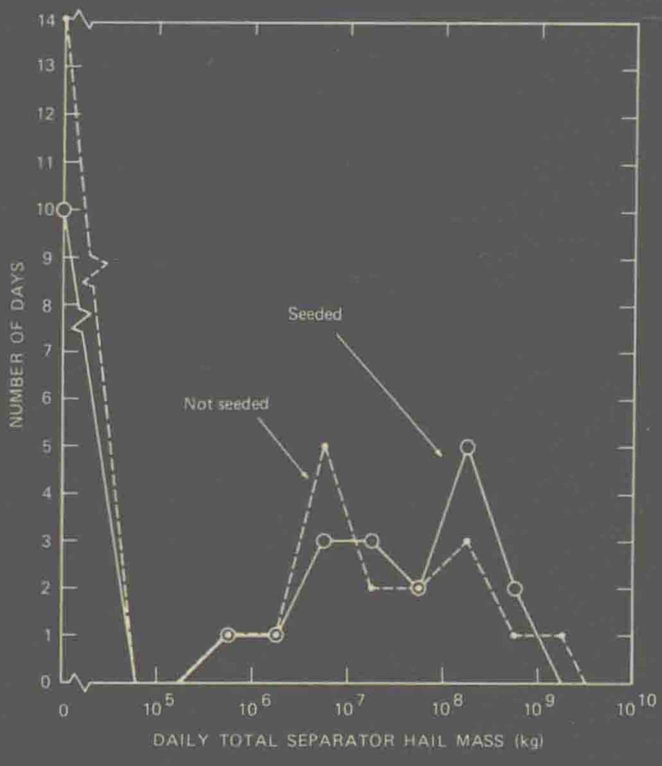
STATISTICS: textbooks and monographs volume 88

LOGNORMAL DISTRIBUTIONS

THEORY AND APPLICATIONS



edited by
EDWIN L. CROW
KUNIO SHIMIZU

LOGNORMAL DISTRIBUTIONS

THEORY AND APPLICATIONS

edited by

Edwin L. Crow
Institute for Telecommunication Sciences
National Telecommunications and
Information Administration
U.S. Department of Commerce
Boulder, Colorado

Kunio Shimizu
Department of Information Sciences
Faculty of Science and Technology
Science University of Tokyo
Noda City, Chiba, Japan

MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging-in-Publication Data

Lognormal distributions : theory and applications / edited by Edwin L. Crow, Kunio Shimizu.
p. cm. — (Statistics, textbooks and monographs ; vol. 88)
Includes indexes.
ISBN 0-8247-7803-0
1. Lognormal distribution. I. Crow, Edwin L. II. Shimizu, Kunio, [date]. III. Series: Statistics, textbooks and monographs ; v. 88.
QA273.6.L64 1988
519.2—dc19 87-24375
CIP

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

Preface

The lognormal distribution may be defined as the distribution of a random variable whose logarithm is normally distributed. It is a typical example of positively skewed frequency curves. It has two parameters but may be generalized by a translation parameter, by truncation and censoring, by adjoining a point probability mass, by extension to two or more dimensions, and by transformation. These generalizations, as well as the two-parameter distribution, are treated in this book.

Aitchison and Brown's 1957 monograph presented a unified treatment of the theory, beginning with a discussion of the genesis of the lognormal distribution from elementary random processes, passing through problems of point and interval estimation of parameters, and ending with a review of applications. Johnson and Kotz, in 1970, concisely summarized the history of theory and application of lognormal distributions and the development of estimation for the two- and three-parameter lognormal and related distributions.

Since the publication of the books by Aitchison and Brown and by Johnson and Kotz, the theory of lognormal distributions has steadily progressed and fields of application have greatly increased. (Johnson and Kotz had stated [1970, page 128], "It is quite likely that the lognormal distribution will be one of the most widely applied distributions in practical statistical work in the near future.") The present book emphasizes, but is not limited to, the more recent developments in the genesis, application, and properties of lognormal distributions, estimation and test theories, and some results for related distributions. The first seven chapters are primarily theory, the last seven primarily application.

This book is directed primarily to applied statisticians and graduate students in statistics, but will also be useful to researchers in a variety of subject-matter fields, including economics, biology, ecology, geology, atmospheric sciences, business, and industry. Subject-matter specialists have often not applied, or even had available, the recent more precise methods in the first seven chapters, while statisticians may find impetus for further development of methods in the discussions and references of the last seven chapters.

