Wiley Series in Probability and Statistics

WILEY

Univariate Discrete Distributions

Third Edition

Norman L. Johnson
Adrienne W. Kemp
Samuel Kotz

Univariate Discrete Distributions

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

Univariate Discrete Distributions

THIRD EDITION

NORMAN L. JOHNSON

University of North Carolina
Department of Statistics
Chapel Hill, North Carolina

ADRIENNE W. KEMP

University of St. Andrews
Mathematical Institute
North Haugh, St. Andrews
United Kingdom

SAMUEL KOTZ

George Washington University
Department of Engineering Management and Systems Engineering
Washington, D.C.

WILEY-
INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

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

Library of Congress Cataloging-in-Publication Data:

ISBN 0-471-27246-9

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To the memory of Norman Lloyd Johnson (1917–2004)

Contents

Preface

This book is dedicated to the memory of Professor N. L. Johnson, who passed away during the production stages. He was my longtime friend and mentor; his assistance with this revision during his long illness is greatly appreciated. His passing is a sad loss to all who are interested in statistical distribution theory.

The preparation of the third edition gave Norman and I the opportunity to substantially revise and reorganize parts of the book. This enabled us to increase the coverage of certain areas and to highlight today's better understanding of interrelationships between distributions. Also a number of errors and inaccuracies in the two previous editions have been corrected and some explanations are clarified.

The continuing interest in discrete distributions is evinced by the addition of over 400 new references, nearly all since 1992. Electronic databases, such as *Statistical Theory and Methods Abstracts* (published by the International Statistical Institute), the *Current Index to Statistics: Applications, Methods and Theory* (published by the American Statistical Association and the Institute of Mathematical Statistics), and the *Thomson ISI Web of Science*, have drawn to our attention papers and articles which might otherwise have escaped notice.

It is important to acknowledge the impact of scholarly, encyclopedic publications such as the *Dictionary and Bibliography of Statistical Distributions in Scientific Work*, Vol. 1: *Discrete Models*, by G. P. Patil, M. T. Boswell, S. W. Joshi, and M. V. Ratnaparkhi (1984) (published by the International Co-operative Publishing House, Fairland, MD), and the *Thesaurus of Univariate Discrete Probability Distributions*, by G. Wimmer and G. Altmann (1999) (published by Stamm Verlag, Essen). The new edition of *Statistical Distributions*, by M. Evans, N. Peacock, and B. Hastings (2000) (published by Wiley, New York), encouraged us to address the needs of occasional readers as distinct from researchers into the theoretical and applied aspects of the subject.

The objectives of this book are far wider. It aims, as before, to give an account of the properties and the uses of discrete distributions at the time of writing, while adhering to the same level and style as previous editions. The 1969 intention to exclude theoretical minutiae of no apparent practical importance has not been forgotten. We have tried to give a balanced account of new developments, especially those in the more accessible statistical journals. There has also been relevant work in related fields, such as econometrics, combinatorics, probability theory, stochastic processes, actuarial studies, operational research, and social sciences. We have aimed to provide a framework within which future research findings can best be understood.

In trying to keep the book to a reasonable length, some material that should have been included was omitted or its coverage curtailed. Comments and criticisms are welcome; I would like to express our gratitude to friends and colleagues for pointing out faults in the last edition and for their input of ideas into the new edition.

The structure of the book is broadly similar to that of the previous edition. The organization of the increased amount of material into the same number of chapters has, however, created some unfamilar bedfellows. An extra chapter would have had an untoward effect on the next two books in the series (*Univariate Continuous Distributions*, Vols. 1 and 2); these begin with Chapter 12.

Concerning numbering conventions, each chapter is divided into sections and within many sections there are subsections. Instead of a separate name index, the listed references end with section numbers enclosed in square brackets.

Chapter 1 has seen some reordering and the inclusion of a small amount of new, relevant material. Sections 1.1 and 1.2 contain mathematical preliminaries and statistical preliminaries, respectively. Material on the computer generation of specific types of random variables is shifted to appropriate sections in other chapters. We chose not to discuss software explicitly—we felt that this is precluded by shortage of space. Some of the major packages are listed at the end of Chapter 1, however. Many contain modules for tasks associated with specific distributions. Websites are given so that readers can obtain further information.

In Chapter 2, most of the material on distributions based on Lagrangian expansions is moved to Chapter 7, which is now entitled Logarithmic and Lagrangian Distributions. There are new short sections in Chapter 2 on order-k and q-series distributions, mentioning their new placement in the book and changes in customary notations since the last edition.

Chapters 3, 4, and 5 are structurally little changed, although new sections on chain binomial models (Chapter 3), the intervened Poisson distribution (Chapter 4), and the minimum and maximum negative binomial distributions and the condensed negative binomial distribution (Chapter 5) are added.

It is hoped that the limited reordering and insertion of new material in Chapter 6 will improve understanding of hypergeometric-type distributions.

Chapter 7 now has a dual role. Logarithmic distributions occupy the first half. The new second part contains a coherent and updated treatment of the previously fragmented material on Lagrangian distributions.

The typographical changes in Chapters 8 and 9 are meant to make them more reader friendly.

Chapter 10 is now much longer. It contains the section on record value Distributions that was previously in Chapter 11. The treatment of order-k distributions is augmented by accounts of recent researches. The chapter ends with a consolidated account of the absorption, Euler, and Heine distributions, as well as new q-series material, including new work on the null distribution of the Wilcoxon–Mann–Whitney test statistic.

Chapter 11 has seen most change; it is now in two parts. The ability of modern computers to gather and analyze very large data sets with many covariates has led to the construction of many regression-type models, both parametric and nonparametric. The first part of Chapter 11 gives an account of certain regression models for discrete data that are probabilistically fully specified, that is, fully parametric. These include the Tweedie–Poisson family, the Poisson lognormal, Poisson inverse Gaussian, and Poisson polynomial distributions. Efron's double Poisson and double binomial and the simplex-binomial mixture model also receive attention.

The remainder of Chapter 11 is on miscellaneous discrete distributions, as before. Those distributions that have fitted better into earlier chapters are replaced with newer ones, such as the discrete Bessel, the discrete Mittag–Leffler, and the Luria–Delbrück distributions. There is a new section on survival distributions. The section on Zipf and zeta distributions is split into two; renewed interest in the literature in Zipf-type distributions is recognized by the inclusion of Hurwitz–zeta and Lerch distributions.

We have been particularly indebted to Professors David Kemp and "Bala" Balakrishnan, who have read the entire manuscript and have made many valuable recommendations (not always implemented). David was particularly helpful with his knowledge of AMS LaTeX and his understanding of the Wiley stylefile. He has also been of immense help with the task of proofreading. It is a pleasure to record the facilities and moral support provided by the Mathematical Institute at the University of St Andrews, especially by Dr. Patricia Heggie.

Norman and I much regretted that Sam Kotz, with his wide-ranging knowledge of the farther reaches of the subject, felt unable to join us in preparing this new edition.

ADRIENNE W. KEMP

St Andrews, Scotland
November 2004

CHAPTER 1

Preliminary Information

Introduction

This work contains descriptions of many different distributions used in statistical theory and applications, each with its own pecularities distinguishing it from others. The book is intended primarily for reference. We have included a large number of formulas and results. Also we have tried to give adequate bibliographical notes and references to enable interested readers to pursue topics in greater depth.

The same general ideas will be used repeatedly, so it is convenient to collect the appropriate definitions and methods in one place. This chapter does just that. The collection serves the additional purpose of allowing us to explain the sense in which we use various terms throughout the work. Only those properties likely to be useful in the discussion of statistical distributions are described. Definitions of exponential, logarithmic, trigonometric, and hyperbolic functions are not given. Except where stated otherwise, we are using real (not complex) variables, and "log," like "ln," means natural logarithm (i.e., to base e).

A further feature of this chapter is material relating to formulas that will be used only occasionally; where appropriate, comparisons are made with other notations used elsewhere in the literature. In subsequent chapters the reader should refer back to this chapter when an unfamiliar and apparently undefined symbol is encountered.

1.1 MATHEMATICAL PRELIMINARIES

1.1.1 Factorial and Combinatorial Conventions

The number of different orderings of n elements is the product of n with all the positive integers less than n; it is denoted by the familiar symbol $n!$ (*factorial* n),

$$n! = n(n-1)(n-2)\cdots 1 = \prod_{j=0}^{n-1}(n-j). \tag{1.1}$$

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

The less familiar semifactorial symbol $k!!$ means

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

where $k = 2n$.

The product of a positive integer with the next $k-1$ smaller positive integers is called a *descending (falling) factorial*; it will in places be denoted by

$$\begin{aligned} n^{(k)} &= n(n-1)\cdots(n-k+1) \\ &= \prod_{j=0}^{k-1}(n-j) = \frac{n!}{(n-k)!}, \end{aligned} \tag{1.2}$$

in accordance with earlier editions of this book. Note that there are k terms in the product and that $n^{(k)} = 0$ for $k > n$, where n is a positive integer. Readers are WARNED that there is no universal notation for descending factorials in the statistical literature. For example, Mood, Graybill, and Boes (1974) use the symbol $(n)_k$ in the sense $(n)_k = n(n-1)\cdots(n-k+1)$, while Stuart and Ord (1987) write $n^{[k]} = n(n-1)\cdots(n-k+1)$; Wimmer and Altmann (1999) use $x_{(n)} = x(x-1)(x-2)\cdots(x-n+1)$, $x \in \mathbb{R}$, $n \in \mathbb{N}$.

Similarly there is more than one notation in the statistical literature for *ascending (rising) factorials*; for instance, Wimmer and Altmann (1999) use $x^{(n)} = x(x+1)(x+2)\cdots(x+n-1)$, $x \in \mathbb{R}$, $n \in \mathbb{N}$. In the first edition of this book we used

$$\begin{aligned} n^{[k]} &= n(n+1)\cdots(n+k-1) \\ &= \prod_{j=0}^{k-1}(n+j) = \frac{(n+k-1)!}{(n-1)!}. \end{aligned} \tag{1.3}$$

There is, however, a standard notation in the mathematical literature, where the symbol $(n)_k$ is known as *Pochhammer's symbol* after the German mathematician L. A. Pochhammer [1841–1920]; it is used to denote

$$(n)_k = n(n+1)\cdots(n+k-1) \tag{1.4}$$

[this definition of $(n)_k$ differs from that of Mood et al. (1974)]. We will use Pochhammer's symbol, meaning (1.4) except where it conflicts with the use of (1.3) in earlier editions.