The book presupposes an introductory course in mathematical statistics and a knowledge of calculus. Thus the book requires some knowledge of basic results for the normal distribution, for many of the properties of the lognormal distribution may immediately be derived from those of the normal. As far as possible, the development is self-contained. A unified treatment is attempted despite the substantial number of authors, but the chapters can be read essentially independently. Each chapter has its own references; in addition, there is a complete author index with the location of full references italicized, as well as a complete subject index.

Chapter 1 gives a brief history (a more extensive history, including 217 references, is available in Aitchison and Brown [1957]), the genesis of the two-parameter distribution from qualitative assumptions (law of proportionate effect and breakage process), and basic properties, including moments, order statistics, and distributions of products and sums. Chapter 2 brings up to date the theory of point estimation of the two-parameter distribution, its multivariate analog, and the delta generalizations, in which a point probability mass is adjoined to the distribution. The formulas for estimates and their variances are often rendered compact and systematized by the use of generalized hypergeometric functions.

Chapter 3 presents the theory of interval estimation and testing of hypotheses for the two-parameter distribution. Chapters 4 and 5 deal with estimation for the three-parameter distribution and for data that have been censored, truncated, or grouped. The Bayesian approach to estimation has been particularly useful in life testing and is presented in Chapter 6.

Chapter 7 on the Poisson-lognormal distribution is unique in this book because that distribution is not strictly a lognormal distribution; it is a mixture of Poisson distributions, the Poisson means being lognormally distributed. It has been found useful in describing the abundance of species in biology.

The last seven chapters describe in detail the extensive applications in economics, business, and human affairs (8 and 9), industry (10), biology, especially growth models (11), ecology (12), atmospheric sciences (13), and geology (14).

The contributions of the ten other authors were invited and reviewed by us. We are grateful to them for their dedication and cooperation and to the chairman and editors of Marcel Dekker, Inc., Maurits Dekker, Vickie Kearn, John K. Cook, and Lila Harris, for their ready and invaluable professional assistance. The book's typesetters, The Bartlett Press, Inc., deserve a hand for dealing elegantly with often unwieldy equations.

While our employers have no responsibility for this work, we are grateful to the Institute for Telecommunication Sciences (ITS), U.S. Department of Commerce (DoC), and the Science University of Tokyo for the use of facilities and the assistance of individuals. In addition, both of us are indebted to the National Center for Atmospheric Research (NCAR) and its sponsor, the National Science Foundation, for positions as visitors in the Convective Storms Division, directed by Edward J. Zipser, during the preparation of the volume. Jean M. Bankhead (DoC), Jane L. Watterson (DoC), and Gayl H. Gray (NCAR) expertly and cheerfully provided bibliographical aid, and Kathy E. Mayeda (ITS), Carmen du Bouchet (ITS), and Sudie J. Kelly (NCAR) similarly provided word-processing aid. We are pleased to express our thanks to all.

Edwin L. Crow
Kunio Shimizu

Contributors

GREGORY CAMPBELL, Ph.D. Senior Staff Fellow, Laboratory of Statistical and Mathematical Methodology, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland

A. CLIFFORD COHEN, Ph.D. Professor Emeritus of Statistics, Department of Statistics, University of Georgia, Athens, Georgia

EDWIN L. CROW, Ph.D. Mathematical Statistician, Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, Colorado

BRIAN DENNIS, Ph.D. Associate Professor, College of Forestry, Wildlife, and Range Sciences, University of Idaho, Moscow, Idaho

CHARLES E. LAND, Ph.D. Health Statistician, Radiation Epidemiology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

RAYMOND J. LAWRENCE, M.A. Professor of Marketing, Department of Marketing, University of Lancaster, Bailrigg, Lancaster, England

JAMES E. MOSIMANN, Ph.D. Chief, Laboratory of Statistical and Mathematical Methodology, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland

WILLIAM J. PADGETT, Ph.D. Professor and Chairman, Department of Statistics, University of South Carolina, Columbia, South Carolina

G. P. PATIL, Ph.D., D.Sc. Professor of Mathematical Statistics, Director, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania; Visiting Professor of Biostatistics, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts

JEAN-MICHEL M. RENDU, Eng. Sc.D. Director, Technical and Scientific Systems Group, Newmont Mining Corporation, Danbury, Connecticut

S. A. SHABAN, Ph.D. Associate Professor, Faculty of Commerce, Economics and Political Science, Department of Insurance and Statistics, Kuwait University, Kuwait

KUNIO SHIMIZU, D.Sc. Assistant Professor, Department of Information Sciences, Faculty of Science and Technology, Science University of Tokyo, Noda City, Chiba, Japan

Contents

1

History, Genesis, and Properties

KUNIO SHIMIZU Department of Information Sciences, Faculty of Science and Technology, Science University of Tokyo, Noda City, Chiba, Japan

EDWIN L. CROW Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, Colorado

1. INTRODUCTION

The lognormal distribution (with two parameters) may be defined as the distribution of a random variable whose logarithm is normally distributed. Such a variable is necessarily positive. Since many variables in real life, from the sizes of organisms and the numbers of species in biology to rainfalls in meteorology and sizes of incomes in economics, are inherently positive, the lognormal distribution has been widely applied in an empirical way for fitting data. In addition, it has been derived theoretically from qualitative assumptions; Gibrat (1930, 1931) did this in 1930, calling it the law of proportionate effect, but Kapteyn (1903) had described a machine that was the mechanical equivalent. Kolmogoroff (1941) derived the distribution as the asymptotic result of an iterative process of successive breakage of a particle into two randomly sized particles.

Thus there is a theoretical basis as well as empirical application of lognormal distributions, but why is there much to say about them if the data analysis can be referred to the intensively studied normal distribution by taking the logarithm? There are several reasons:

(1) The parameter estimates resulting from the inverse transformation are biased.
(2) The two-parameter distribution is often not a sufficient description; a third parameter, the threshold or location parameter, is needed, for example, for the distribution of ages at first marriage.
(3) The distribution may be censored or truncated (e.g., low income data are often missing), or the data may be classified into groups, so that special methods are needed. (See Chapter 5.)

Aitchison and Brown (1957) and Johnson and Kotz (1970) have described the early history of lognormal distributions, but a brief summary is desirable here. Galton (1879) and McAlister (1879) initiated the study of the distribution in papers published together, relating it to the use of the geometric mean as an estimate of location. Much later Kapteyn (1903) discussed the genesis of the distribution, and Kapteyn and Van Uven (1916) gave a graphical method for estimating the parameters. Wicksell (1917) used the method of moments for three-parameter estimation, introducing a third parameter, the threshold, to fit the distribution of ages at first marriage. Nydell (1919) obtained asymptotic standard errors for the moment estimates. The distribution appeared in papers of the 1930s that developed probit analysis in bioassay. Yuan (1933) introduced the bivariate lognormal distribution.

Later work on genesis of the lognormal distribution is most appropriately recorded in connection with exposition of the basic methods of genesis in Section 3. Similarly, later developments of properties of lognormal distributions, including distributions of products, quotients, and sums, are discussed in Section 4. The two- and three-parameter univariate (i.e., one-dimensional) lognormal distributions and the multivariate lognormal distribution are defined in Section 2. The later developments in inference and applications of lognormal distributions are considered in the later chapters.

2. DEFINITION AND NOTATION

A positive random variable X is said to be lognormally distributed with two parameters μ and σ^2 if $Y = \ln X$ is normally distributed with mean μ and variance σ^2. The two-parameter lognormal distribution is denoted by $\Lambda(\mu, \sigma^2)$; the corresponding normal distribution is denoted by $N(\mu, \sigma^2)$. The probability density function of X having $\Lambda(\mu, \sigma^2)$ is

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\dfrac{(\ln x - \mu)^2}{2\sigma^2}\right\} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{2.1}$$

Figure 1 illustrates the probability density function of $\Lambda(0, \sigma^2)$ for $\sigma = 0.1, 0.3, 1.0$. The probability density function of $\Lambda(\mu, \sigma^2)$ can be written as $f(x \mid \mu, \sigma^2) = e^{-\mu} f(xe^{-\mu} \mid 0, \sigma^2)$, so it has the same shape as that of $\Lambda(0, \sigma^2)$. The distribution is unimodal and positively skew.

In addition, a random variable X which can take any value exceeding a fixed value τ is said to be lognormally distributed with three parameters τ, μ, and σ^2 if $Y = \ln(X - \tau)$ is $N(\mu, \sigma^2)$. The three-parameter lognormal distribution is denoted by $\Lambda(\tau, \mu, \sigma^2)$. The parameter τ is called the *threshold* parameter. Thus the two-parameter lognormal distribution $\Lambda(\mu, \sigma^2)$ is a special case of the three-parameter lognormal distribution $\Lambda(\tau, \mu, \sigma^2)$ for which $\tau = 0$. But estimation procedures developed for the two-parameter case are not directly applicable to the three-parameter case. (See Chapter 4).