The *binomial coefficient* $\binom{n}{r}$ denotes the number of different possible combinations of r items from n different items. We have

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{n}{n-r}; \tag{1.5}$$

also

$$\binom{n}{0} = \binom{n}{n} = 1 \qquad \text{and} \qquad \binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}. \tag{1.6}$$

It is usual to define $\binom{n}{r} = 0$ if $r < 0$ or $r > n$. However,

$$\begin{aligned}\binom{-n}{r} &= \frac{(-n)(-n-1)\cdots(-n-r+1)}{r!} \\ &= (-1)^r \binom{n+r-1}{r}.\end{aligned} \tag{1.7}$$

The *binomial theorem* for a positive integer power n is

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^{n-j} b^j. \tag{1.8}$$

Putting $a = b = 1$ gives

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$$

and putting $a = 1$, $b = -1$ gives

$$\binom{n}{0} - \binom{n}{1} + \cdots + (-1)^n \binom{n}{n} = 0.$$

More generally, for any real power k

$$(1+b)^k = \sum_{j=0}^{\infty} \binom{k}{j} a^j, \qquad -1 < b < 1. \tag{1.9}$$

By equating coefficients of x in $(1+x)^{a+b} = (1+x)^a(1+x)^b$, we obtain the well-known and useful identity known as *Vandermonde's theorem* (A. T. Vandermonde [1735–1796]):

$$\binom{a+b}{n} = \sum_{j=0}^{n} \binom{a}{j}\binom{b}{n-j}. \tag{1.10}$$

Hence

$$\binom{2n}{n} = \binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2.$$

The *multinomial coefficient* is

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}, \tag{1.11}$$

where $r_1 + r_2 + \cdots + r_k = n$.

The *multinomial theorem* is a generalization of the binomial theorem:

$$\left(\sum_{j=1}^{k} a_j\right)^n = \sum \left(\frac{n! \prod_{i=1}^{k} a_i^{n_i}}{\prod_{i=1}^{k} n_i!}\right), \tag{1.12}$$

where summation is over all sets of nonnegative integers $n_1, n_2, \ldots, n_k$ that sum to n.

There are four ways in which a sample of k elements can be selected from a set of n distinguishable elements:

Order Important?	Repetitions Allowed?	Name of Sample	Number of Ways to Select Sample
No	No	k-Combination	$C(n,k)$
Yes	No	k-Permutation	$P(n,k)$
No	Yes	k-Combination with replacement	$C^R(n,k)$
Yes	Yes	k-Permutation with replacement	$P^R(n,k)$

where

$$\begin{aligned} C(n,k) &= \frac{n!}{k!(n-k)!}, & P(n,k) &= \frac{n!}{(n-k)!}, \\ C^R(n,k) &= \frac{(n+k-1)!}{k!(n-1)!}, & P^R(n,k) &= n^k. \end{aligned} \tag{1.13}$$

The number of ways to arrange n distinguishable items in a row is $P(n,n) = n!$ (the number of permutations of n items).

The number of ways to arrange n items in a row, assuming that there are k types of items with n_i nondistinguishable items of type i, $i = 1, 2, \ldots, k$, is the multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_k}$.

The number of derangements of n items (permutations of n items in which item i is not in the ith position) is

$$D_n = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots (-1)^n \frac{1}{n!}\right).$$

The signum function, $\operatorname{sgn}(\cdot)$, shows whether an argument is greater or less than zero:

$$\operatorname{sgn}(x) = 1 \text{ when } x > 0; \quad \operatorname{sgn}(0) = 0; \quad \operatorname{sgn}(x) = -1 \quad \text{when} \quad x < 0.$$

The ceiling function, $\lceil x \rceil$, is the least integer that is not smaller than x, for example,

$$\lceil e \rceil = 3, \qquad \lceil 7 \rceil = 7, \qquad \lceil -2.4 \rceil = -2.$$

The floor function, $\lfloor x \rfloor$, is the greatest integer that is not greater than x, for example,

$$\lfloor e \rfloor = 2, \qquad \lfloor 7 \rfloor = 7, \qquad \lfloor -2.4 \rfloor = -3.$$

The notation $[\cdot] = \lfloor \cdot \rfloor$ is called the integer part.

$$\pi = 4 \sum_{j=0}^{\infty} \frac{(-1)^j}{2j+1} = 3.1415926536,$$

$$e = \sum_{j=0}^{\infty} \frac{1}{j!} = 2.7182818285,$$

$$\ln 2 = \sum_{j=0}^{\infty} \frac{(-1)^{j-1}}{j} = 0.6931471806.$$

1.1.2 Gamma and Beta Functions

When n is real but is *not* a positive integer, meaning can be given to $n!$, and hence to (1.2), (1.3), (1.5), (1.7), and (1.11), by defining

$$(n-1)! = \Gamma(n), \qquad n \in \mathbb{R}^+, \tag{1.14}$$

where $\Gamma(n)$ is the *gamma function*.

The binomial theorem can thereby be shown to hold for any real power.

There are three equivalent definitions of the gamma function, due to L. Euler [1707–1783], C. F. Gauss [1777–1855], and K. Weierstrass [1815–1897], respectively:

Definition 1 *(Euler)*:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t}\, dt, \qquad x > 0. \tag{1.15}$$

Definition 2 *(Gauss)*:

$$\Gamma(x) = \lim_{n\to\infty}\left[\frac{n!n^x}{x(x+1)\cdots(x+n)}\right], \qquad x \neq 0, -1, -2, \ldots. \tag{1.16}$$

Definition 3 *(Weierstrass)*:

$$\frac{1}{\Gamma(x)} = xe^{\gamma x}\prod_{n=1}^{\infty}\left[\left(1+\frac{x}{n}\right)\exp\left(-\frac{x}{n}\right)\right], \qquad x > 0, \tag{1.17}$$

where γ is *Euler's constant*

$$\gamma = \lim_{n\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\ln n\right) \cong 0.5772156649\ldots. \tag{1.18}$$

From Definition 1, $\Gamma(1) = 0! = 1$.
Using integration by parts, Definition 1 gives the recurrence relation for $\Gamma(x)$:

$$\Gamma(x+1) = x\Gamma(x) \tag{1.19}$$

[when x is a positive integer, $\Gamma(x+1) = x!$]. This enables us to define $\Gamma(x)$ over the entire real line, except where x is zero or a negative integer, as

$$\Gamma(x) = \begin{cases} \displaystyle\int_0^{\infty} t^{x-1}e^{-t}\,dt, & x > 0, \\ x^{-1}\Gamma(x+1), & x < 0, \qquad x \neq -1, -2, \ldots. \end{cases} \tag{1.20}$$

From Definition 3 it can be shown that $\Gamma\left(\frac{1}{2}\right) = \pi^{1/2}$; this implies that

$$\int_0^{\infty}\frac{e^{-t}}{t^{1/2}}\,dt = \sqrt{\pi};$$

hence, by taking $t = u^2$, we obtain

$$\int_0^{\infty}\exp\left(\frac{-u^2}{2}\right)du = \sqrt{\frac{\pi}{2}}. \tag{1.21}$$

Also, from $\Gamma\left(\frac{1}{2}\right) = \pi^{1/2}$, we have

$$\Gamma\left(n+\tfrac{1}{2}\right) = \frac{(2n)!\pi^{1/2}}{n!2^{2n}}, \tag{1.22}$$

Definition 3 and the product formula

$$\sin(\pi x) = \pi x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2}\right) \tag{1.23}$$

together imply that

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}, \qquad x \neq 0, -1, -2, \ldots. \tag{1.24}$$

Legendre's duplication formula [A.-M. Legendre, 1752–1833] is

$$\sqrt{\pi}\,\Gamma(2x) = 2^{2x-1}\Gamma(x)\Gamma\left(x+\tfrac{1}{2}\right), \qquad x \neq 0, -\tfrac{1}{2}, -1, -\tfrac{3}{2}, \ldots. \tag{1.25}$$

Gauss's multiplication theorem is

$$\begin{aligned}\Gamma(mx) = (2\pi)^{(1-m)/2} m^{mx-1/2} \prod_{j=1}^{m} \Gamma\left(x + \frac{j-1}{m}\right),&\\ x \neq 0, -\frac{1}{m}, -\frac{2}{m}, -\frac{3}{m}, \ldots,&\end{aligned} \tag{1.26}$$

where $m = 1, 2, 3, \ldots$. This clearly reduces to Legendre's duplication formula when $m = 2$.

Many approximations for probabilities and cumulative probabilities have been obtained using various forms of *Stirling's expansion* [J. Stirling, 1692–1770] for the gamma function:

$$\begin{aligned}\Gamma(x+1) &\sim (2\pi)^{1/2}(x+1)^{x+1/2}e^{-x-1}\\ &\quad \times \exp\left(\frac{1}{12(x+1)} - \frac{1}{360(x+1)^3} + \frac{1}{1260(x+1)^5} - \cdots\right),\end{aligned} \tag{1.27}$$

$$\begin{aligned}\Gamma(x+1) &\sim (2\pi)^{1/2}x^{x+1/2}e^{-x}\\ &\quad \times \exp\left(\frac{1}{12x} - \frac{1}{360x^3} + \frac{1}{1260x^5} - \frac{1}{1680x^7} + \cdots\right),\end{aligned} \tag{1.28}$$

$$\begin{aligned}\Gamma(x+1) &\sim (2\pi)^{1/2}(x+1)^{x+1/2}e^{-x-1}\\ &\quad \times \left(1 + \frac{1}{12(x+1)} + \frac{1}{288(x+1)^2} - \cdots\right),\end{aligned} \tag{1.29}$$

$$\begin{aligned}\Gamma(x+1) &\sim (2\pi)^{1/2}x^{x+1/2}e^{-x}\\ &\quad \times \left(1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51{,}840x^3} - \frac{571}{2{,}488{,}320x^4} + \cdots\right).\end{aligned} \tag{1.30}$$

These are divergent asymptotic expansions, yielding extremely good approximations. The remainder terms for (1.27) and (1.28) are each less in absolute value than the first term that is neglected, and they have the same sign.