Further consider a vector $(X_1, \ldots, X_n)'$ of positive random variables such that $(Y_1, \ldots, Y_n)' = (\ln X_1 \ldots, \ln X_n)'$ has an n-dimensional normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_n)'$ and variance-

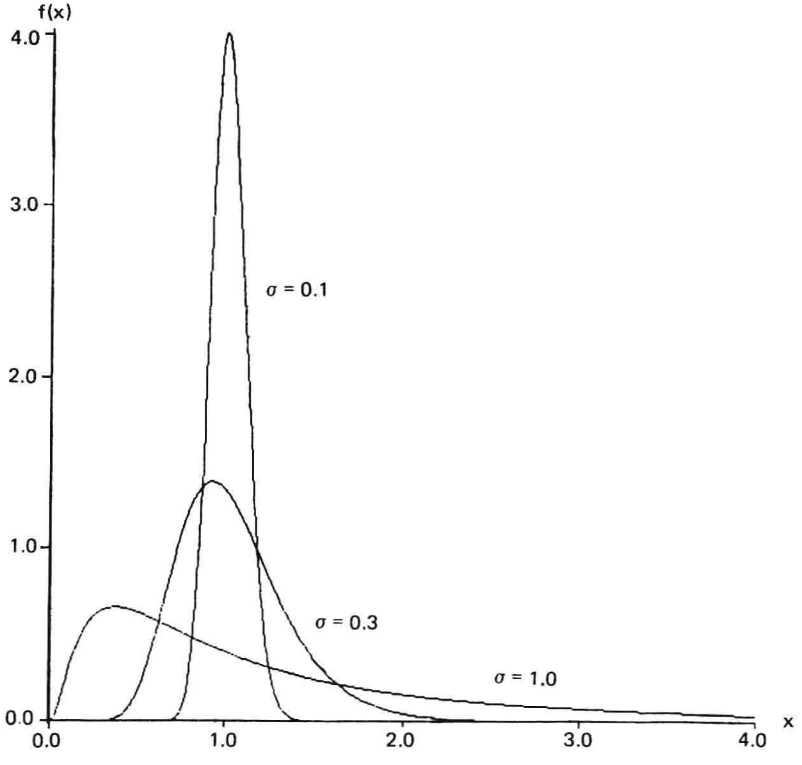


Figure 1 Probability density function of $\Lambda(0, \sigma^2)$.

covariance matrix $\Sigma = (\sigma_{ij})$, $i, j = 1, \ldots, n$, with $\sigma_{ij} = \sigma_{ji}$ and $\sigma_{ii} = \sigma_i^2$. The distribution of $(X_1, \ldots, X_n)'$ is said to be an n-dimensional lognormal distribution with parameters μ and Σ and denoted by $\Lambda_n(\mu, \Sigma)$. The corresponding n-dimensional normal distribution is denoted by $N_n(\mu, \Sigma)$. The probability density function of $(X_1, \ldots, X_n)'$ having $\Lambda_n(\mu, \Sigma)$ is

$$f(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}x_1 \cdots x_n} \exp\left\{-\dfrac{1}{2}(\ln x - \mu)'\Sigma^{-1}(\ln x - \mu)\right\} \\ \qquad x \in R^{n+} \\ 0, \qquad x \in (R^{n+})^c \end{cases}$$

where $x = (x_1, \ldots, x_n)', \ln x = (\ln x_1, \ldots, \ln x_n)'$, and $R^{n+} = \{x \mid x_i > 0$ for all $i = 1, \ldots, n\}$.

3. GENESIS

3.1 The Law of Proportionate Effect

Numerous processes have been devised for generating the lognormal distribution. One of them is the *law of proportionate effect*, so called by Gibrat (1930, 1931).

Suppose that an initial variable X_0 is positive. The equation considered by Kapteyn (1903), which successively calculates the jth step variable X_j, is

$$X_j - X_{j-1} = \varepsilon_j \phi(X_{j-1}) \tag{3.1}$$

where $\{\varepsilon_j\}$ is a set of mutually independent and identically distributed random variables and is also statistically independent of $\{X_j\}$, and ϕ is a certain function. We consider here the important special case $\phi(X) = X$; then the process $\{X_j\}$ is said to obey the law of proportionate effect. Thus (3.1) reduces to

$$X_j = X_{j-1}(1 + \varepsilon_j) \tag{3.2}$$