Barnes's expansion [E. W. Barnes, 1874–1953] is less well known, but it is useful for half integers:

$$\Gamma\left(x+\tfrac{1}{2}\right) \sim (2\pi)^{1/2} x^x e^{-x} \exp\left(-\frac{1}{24x} + \frac{7}{2880x^3} - \frac{31}{40320x^5} + \cdots\right). \quad (1.31)$$

Also

$$\frac{\Gamma(x+a)}{\Gamma(x+b)} \sim x^{a-b}\left(1 + \frac{(a-b)(a+b-1)}{2x} + \cdots\right). \quad (1.32)$$

These also are divergent asymptotic expansions. Series (1.31) has accuracy comparable to (1.27) and (1.28).

The *beta function* $B(a, b)$ is defined by the *Eulerian integral of the first kind*:

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt, \qquad a>0, \qquad b>0. \quad (1.33)$$

Clearly $B(a, b) = B(b, a)$. Putting $t = u/(1+u)$ gives

$$B(a,b) = \int_0^\infty \frac{u^{a-1}\,du}{(1+u)^{a+b}}\,du, \qquad a>0, \qquad b>0. \quad (1.34)$$

The relationship between the beta and gamma functions is

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \qquad a, b \neq 0, -1, -2 \ldots. \quad (1.35)$$

The derivatives of the logarithm of $\Gamma(a)$ are also useful, though they are not needed as often as the gamma function itself. The function

$$\psi(x) = \frac{d}{dx}[\ln \Gamma(x)] = \frac{\Gamma'(x)}{\Gamma(x)} \quad (1.36)$$

is called the *digamma function* (with argument x) or the *psi function*. Similarly

$$\psi'(x) = \frac{d}{dx}[\psi(x)] = \frac{d^2}{dx^2}[\ln \Gamma(x)]$$

is called the *trigamma function*, and generally

$$\psi^{(s)}(x) = \frac{d^s}{dx^s}[\psi(x)] = \frac{d^{s+1}}{dx^{s+1}}[\ln\Gamma(x)] \tag{1.37}$$

is called the $(s+2)$-*gamma function*. Extensive tables of the digamma, trigamma, tetragamma, pentagamma, and hexagamma functions are contained in Davis (1933, 1935). Shorter tables are in Abramowitz and Stegun (1965).

The recurrence formula (1.19) for the gamma function yields the following recurrence formulas for the psi function:

$$\psi(x+1) = \psi(x) + x^{-1}$$

and

$$\psi(x+n) = \psi(x) + \sum_{j=1}^{n}(x+j-1)^{-1}, \qquad n = 1, 2, 3, \ldots. \tag{1.38}$$

Also

$$\begin{aligned}\psi(x) &= \lim_{n\to\infty}\left[\ln(n) - \sum_{j=0}^{n}(x+j)^{-1}\right] \\ &= -\gamma - \frac{1}{x} + \sum_{j=1}^{\infty}\frac{x}{j(x+j)} &(1.39) \\ &= -\gamma + (x-1)\sum_{j=0}^{\infty}[(j+1)(j+x)]^{-1} &(1.40)\end{aligned}$$

and

$$\psi(mx) = \ln(m) + \frac{1}{m}\sum_{j=0}^{m-1}\psi\left(x+\frac{j}{m}\right), \qquad m = 1, 2, 3, \ldots, \tag{1.41}$$

where γ is Euler's constant ($\cong 0.5772156649\ldots$).

An asymptotic expansion for $\psi(x)$ is

$$\psi(x) \sim \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \cdots, \tag{1.42}$$

and hence a very good approximation for $\psi(x)$ is $\psi(x) \approx \ln(x-0.5)$, provided that $x \geq 2$. Particular values of $\psi(x)$ are

$$\psi(1) = -\gamma, \qquad \psi\left(\tfrac{1}{2}\right) = -\gamma - 2\ln(2) \approx -1.963510\ldots.$$

1.1.3 Finite Difference Calculus

The *displacement operator* E increases the argument of a function by unity:

$$E[f(x)] = f(x+1),$$
$$E[E[f(x)]] = E[f(x+1)] = f(x+2).$$

More generally,

$$E^n[f(x)] = f(x+n) \tag{1.43}$$

for any positive integer n, and we interpret $E^h[f(x)]$ as $f(x+h)$ for any real h.

The *forward-difference operator* Δ is defined by

$$\Delta f(x) = f(x+1) - f(x). \tag{1.44}$$

Noting that

$$f(x+1) - f(x) = E[f(x)] - f(x) = (E-1)f(x),$$

we have the *symbolic* (or *operational*) relation

$$\Delta \equiv E - 1. \tag{1.45}$$

If n is an integer, then the nth *forward difference* of $f(x)$ is

$$\begin{aligned}\Delta^n f(x) = (E-1)^n f(x) &= \sum_{j=0}^{n} \binom{n}{j} (-1)^j E^{n-j} f(x) \\ &= \sum_{j=0}^{n} \binom{n}{j} (-1)^j f(x+n-j).\end{aligned} \tag{1.46}$$

Also, rewriting (1.45) as $E = 1 + \Delta$, we have

$$f(x+n) = (1+\Delta)^n f(x) = \sum_{j=0}^{n} \binom{n}{j} \Delta^j f(x). \tag{1.47}$$

Newton's forward-difference (interpolation) formula [I. Newton, 1642–1727] is obtained by replacing n by h, where h may be any real number, and using the interpretation of $E^h[f(x)]$ as $f(x+h)$:

$$f(x+h) = (1+\Delta)^h = f(x) + h\,\Delta f(x) + \frac{h(h-1)}{2!}\Delta^2 f(x) + \cdots. \tag{1.48}$$

The series on the right-hand side need not terminate. However, if h is small and $\Delta^n f(x)$ decreases rapidly enough as n increases, then a good approximation to $f(x+h)$ may be obtained with but few terms of the expansion. This expansion may then be used to interpolate values of $f(x+h)$, given values $f(x)$, $f(x+1), \ldots,$ at unit intervals.

The *backward-difference operator* ∇ is defined similarly, by the equation

$$\nabla f(x) = f(x) - f(x-1) = (1 - E^{-1}) f(x). \tag{1.49}$$

Note that $\nabla \equiv \Delta E^{-1} \equiv E^{-1}\Delta$. There is a backward-difference interpolation formula analogous to Newton's forward-difference formula.

The *central-difference operator* δ is defined by

$$\begin{aligned} \delta f(x) &= f\left(x + \tfrac{1}{2}\right) - f\left(x - \tfrac{1}{2}\right) \\ &= (E^{1/2} - E^{-1/2}) f(x). \end{aligned} \tag{1.50}$$

Note that $\delta \equiv \Delta E^{-1/2} \equiv E^{-1/2}\Delta$. *Everett's central-difference interpolation formula* [W. N. Everett, 1924–]

$$\begin{aligned} f(x+h) = (1-h)f(x) &+ hf(x+1) - \tfrac{1}{6}(1-h)[1-(1-h)^2]\delta^2 f(x) \\ &- \tfrac{1}{6}h(1-h^2)\delta^2 f(x+1) + \cdots \end{aligned}$$

is especially useful for computation.

Newton's forward-difference formula (1.48) can be rewritten as

$$f(x+h) = \sum_{j=0}^{\infty} \binom{h}{j} \Delta^j f(x). \tag{1.51}$$

If $f(x)$ is a polynomial of degree N, this expansion ends with the term containing $\Delta^N f(x)$.

Applying the difference operator Δ to the descending factorial $x^{(N)}$ gives

$$\begin{aligned} \Delta x^{(N)} &= (x+1)^{(N)} - x^{(N)} \\ &= (x+1)x(x-1)\cdots(x-N+2) - x(x-1)(x-2)\cdots(x-N+1) \\ &= [(x+1) - (x-N+1)]x(x-1)\cdots(x-N+2) \\ &= Nx^{(N-1)}. \end{aligned} \tag{1.52}$$

Repeating the operation, we have

$$\Delta^j x^{(N)} = N^{(j)} x^{(N-j)}, \qquad j \le N. \tag{1.53}$$

For $j > N$ we have $\Delta^j x^{(N)} = 0$.

Putting $x = 0$, $h = x$, and $f(x) = x^n$ in (1.51) gives

$$x^n = \sum_{k=0}^{n} \binom{x}{k} \Delta^k 0^n = \sum_{k=0}^{n} \frac{S(n,k)x!}{(x-k)!}, \tag{1.54}$$

where $\Delta^k 0^n / k!$ in (1.54) means $\Delta^k x^n / k!$ evaluated at $x = 0$ and is called a *difference of zero*. The multiplier $S(n,k) = \Delta^k 0^n / k!$ of the descending factorials in (1.54) is called a *Stirling number of the second kind.*

Equation (1.54) can be inverted to give the descending factorials as polynomials in x with coefficients called *Stirling numbers of the first kind*:

$$\frac{x!}{(x-n)!} = \sum_{j=0}^{n} s(n,j)x^j. \tag{1.55}$$

These notations for the Stirling numbers of the first and second kinds have won wide acceptance in the statistical literature. However, there are no standard symbols in the mathematical literature. Other notations for the Stirling numbers are as follows:

First Kind	Second Kind	Reference
$s(n,j)$	$S(n,k)$	Riordan (1958)
$\binom{n-1}{j-1} B_{n-j}^{(n)}$	$\binom{n}{n} k B_{n-k}^{(-k)}$	Milne-Thompson (1933)
	$\Delta^k 0^n / k!$	David and Barton (1962)
$S_n^{(j)}$	$\mathfrak{S}_n^{(m)}$	Abramowitz and Stegun (1965)
S_n^j	$\mathfrak{S}_k^n$	Jordan (1950)
S_n^j	σ_n^k	Patil et al. (1984)
$S(n,j)$	$Z(n,k)$	Wimmer and Altmann (1999)

Both sets of numbers are nonzero only for $j = 0, 1, 2, \ldots, n$, $k = 0, 1, 2, \ldots, n$, $n > 0$. For given n or given k, the Stirling numbers of the first kind alternate in sign. The Stirling numbers of the second kind are always positive. An extensive tabulation of the numbers and details of their properties appear in Abramowitz and Stegun (1965) and in Goldberg et al. (1976). The numbers increase very rapidly as their parameters increase.

Useful properties are

$$[\ln(1+x)]^j = j! \sum_{n=j}^{\infty} \frac{s(n,j)x^n}{n!}, \tag{1.56}$$

$$(e^x - 1)^k = k! \sum_{n=k}^{\infty} \frac{S(n, k)x^n}{n!}. \tag{1.57}$$

Also

$$s(n+1, j) = s(n, j-1) - ns(n, j), \tag{1.58}$$

$$S(n+1, k) = kS(n, k) + S(n, k-1), \tag{1.59}$$

and

$$\sum_{j=m}^{n} S(n, j)s(j, m) = \sum_{j=m}^{n} s(n, j)S(j, m) = \delta_{m,n}, \tag{1.60}$$

where $\delta_{m,n}$ is *Kronecker delta* [L. Kronecker, 1823–1891]; that is, $\delta_{m,n} = 1$ for $m = n$ and zero otherwise.

Charalambides and Singh (1988) have written a useful review and bibliography concerning the Stirling numbers and their generalizations. Charalambides's (2002) book deals in depth with many types of special numbers that occur in combinatorics, including generalizations and modifications of the Stirling numbers and the Carlitz, Carlitz–Riordan, Eulerian, and Lah numbers.

The *Bell numbers* are partial sums of Stirling numbers of the second kind,

$$B_m = \sum_{j=0}^{m} S(m, j).$$

The *Catalan numbers* are

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

The *Fibonacci numbers* are

$$\begin{aligned} F_0 &= F_1 = 1, \\ F_2 &= F_0 + F_1 = 2, \\ F_3 &= F_1 + F_2 = 3, \\ F_4 &= F_2 + F_3 = 5, \\ &\vdots \end{aligned}$$

Their generating function is $g(t) = 1/(1 - t - t^2)$.

The *Narayana numbers* are

$$N(n, k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1}.$$

1.1.4 Differential Calculus

Next we introduce from the differential calculus the *differential operator* D, defined by

$$Df(x) = f'(x) = \frac{df(x)}{dx}. \tag{1.61}$$

More generally

$$D^j x^N = N^{(j)} x^{N-j}, \qquad j \leq N. \tag{1.62}$$

Note the analogy between (1.53) and (1.62). If the function $f(x)$ can be expressed in terms of a Taylor series, then the Taylor series is

$$f(x+h) = \sum_{j=0}^{\infty} \left(\frac{h^j}{j!} \right) D^j f(x). \tag{1.63}$$

The operator D acting on $f(x)$ formally satisfies

$$\sum_{j=0}^{\infty} \frac{(hD)^j}{j!} \equiv e^{hD}. \tag{1.64}$$

Comparing (1.48) with (1.63), we have (again formally)

$$e^{hD} \equiv (1+\Delta)^h \qquad \text{and} \qquad e^D \equiv 1 + \Delta. \tag{1.65}$$

Although this is only a *formal* relation between operators, it gives exact results when $f(x)$ is a polynomial of finite order; it gives useful approximations in many other cases, especially when $D^j f(x)$ and $\Delta^j f(x)$ decrease rapidly as j increases.

Rewriting $e^D \equiv 1 + \Delta$ as $D \equiv \ln(1+\Delta)$, we obtain a *numerical differentiation* formula

$$f'(x) = Df(x) = \Delta f(x) - \tfrac{1}{2}\,\Delta^2 f(x) + \tfrac{1}{3}\,\Delta^3 f(x) - \cdots. \tag{1.66}$$

(This is not the only numerical differentiation formula. There are others that are sometimes more accurate. This one is quoted as an example.)

Given a change of variable, $x = (1+t)$, we have

$$[D^k f(x)]_{x=1+t} = D^k f(1+t). \tag{1.67}$$

Consider now the *differential operator* θ, defined by

$$\theta f(x) = xDf(x) = xf'(x) = x\frac{df(x)}{dx}. \tag{1.68}$$

This satisfies

$$\theta^k f(x) = \sum_{j=1}^{k} S(k, j) x^j D^j f(x) \tag{1.69}$$

and

$$x^k D^k f(x) = \theta(\theta - 1) \cdots (\theta - k + 1) f(x). \tag{1.70}$$

Also

$$[\theta^k f(x)]_{x=e^t} = D^k f(e^t), \tag{1.71}$$

$$e^{-ct}[\theta^k f(x)]_{x=e^t} = (D + c)^k [e^{-ct} f(e^t)], \tag{1.72}$$

and

$$\begin{aligned} x^c \theta^k [x^{-c} f(x)] &= [e^{ct} D^k \{e^{-ct} f(e^t)\}]_{e^t = x} \\ &= [(D - c)^k f(e^t)]_{e^t = x} \\ &= (\theta - c)^k f(x). \end{aligned} \tag{1.73}$$

The D and θ operators are useful for handling moment properties of distributions.

Lagrange's expansion [J. L. Lagrange, 1736–1813] for the reversal of a power series assumes that if (1) $y = f(x)$, where $f(x)$ is regular in the neighborhood of x_0, (2) $y_0 = f(x_0)$, and (3) $f'(x_0) \neq 0$, then

$$x = x_0 + \sum_{k=1}^{\infty} \frac{(y - y_0)^k}{k!} \left[\frac{d^{k-1}}{dx^{k-1}} \left(\frac{x - x_0}{f(x) - y_0} \right)^k \right]_{x=x_0}. \tag{1.74}$$

More generally

$$h(x) = h(x_0) + \sum_{k=1}^{\infty} \frac{(y - y_0)^k}{k!} \left[\frac{d^{k-1}}{dx^{k-1}} \left\{ h'(x) \left(\frac{x - x_0}{f(x) - y_0} \right)^k \right\} \right]_{x=x_0}, \tag{1.75}$$

where $h(x)$ is infinitely differentiable. (This expansion plays an important role in the theory of Lagrangian distributions; see Section 2.5.)

L'Hôpital's rule [G. F. A. de L'Hôpital, 1661–1704] is useful for finding the limit of an indeterminate form. If $f(x)$ and $g(x)$ are functions of x for which $\lim_{x \to b} f(x) = \lim_{x \to b} g(x) = 0$, and if $\lim_{x \to b}[f'(x)/g'(x)]$ exists, then

$$\lim_{x \to b} \frac{f(x)}{g(x)} = \lim_{x \to b} \frac{f'(x)}{g'(x)}. \tag{1.76}$$

The use of the O, o notation (*Landau's notation*) [E. Landau, 1877–1938] is standard. We say that

$$f(x) = o(g(x)) \qquad \text{as } x \to \infty \qquad \text{if } \lim_{x \to \infty} \left(\frac{f(x)}{g(x)} \right) = 0$$

and

$$f(x) = O(g(x)) \qquad \text{as } x \to \infty \qquad \text{if } \left|\frac{f(x)}{g(x)}\right| < C \tag{1.77}$$

for some constant C and large x.

1.1.5 Incomplete Gamma and Beta Functions and Other Gamma-Related Functions

In statistical work we often encounter the *incomplete gamma function* $\gamma(a, x)$ and its complement $\Gamma(a, x)$; see Khamis (1960) for a discussion of incomplete gamma function expansions of statistical distribution functions. These functions are defined by

$$\begin{aligned} \gamma(a, x) &= \int_0^x t^{a-1} e^{-t}\, dt, \\ \Gamma(a, x) &= \int_x^\infty t^{a-1} e^{-t}\, dt, \qquad x > 0; \end{aligned} \tag{1.78}$$

that is,

$$\gamma(a, x) + \Gamma(a, x) = \Gamma(a).$$

The notation $\Gamma_x(a) = \gamma(a, x)$ is also in use.

Infinite-series formulas are

$$\begin{aligned} \gamma(a, x) &= a^{-1} x^a\, {}_1F_1[a; a + 1; -x] = \sum_{n=0}^{\infty} \frac{(-1)^n x^{a+n}}{n!(a + n)} \\ &= a^{-1} x^a e^{-x}\, {}_1F_1[1; a + 1; x] = e^{-x} \sum_{n=0}^{\infty} \frac{(a - 1)! x^{a+n}}{(a + n)!}, \end{aligned} \tag{1.79}$$

$a \neq 0, -1, -2, \ldots,$ where ${}_1F_1[\cdot]$ is a confluent hypergeometric function; see Section 1.1.7.

The following recursion formulas are useful:

$$\begin{aligned} \gamma(a + 1, x) &= a\gamma(a, x) - x^a e^{-x}, \\ \Gamma(a + 1, x) &= a\Gamma(a, x) + x^a e^{-x}. \end{aligned} \tag{1.80}$$

For x real, $x \to \infty$,

$$\Gamma(a, x) \sim x^{a-1} e^{-x} \left[1 + \frac{a - 1}{x} + \frac{(a - 1)(a - 2)}{x^2} + \cdots\right].$$

The *incomplete gamma function ratio*

$$\frac{\Gamma_x(a)}{\Gamma(a)} = \frac{\gamma(a, x)}{\Gamma(a)}$$

is used in the statistical literature more than $\Gamma_x(a) = \gamma(a, x)$ itself. (The word "ratio" is, alas, sometimes omitted.)

The function tabulated in Pearson's (1922) tables is

$$I(u, p) = \frac{\Gamma_{u\sqrt{p+1}}(p+1)}{\Gamma(p+1)}; \tag{1.81}$$

it is given to seven decimal places for $p = -1(0.05)0(0.1)5(0.2)50$, with u at intervals of 0.1. Harter (1964) gave $I(u, p)$ to nine decimal places for $p = -0.5(0.5)74(1)164$ and u at intervals of 0.1. We note also the extensive tables of Khamis and Rudert (1965).

Pearson and Hartley (1976) [see also Abramowitz and Stegun (1965)] tabulated the function

$$Q(\chi^2|\nu) = \frac{\Gamma(\nu/2, \chi^2/2)}{\Gamma(\nu/2)}$$

(the upper tail of a χ^2 distribution) for

$$\chi^2 = 0.001(0.001)0.01(0.01)0.1(0.1)2(0.2)10(0.5)20(1)40(2)76,$$
$$\nu = 1(1)30$$

to five decimal places.

Just as we often need the incomplete gamma function, so we need also the *incomplete beta function*

$$B_p(a, b) = \int_0^p t^{a-1}(1-t)^{b-1}\, dt, \qquad 0 < p < 1, \tag{1.82}$$

and the *incomplete beta function ratio*

$$I_p(a, b) = \frac{B_p(a, b)}{B(a, b)}. \tag{1.83}$$

Again the word "ratio" is often omitted. In terms of the hypergeometric function ${}_2F_1[\cdot]$ (cf. Section 1.1.6) we have

$$B_p(a, b) = a^{-1}p^a\, {}_2F_1[a, 1-b; a+1; p] = \sum_{n=0}^{\infty} \frac{(1-b)_n p^{a+n}}{n!(a+n)}. \tag{1.84}$$

The incomplete beta function ratio $I_p(a, b)$ has the following properties:

$$\begin{aligned}
I_p(a, b) &= 1 - I_{1-p}(b, a), \\
I_p(k, n-k+1) &= \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}, \qquad 1 \le k \le n, \\
I_p(a, b) &= pI_p(a-1, b) + (1-p)I_p(a, b-1), \qquad (1.85) \\
(a+b-ap)I_p(a, b) &= a(1-p)I_p(a+1, b-1) + bI_p(a, b+1), \\
(a+b)I_p(a, b) &= aI_p(a+1, b) + bI_p(a, b+1).
\end{aligned}$$

Extensive tables of $I_p(a, b)$ to seven decimal places are contained in Pearson (1934) for $p = 0.01(0.01)1$; $a, b = 0.5(0.5)11(1)50$, $a \ge b$. These may be supplemented for small values of a by the tables of Vogler (1964). Both Pearson and Vogler give values for the complete beta function $B(a, b)$.

Pearson and Hartley (1976) have tabulated the percentage points of the F distribution with upper tail

$$Q(F|\nu_1, \nu_2) = I_p\left(\tfrac{1}{2}\nu_2, \tfrac{1}{2}\nu_1\right),$$

where $p = \nu_2/(\nu_2 + \nu_1 F)$ for

$$\begin{aligned}
Q(F|\nu_1, \nu_2) &= 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, \\
\nu_1 &= 1(1)6, 8, 12, 15, 20, 30, 60, \infty, \\
\nu_2 &= 1(1)30, 40, 60, 120, \infty,
\end{aligned}$$

to at least three significant digits; this table is quoted in Abramowitz and Stegun (1965).

The *Laplace transform* [P. S. Laplace, 1749–1827] of a function $f(t)$ is defined as

$$F(p) = \int_0^\infty f(t)e^{-pt}\, dt. \qquad (1.86)$$

The *error function* $\operatorname{erf}(x)$ is defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt. \qquad (1.87)$$

It is closely related to the *normal distribution function*,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt = 0.5\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \qquad (1.88)$$

Its complement is $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$. Sometimes one sees

$$\mathrm{Erf}(x) = 0.5\sqrt{\pi}\,\mathrm{erf}(x), \qquad \mathrm{Erfc}(x) = 0.5\sqrt{\pi}\,\mathrm{erfc}(x).$$

The *Bessel function of the first kind* $J_\nu(x)$ is

$$J_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{(-x^2/4)^j}{j!\Gamma(\nu+j+1)}, \tag{1.89}$$

where ν is the order of the function. The *modified Bessel function of the first kind* is

$$I_\nu(x) = (-i)^\nu J_\nu(ix) = \left(\frac{x}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{(x^2/4)^j}{j!\Gamma(\nu+j+1)}, \tag{1.90}$$

where $i = \sqrt{-1}$.

The *modified Bessel function of the third kind*, $K_\nu(y)$, is defined as

$$K_\nu(y) = \frac{\pi}{2} \cdot \frac{I_{-\nu}(y) - I_\nu(y)}{\sin(\nu\pi)} \tag{1.91}$$

when ν is not an integer or zero. When ν is an integer or zero, the right-hand side of this definition is replaced by its limiting value; see, for example, Abramowitz and Stegun (1965). Sometimes $K_\nu(y)$ is called the modified Bessel function of the *second* kind in the statistical literature.

Useful properties are

$$K_{-\nu}(y) = K_\nu(y) \tag{1.92}$$

and the recurrence relation

$$K_{\nu+1}(y) = \frac{2\nu}{y} K_\nu(y) + K_{\nu-1}(y). \tag{1.93}$$

The *Riemann zeta function* [G. F. B. Riemann, 1826–1866] is defined by the equation

$$\zeta(x) = \sum_{j=1}^{\infty} j^{-x}. \tag{1.94}$$

The series is convergent for $x > 1$, and it is only for these values of x that we shall use the function. A *generalized form of the Riemann zeta function* is defined by

$$\zeta(x, a) = \sum_{j=1}^{\infty} (j+a)^{-x}, \tag{1.95}$$

where $x > 1$ and $a > 0$.

An approximate formula for $\zeta(x)$ is

$$\zeta(x) \approx 1 + \frac{2x^2 + 8.4x + 21.6}{(x-1)(x+7)2^{x+1}}.$$

Particular values are

$$\zeta(2) = \frac{\pi^2}{6} \qquad \text{and} \qquad \zeta(4) = \frac{\pi^4}{90}.$$

Values of $\zeta(n)$ for $n = 2(1)42$ to 20 decimal places are given in Abramowitz and Stegun (1965).

A general formula for even values of the argument is

$$\zeta(2r) = \frac{(2\pi)^{2r}}{2[(2r)!]}|B_{2r}|, \tag{1.96}$$

where B_{2r} is a Bernoulli number (see Section 1.1.9).

The *Lerch function* is

$$\Phi(z, s, \nu) = \sum_{j=0}^{\infty} \frac{z^j}{(\nu + j)^s}, \qquad \nu, z, s \text{ real}, \qquad \nu \neq 0, -1, -2, \ldots. \tag{1.97}$$

1.1.6 Gaussian Hypergeometric Functions

The *hypergeometric function*, or more precisely the *Gaussian hypergeometric function*, has the form

$$\begin{aligned} {}_2F_1[a, b; c; x] &= 1 + \frac{ab}{c\ 1!}x + \frac{a(a+1)b(b+1)}{c(c+1)\ 2!}x^2 + \cdots \\ &= \sum_{j=0}^{\infty} \frac{(a)_j (b)_j x^j}{(c)_j\ j!}, \qquad c \neq 0, -1, -2, \ldots, \end{aligned} \tag{1.98}$$

where $(a)_j$ is Pochhammer's symbol (1.4). The suffixes refer to the numbers of numerator and denominator parameters—there are two numerator parameters and one denominator parameter. Clearly ${}_2F_1[b, a; c; x] = {}_2F_1[a, b; c; x]$.

We will only be interested in the case where a, b, c, and x are real. If a is a nonpositive integer, then $(a)_j$ is zero for $j > -a$, and the series terminates. When the series is infinite, it is absolutely convergent for $|x| < 1$ and divergent for $|x| > 1$. For $|x| = 1$, it is

1. absolutely convergent if $c - a - b > 0$;
2. conditionally convergent if $-1 < c - a - b \leq 0$, $x = -1$; and
3. divergent if $c - a - b \leq -1$.

When $a = 1$ and $b = c$ (or $b = 1$ and $a = c$), the series becomes $1 + x + x^2 + \cdots$; hence the name "hypergeometric."

Gauss's summation theorem states that, when $x = 1$,

$$ {}_2F_1[a, b; c; x] = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} = \frac{B(c, c-a-b)}{B(c-a, c-b)}, \tag{1.99} $$

where $c - a - b > 0$, $c \neq 0, -1, -2, \ldots$.

When a is a nonpositive integer, $a = -n$ say, and $b = -u$, $c = v - n + 1$, this becomes *Vandermonde's theorem* (see Section 1.1.1):

$$ \sum_{j=0}^{n} \binom{u}{j}\binom{v}{n-j} = \binom{u+v}{n}. \tag{1.100} $$

The Gaussian hypergeometric function satisfies the second-order linear differential equation

$$ x(1-x)\frac{d^2y}{dx^2} + [c - (a+b+1)x]\frac{dy}{dx} - aby = 0, \tag{1.101} $$

or, equivalently,

$$ [\theta(\theta + c - 1) - x(\theta + a)(\theta + b)]y = 0, \tag{1.102} $$

where θ is the differential operator $x(d/dx)$; see Section 1.1.4.

The Gaussian hypergeometric function has been described as "the wooden plough of the nineteenth century"; it occurs frequently in mathematical applications because every linear differential equation of the second order, whose singularities are regular and at most three in number, can be transformed into the hypergeometric equation.

The derivatives are

$$ \begin{aligned} \frac{d}{dx}\,{}_2F_1[a, b; c; x] &= \frac{ab}{c}\,{}_2F_1[a+1, b+1; c+1; x], \\ D^n\,{}_2F_1[a, b; c; x] &= \frac{(a)_n(b)_n}{(c)_n}\,{}_2F_1[a+n, b+n; c+n; x]. \end{aligned} \tag{1.103} $$

Euler's integral for the function is

$$ {}_2F_1[a, b; c; x] = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{a-1}(1-u)^{c-a-1}(1-xu)^{-b}\,du, \tag{1.104} $$

where $c > a > 0$. The function is also a Laplace transform:

$$ {}_2F_1\left[a, b; c; \frac{k}{s}\right] = \frac{s^b}{\Gamma(b)} \int_0^\infty e^{-su} u^{b-1}\,{}_1F_1[a; c; ku]\,du. \tag{1.105} $$

The *Euler transformations* are

$$
\begin{aligned}
{}_2F_1[a, b; c; x] &= (1-x)^{-a}\ {}_2F_1\left[a, c-b; c; \frac{x}{x-1}\right] \\
&= (1-x)^{-b}\ {}_2F_1\left[c-a, b; c; \frac{x}{x-1}\right] \\
&= (1-x)^{c-a-b}\ {}_2F_1[c-a, c-b; c; x]. \qquad (1.106)
\end{aligned}
$$

Hypergeometric representations of elementary functions are

$$
\begin{aligned}
(1-x)^{-a} &= {}_2F_1[a, b; b; x], \\
\ln(1+x) &= x\ {}_2F_1[1, 1; 2; -x], \\
\ln\left(\frac{1+x}{1-x}\right) &= 2x\ {}_2F_1\left[\frac{1}{2}, 1; \frac{3}{2}; x^2\right], \\
\arcsin(x) &= x\ {}_2F_1\left[\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; x^2\right], \\
\arctan(x) &= x\ {}_2F_1\left[\frac{1}{2}, 1; \frac{3}{2}; -x^2\right], \qquad (1.107) \\
\left(\frac{1}{2} + \frac{\sqrt{1-x}}{2}\right)^{1-2a} &= {}_2F_1\left[a, a-\frac{1}{2}; 2a; x\right], \\
&= (1-x)^{1/2}\ {}_2F_1\left[a, a+\frac{1}{2}; 2a; x\right], \\
(1+x)^{-2a} + (1-x)^{-2a} &= 2\ {}_2F_1\left[a, a+\frac{1}{2}; \frac{1}{2}; x^2\right], \\
(1-x)^{-2a-1}(1+x) &= {}_2F_1[a+1, 2a; a; x].
\end{aligned}
$$

A large number of special functions can also be represented as Gaussian hypergeometric functions. The incomplete beta function is

$$
B_p(a, b) = a^{-1} p^a\ {}_2F_1[a, 1-b; a+1; p], \qquad (1.108)
$$

the *Legendre polynomials* are

$$
P_n(x) = {}_2F_1\left[-n, n+1; 1; \tfrac{1}{2}(1-x)\right], \qquad (1.109)
$$

the *Chebyshev polynomials* [P. L. Chebyshev, 1821–1894] are

$$T_n(x) = {}_2F_1\left[-n, n; \tfrac{1}{2}; \tfrac{1}{2}(1-x)\right], \tag{1.110}$$

$$U_n(x) = (n+1)\,{}_2F_1\left[-n, n+2; \tfrac{3}{2}; \tfrac{1}{2}(1-x)\right], \tag{1.111}$$

and the *Jacobi polynomials* [C. G. J. Jacobi, 1804–1851] are

$$P_n^{(a,b)}(x) = \binom{a+n}{n}\,{}_2F_1\left[-n, a+b+n+1; a+1; \tfrac{1}{2}(1-x)\right]. \tag{1.112}$$

For detailed studies of the Gaussian hypergeometric function, including recurrence relationships between contiguous functions, see Bailey (1935), Erdélyi et al. (1953, Vol. 1), Slater (1966), and Luke (1975).

1.1.7 Confluent Hypergeometric Functions (Kummer's Functions)

Notations vary for the *confluent hypergeometric function* (also known as *Kummer's series* [E. E. Kummer, 1810–1893]). We have

$$\begin{aligned}
{}_1F_1[a; c; x] &= 1 + \frac{a}{c1!}x + \frac{a(a+1)}{c(c+1)2!}x^2 + \cdots \\
&= \sum_{j=0}^{\infty} \frac{(a)_j x^j}{(c)_j j!} \\
&= \lim_{|b|\to\infty} {}_2F_1\left[a, b; c; \frac{x}{b}\right], \qquad c \neq 0, -1, -2, \ldots,
\end{aligned} \tag{1.113}$$

where $(a)_j$ is Pochhammer's symbol. Other notations for ${}_1F_1[a; c; x]$ are $M(a; c; x)$ and $\phi(a; c; x)$. The suffixes in ${}_1F_1[a; c; x]$ emphasize that there is one numerator parameter and one denominator parameter. If a is a nonpositive integer, the series terminates. The series converges for all real values of a, c, and x, provided that c is not a nonpositive integer. When $a = c$, $c > 0$, the series becomes the exponential series $1 + x + x^2/2! + x^3/3! + \cdots$.

The confluent hypergeometric function satisfies Kummer's differential equation

$$x\frac{d^2y}{dx^2} + (c-x)\frac{dy}{dx} - ay = 0, \tag{1.114}$$

that is,

$$[\theta(\theta + c - 1) - x(\theta + a)]y = 0, \tag{1.115}$$

where $\theta \equiv x(d/dx)$.

The derivatives of the confluent hypergeometric function are

$$\begin{aligned}\frac{d}{dx}\,{}_1F_1[a;c;x] &= \frac{a}{c}\,{}_1F_1[a+1;c+1;x],\\ D^n\,{}_1F_1[a;c;x] &= \frac{(a)_n}{(c)_n}\,{}_1F_1[a+n;c+n;x].\end{aligned} \tag{1.116}$$

The following integral representation is useful:

$${}_1F_1[a;c;x] = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 u^{a-1}(1-u)^{c-a-1}e^{xu}\,du, \tag{1.117}$$

where $c > a > 0$.

Kummer's first theorem yields the transformation

$${}_1F_1[a;c;x] = e^x\,{}_1F_1[c-a;c;-x]. \tag{1.118}$$

Kummer's second theorem is

$$e^{-x}\,{}_1F_1[a;2a;2x] = {}_0F_1\left[\,;a+\tfrac{1}{2};\tfrac{1}{4}x^2\right], \tag{1.119}$$

where $a + \frac{1}{2}$ is not a negative integer and

$${}_0F_1[\,;c;x] = \lim_{|a|\to\infty}\,{}_1F_1\left[a;c;\frac{x}{a}\right] = \sum_{j=0}^{\infty}\frac{x^j}{(c)_j\,j!},$$

where $c \neq 0, 1, 2, \ldots$ and x is finite.

Kummer's differential equation is also satisfied by

$$\Psi(a,c;x) = x^{-a}\,{}_2F_0\left[a, a-c+1;\ ;-\frac{1}{x}\right], \tag{1.120}$$

$$= \frac{1}{\Gamma(a)}\int_0^{\infty} e^{-xu}u^{a-1}(1+u)^{c-a-1}\,du, \tag{1.121}$$

where $a > 0$, $x > 0$.

The following relationship holds:

$$\begin{aligned}\Psi(a,c;x) = {}& \frac{\Gamma(1-c)}{\Gamma(a-c+1)}\,{}_1F_1[a;c;x]\\ &+ \frac{\Gamma(c-1)x^{1-c}}{\Gamma(a)}\,{}_1F_1[a-c+1;2-c;x],\end{aligned} \tag{1.122}$$

provided that $c \neq 0, \pm 1, \pm 2, \ldots$. Also

$$\Psi(a, c; x) = x^{1-c}\Psi(a - c + 1, 2 - c; x). \tag{1.123}$$

Many functions that are important in distribution theory can be expressed in terms of the confluent hypergeometric function; for example, the incomplete gamma functions are

$$\gamma(a, x) = a^{-1}x^a \ {}_1F_1[a; a + 1; -x], \tag{1.124}$$

$$\Gamma(a, x) = \Gamma(a) - a^{-1}x^a \ {}_1F_1[a; a + 1; -x], \tag{1.125}$$

and the *error functions* are

$$\begin{aligned}\mathrm{Erf}(x) &= \frac{\sqrt{\pi}}{2}\mathrm{erf(x)} = 0.5\ \gamma\left(\frac{1}{2}, x^2\right)\\ &= x\ {}_1F_1\left[\frac{1}{2}; \frac{3}{2}; -x^2\right],\end{aligned} \tag{1.126}$$

$$\mathrm{Erfc}(x) = \frac{\sqrt{\pi}}{2}\mathrm{erfc(x)} = \frac{\sqrt{\pi}}{2} - x\ {}_1F_1\left[\frac{1}{2}; \frac{3}{2}; -x^2\right]. \tag{1.127}$$

The *Hermite polynomials* [Ch. Hermite, 1822–1901] as used in statistics are defined as

$$H_n(x) = \sum_{j=0}^{[n/2]} \frac{(-1)^j n! x^{n-2j}}{(n - 2j)! j! 2^j}, \tag{1.128}$$

where [·] denotes the integer part. Hence

$$\begin{aligned}H_{2n}(x) &= \frac{(-1)^n(2n)!}{n!2^n}\ {}_1F_1\left[-n; \frac{1}{2}; \frac{x^2}{2}\right],\\ H_{2n+1}(x) &= \frac{(-1)^n(2n + 1)!x}{n!2^n}\ {}_1F_1\left[-n; \frac{3}{2}; \frac{x^2}{2}\right];\end{aligned} \tag{1.129}$$

see Stuart and Ord (1987, Sections 6.14–6.15). Fisher (1951, p. xxxi) used the "modified" Hermite polynomials

$$H_n^*(x) = i^{-n}H_n(ix), \qquad \text{where } i = \sqrt{-1}. \tag{1.130}$$

The Bessel functions $J_\nu(x)$, $I_\nu(x)$, and $K_\nu(x)$ [F. W. Bessel, 1784–1846], Whittaker functions [E. T. Whittaker, 1873–1956], Laguerre functions and polynomials [E. N. Laguerre, 1834–1886], and Poisson–Charlier polynomials (S. D. Poisson [1781–1840] and C. L. Charlier [1862–1939]) can also all be represented as confluent hypergeometric functions.

Further details concerning some of these functions can be found in Section 1.1.11. Thorough coverage is in Erdélyi et al. (1953, Vols. 1 and 2) and in the book devoted to confluent hypergeometric functions by Slater (1960). Readers are WARNED, however, that most mathematical texts, including those by Erdélyi and by Abramowitz and Stegun, use slightly different notations for the Hermite polynomials (differing by powers of 2). Slater (1960), Rushton and Lang (1954), and Abramowitz and Stegun (1965) give useful tables.

1.1.8 Generalized Hypergeometric Functions

The *generalized hypergeometric function* is a natural generalization of the Gaussian hypergeometric function. The series is defined as

$$
\begin{aligned}
{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q;\ x] &= {}_pF_q\left[\begin{matrix} a_1, \ldots, a_p;\ x \\ b_1, \ldots, b_q \end{matrix}\right] \\
&= \sum_{j=0}^{\infty} \frac{(a_1)_j \ldots (a_p)_j x^j}{(b_1)_j \ldots (b_q)_j\ j!}, \qquad (1.131)
\end{aligned}
$$

where $b_i \neq 0, -1, -2, \ldots,\ i = 1, \ldots, q$.

There are p numerator parameters and q denominator parameters. Clearly the orderings of the numerator parameters and of the denominator parameters are immaterial. The simplest generalized hypergeometric series is

$$
{}_0F_0[-; -; x] = {}_0F_0[\ ;\ ; x] = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = e^x \qquad (1.132)
$$

(a blank indicates the absence of a parameter).

If one of the numerator parameters a_i, $i = 1, \ldots, p$, is a negative integer, $a_1 = -n$ say, the series terminates and

$$
\begin{aligned}
&{}_pF_q\left[\begin{matrix} -n, a_2, \ldots, a_p; x \\ b_1, \ldots, b_q \end{matrix}\right] \\
&= \sum_{j=0}^{n} \frac{(-n)_j (a_2)_j \ldots (a_p)_j x^j}{(b_1)_j \ldots (b_q)_j\ j!}, \qquad (1.133) \\
&= \frac{(a_2)_n \ldots (a_p)_n (-x)^n}{(b_1)_n \ldots (b_q)_n} \\
&\quad \times {}_{q+1}F_{p-1}\left[\begin{matrix} -n, 1 - b_1 - n, \ldots, 1 - b_q - n; (-1)^{p+q-1} x^{-1} \\ 1 - a_2 - n, \ldots, 1 - a_p - n \end{matrix}\right]. \qquad (1.134)
\end{aligned}
$$

When the series is infinite, it converges for $|x| < \infty$ if $p \leq q$, it converges for $|x| < 1$ if $p = q + 1$, and it diverges for all x, $x \neq 0$ if $p > q + 1$. Furthermore, if

$$s = \sum_{i=1}^{q} b_i - \sum_{i=1}^{p} a_i,$$

then the series with $p = q + 1$ is absolutely convergent for $|x| = 1$ if $s > 0$, is conditionally convergent for $|x| = 1$, $x \neq 1$ if $-1 < s \leq 0$, and is divergent for $|x| = 1$ if $s \leq -1$.

The function is characterized as a power series $\sum_{j=0}^{\infty} A_j x^j$ by the property that A_{j+1}/A_j is a rational function of j.

The function satisfies the differential equation

$$\theta(\theta + b_1 - 1) \cdots (\theta + b_q - 1)y = x(\theta + a_1) \cdots (\theta + a_p)y, \tag{1.135}$$

where θ is the differential operator $x(d/dx)$.

The derivatives are

$$\begin{aligned} &\frac{d}{dx}\, {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; x] \\ &\qquad = \frac{a_1 \ldots a_p}{b_1 \ldots b_q}\, {}_pF_q[a_1 + 1, \ldots, a_p + 1; b_1 + 1, \ldots, b_q + 1; x], \end{aligned} \tag{1.136}$$

$$\begin{aligned} &D^n\, {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; x] \\ &\qquad = \frac{(a_1)_n \ldots (a_p)_n}{(b_1)_n \ldots (b_q)_n}\, {}_pF_q[a_1 + n, \ldots, a_p + n; b_1 + n, \ldots, b_q + n; x]. \end{aligned} \tag{1.137}$$

The *Eulerian integral* generalizes to

$$\begin{aligned} &{}_{p+1}F_{q+1}[a_1, \ldots, a_p, c; b_1, \ldots, b_q, d; x] \\ &\quad = \frac{\Gamma(d)}{\Gamma(c)\Gamma(d - c)} \int_0^1 u^{c-1}(1 - u)^{d-c-1}\, {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; xu]\, du. \end{aligned} \tag{1.138}$$

Also

$$\begin{aligned} &{}_{p+1}F_q[a_1, \ldots, a_p, c; b_1, \ldots, b_q; x] \\ &\qquad = \frac{1}{\Gamma(c)} \int_0^{\infty} e^{-u} u^{c-1}\, {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; xu]\, du. \end{aligned} \tag{1.139}$$

The product of two generalized hypergeometric functions can be expressed as a series in other generalized hypergeometric functions. So can generalized hypergeometric functions with arguments of the form $x = y + z$.

A generalized hypergeometric series tail truncated after $m + 1$ terms can be represented as

$$ {}_{p+1}F_{q+1}[a_1, \ldots, a_p, -m; b_1, \ldots, b_q, -m; x]. $$

Head truncation of the first k terms gives

$$ \frac{(a_1)_k \cdots (a_p)_k x^k}{(b_1)_k \cdots (b_q)_k k!}\ {}_{p+1}F_{q+1}[a_1 + k, \ldots, a_p + k, 1; b_1 + k, \ldots, b_q + k, 1 + k; x]. \tag{1.140} $$

Generalized hypergeometric representations of elementary functions include

$$ \begin{aligned} e^x &= {}_0F_0[-; -; x] = {}_0F_0[\ ;\ ; x], \\ (1 - x)^{-a} &= {}_1F_0[a; -; x] = {}_1F_0[a;\ ; x], \\ \cos(x) &= {}_0F_1\left[-; \tfrac{1}{2}; -\tfrac{1}{4}x^2\right] = {}_0F_1\left[\ ; \tfrac{1}{2}; -\tfrac{1}{4}x^2\right], \\ \sin(x) &= x\,{}_0F_1\left[-; \tfrac{3}{2}; -\tfrac{1}{4}x^2\right] = x\,{}_0F_1\left[\ ; \tfrac{3}{2}; -\tfrac{1}{4}x^2\right], \\ \arctan(x) &= x\,{}_2F_1\left[\tfrac{1}{2}, 1; \tfrac{3}{2}; -x^2\right]. \end{aligned} \tag{1.141} $$

Bessel functions can also be stated this way; for example,

$$ \begin{aligned} J_\nu(x) &= \frac{(x/2)^\nu}{\Gamma(\nu + 1)}\ {}_0F_1\left[\ ; \nu + 1; -\frac{x^2}{4}\right], \\ I_\nu(x) &= \frac{(x/2)^\nu}{\Gamma(\nu + 1)}\ {}_0F_1\left[\ ; \nu + 1; \frac{x^2}{4}\right] \end{aligned} \tag{1.142} $$

(see also Sections 1.1.5 and 1.1.7).

The *Horn–Appell functions* are generalized hypergeometric functions in two variables; they include

$$ \begin{aligned} F_1(a, b, b'; c; x, y) &= \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(b)_m(b')_n}{(c)_{m+n}} \frac{x^m y^n}{m!n!} \\ \Phi_1(a, b; c; x, y) &= \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(b)_n}{(c)_{m+n}} \frac{x^m y^n}{m!n!}, \qquad |x| < 1, \\ \Psi_1(a, b; c, c'; x, y) &= \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(b)_m}{(c)_m(c')_n} \frac{x^m y^n}{m!n!}, \qquad |x| < 1, \\ \Xi_1(a, a', b; c; x, y) &= \sum_{m,n=0}^{\infty} \frac{(a)_m(a')_n(b)_m}{(c)_{m+n}} \frac{x^m y^n}{m!n!}, \qquad |x| < 1, \end{aligned} \tag{1.143} $$

where $(a)_m$ is Pochhammer's symbol.

Extensive treatments of generalized hypergeometric functions (including further references) are provided in the books by Erdélyi et al. (1953, Vol. 1), Rainville (1960), and Slater (1966). Certain useful integrals are in Erdélyi et al. (1954, Vols. 1 and 2) and Exton (1978). More advanced special functions and their statistical applications have been studied by Mathai and Saxena (1973, 1978).

1.1.9 Bernoulli and Euler Numbers and Polynomials

The *Bernoulli numbers* [J. Bernoulli, 1654–1705] $B_0, B_1, \ldots, B_r, \ldots$ are defined by the identity

$$\frac{t}{e^t - 1} = \sum_{n=0}^{\infty} \frac{B_n t^n}{n!}, \tag{1.144}$$

giving

$$B_0 = 1, \qquad B_1 = -\tfrac{1}{2}, \qquad B_2 = \tfrac{1}{6}, \qquad B_4 = -\tfrac{1}{30}, \qquad B_6 = \tfrac{1}{42}, \qquad B_8 = -\tfrac{1}{30},$$

with $B_{2r+1} = 0$ for $r > 0$.

The *Bernoulli polynomials* $B_0(x), B_1(x), \ldots, B_r(x), \ldots$ are defined by the identity

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} \frac{B_n(x) t^n}{n!}. \tag{1.145}$$

Clearly $B_r(0) = B_r$. A useful formula is

$$\sum_{j=1}^{n} j^r = (r+1)^{-1}[B_{r+1}(n+1) - B_{r+1}]. \tag{1.146}$$

The polynomials have the properties that

$$\begin{aligned} \frac{dB_r(x)}{dx} &= rB_{r-1}(x), \\ B_r(x+h) &= \sum_{j=0}^{r} \binom{r}{j} B_j(x) h^{r-j} \end{aligned} \tag{1.147}$$

[symbolically $B_r(x+h) = (E+h)^r B_0(x)$ with the displacement operator E applying to the subscript].

The first seven Bernoulli polynomials are

$$
\begin{aligned}
B_0(x) &= 1,\\
B_1(x) &= x - \tfrac{1}{2},\\
B_2(x) &= x^2 - x + \tfrac{1}{6},\\
B_3(x) &= x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x,\\
B_4(x) &= x^4 - 2x^3 + x^2 - \tfrac{1}{30},\\
B_5(x) &= x^5 - \tfrac{5}{2}x^4 + \tfrac{5}{3}x^3 - \tfrac{1}{6}x,\\
B_6(x) &= x^6 - 3x^5 + \tfrac{5}{2}x^4 - \tfrac{1}{2}x^2 + \tfrac{1}{42}.
\end{aligned}
\tag{1.148}
$$

David et al. (1966) have tabulated the Bernoulli polynomials $B_n(x)$ for $n = 0(1)12$ and the Bernoulli numbers B_n for $n = 1(1)12$.

Let $T_k(n) = 1^k + 2^k + \cdots + n^k = \sum_{j=1}^{n} j^k$. Then

$$T_k(n) = \frac{B_{k+1}(n+1) - B_{k+1}(0)}{k+1}.$$

In particular,

$$
\begin{aligned}
T_1(n) &= \tfrac{1}{2}n(n+1),\\
T_2(n) &= \tfrac{1}{6}n(n+1)(2n+1) = \tfrac{1}{3}T_1(n)(2n+1),\\
T_3(n) &= \tfrac{1}{4}n^2(n+1)^2 = \tfrac{1}{2}T_1(n)(n^2+n),\\
T_4(n) &= \tfrac{1}{5}T_2(n)(3n^2+3n-1),\\
T_5(n) &= \tfrac{1}{3}T_3(n)(2n^2+2n-1),\\
T_6(n) &= \tfrac{1}{7}T_2(n)(3n^4+6n^3-3n+1).
\end{aligned}
$$

The *Euler numbers* E_r are defined by the identity

$$\frac{2e^x}{e^{2x}+1} = \sum_{n=0}^{\infty} \frac{E_n x^n}{n!}. \tag{1.149}$$

They satisfy the symbolic formula

$$(E+1)^n + (E-1)^n = 0, \tag{1.150}$$

with powers of E^m replaced by E_m. We find that $E_{2n+1} = 0$ and that the Euler numbers are all integers for r even:

$$\begin{aligned} E_0 &= 1, \\ E_2 &= -1, \\ E_4 &= 5, \\ E_6 &= -61, \\ E_8 &= 1{,}385, \\ E_{10} &= -50{,}521 \\ &\vdots \end{aligned}$$

Further values are given in Abramowitz and Stegun (1965).

The *Euler polynomials* $E_r(x)$ are defined by the identity

$$\frac{2e^{tx}}{e^t + 1} \equiv \sum_{j=0}^{\infty} E_j(x) \frac{t^j}{j!}. \tag{1.151}$$

Their properties include

$$E_n(x) + E_n(x+1) = 2x^n, \tag{1.152}$$

$$\frac{dE_n(x)}{dx} = nE_{n-1}(x). \tag{1.153}$$

The following symbolic relationships connect the Bernoulli and the Euler numbers:

$$\begin{aligned} E^{n-1} &\equiv \frac{(4B-1)^n - (4B-3)^n}{2n}, \\ E^{2n} &\equiv \frac{4^{2n+1}(B - 1/4)^{2n+1}}{2n+1}. \end{aligned} \tag{1.154}$$

If $m + n$ is odd, then

$$\int_0^1 B_m(x) B_n(x)\, dx = 0 = \int_0^1 E_m(x) E_n(x)\, dx. \tag{1.155}$$

Both the polynomials $B_m(x)$, $B_n(x)$ and the polynomials $E_m(x)$, $E_n(x)$ are *orthogonal* over the interval (0, 1) (see Section 1.1.11), with uniform weight function. For a full discussion of Bernoulli and Euler polynomials, we refer the reader to Nörlund (1923) and Milne-Thompson (1933). Abramowitz and Stegun (1965) give an excellent summary.

1.1.10 Integral Transforms

The *exponential Fourier transform* [J. B. J. Fourier, 1768–1830]

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx \tag{1.156}$$

gives the *characteristic function* of a distribution.

The *Laplace transform*

$$L(p) = \int_{0}^{\infty} e^{-px} f(x)\, dx \tag{1.157}$$

(if it exists) yields the moment generating function $M(t)$ of a distribution with probability density function (pdf) $f(x)$ on the nonnegative real line by setting $t = -p$; that is, $M(t) = L(-t)$.

The *Mellin transform* [R. H. Mellin, 1854–1933] and its *inverse* are

$$H(s) = \int_{0}^{\infty} x^{s-1} f(x)\, dx, \tag{1.158}$$

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} H(s)\, ds. \tag{1.159}$$

If $f(x)$ is a pdf, then (1.158) gives the $(s-1)$th moment about the origin of a distribution on the nonnegative real line. Springer (1979) has demonstrated the key role of the Mellin transform and its inverse in the derivation of distributions of products, quotients, and other algebraic functions of independent random variables.

For a comprehensive coverage of these and other types of integral transforms, see Erdélyi et al. (1954, Vols. 1 and 2).

1.1.11 Orthogonal Polynomials

If the polynomial $P_r(x)$ of degree r is a member of a family of polynomials $\{P_j(x)\}$, $j = 0, 1, \ldots$, and

$$\int_{-\infty}^{\infty} w(x) P_m(x) P_n(x)\, dx = 0 \tag{1.160}$$

is satisfied whenever $m \neq n$, then the family of polynomials is said to be *orthogonal* with respect to the weight function $w(x)$. In particular cases, $w(x)$ may be zero outside certain intervals.

Two families of *orthogonal polynomials* have special importance in distribution theory. These are the Hermite polynomials and the generalized Laguerre polynomials. The *Hermite polynomials* have the weight function

$$w(x) = e^{-x^2/2}. \tag{1.161}$$

The rth Hermite polynomial is defined by

$$H_r(x) = (-1)^r e^{x^2/2} D^r (e^{-x^2/2}), \qquad r = 0, 1, \ldots, \tag{1.162}$$

It follows that

$$\begin{aligned}
H_0(x) &= 1,\\
H_1(x) &= x,\\
H_2(x) &= x^2 - 1,\\
H_3(x) &= x^3 - 3x,\\
H_4(x) &= x^4 - 6x^2 + 3,\\
H_5(x) &= x^5 - 10x^3 + 15x,
\end{aligned}$$

and generally

$$\begin{aligned}
H_r(x) = x^r - \frac{r(r-1)}{1! \cdot 2} x^{r-2} + \frac{r(r-1)(r-2)(r-3)}{2! \cdot 2^2} x^{r-4} - \cdots \\
+ (-1)^j \frac{r!}{(r-2j)!\,j! \cdot 2^j} x^{r-2j} + \cdots
\end{aligned} \tag{1.163}$$

(cf. Section 1.1.7). The series terminates after $j = [r/2]$, where $[r]$ denotes the largest integer less than or equal to r.

The *generalized Laguerre polynomials* have the weight function

$$w(x) = \begin{cases} x^a e^{-x}, & x \geq 0, \qquad a > -1, \\ 0, & x < 0. \end{cases} \tag{1.164}$$

The rth generalized Laguerre polynomial of order a is

$$\begin{aligned}
L_r^{(a)}(x) &= \sum_{j=0}^{r} (-1)^j \binom{r+a}{r-j} \frac{x^j}{j!} \\
&= \binom{r+a}{r} {}_1F_1[-r; a+1; x].
\end{aligned} \tag{1.165}$$

The recurrence formula

$$x L_r^{(a+1)}(x) = (x - r) L_r^{(a)}(x) + (a + r) L_{r-1}^{(a)}(x) \tag{1.166}$$

is useful in computation.

The *Jacobi, Chebyshev, Krawtchouk, and Charlier polynomials* are other families of orthogonal polynomials that are occasionally used in statistical theory. The weight function for the Jacobi polynomial $P_n^{(a,b)}(x)$ is

$$w(x) = \begin{cases} (1-x)^a(1+x)^b, & -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.167}$$

The other three families have the following weight functions:

Chebyshev polynomial $T_n(x)$:

$$w(x) = \begin{cases} (1-x^2)^{-1/2}, & -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.168}$$

Chebyshev polynomial $U_n(x)$:

$$w(x) = \begin{cases} (1-x^2)^{1/2}, & -1 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{1.169}$$

Krawtchouk polynomials:

$$w(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n, \\ 0, & \text{otherwise.} \end{cases} \tag{1.170}$$

Charlier polynomials:

$$w(x) = \begin{cases} e^{-\theta}\theta^x/x!, & x = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases} \tag{1.171}$$

Szegö (1939, 1959, 1967) is a standard work on orthogonal polynomials. Their properties are summarized in Abramowitz and Stegun (1965). Stuart and Ord (1987, Chapter 6) demonstrate some of their statistical uses.

1.1.12 Basic Hypergeometric Series

Heine's generalization of the hypergeometric series [H. E. Heine, 1821–1881] is known as a *basic hypergeometric series*, also as a *q-series* and as a *q-hypergeometric series*; it is defined as

$$
\begin{aligned}
&{}_2\phi_1(a, b; c; q, z) \\
&\quad = 1 + \frac{(1-a)(1-b)z}{(1-c)(1-q)} + \frac{(1-a)(1-aq)(1-b)(1-bq)z^2}{(1-c)(1-cq)(1-q)(1-q^2)} + \cdots \\
&\quad = \sum_{j=0}^{\infty} \frac{(a;q)_j (b;q)_j z^j}{(c;q)_j (q;q)_j},
\end{aligned}
\tag{1.172}
$$

where $|q| < 1$, $|z| < 1$; there are two numerator parameters and one denominator parameter. By $(a; q)_j$ we mean

$$(a;q)_0 = 1, \qquad (a;q)_j = (1-a)(1-aq)\cdots(1-aq^{j-1}).$$

Readers are WARNED that there are several differing notations for this expression in the literature; for example:

$(a;q)_j$	Slater (1966), Andrews (1986), Gasper and Rahman (1990),
$(a)_{q,j}$	Bailey (1935)
$[a]_j$	Jackson (1921)
$[a;q,j]$	Exton (1983)

[For a complete list of F. H. Jackson's numerous publications over the period 1904–1954, see Chaundy (1962).]

The Gaussian (basic) binomial coefficient is

$$\begin{bmatrix} n \\ 0 \end{bmatrix}_q = 1, \qquad \begin{bmatrix} n \\ x \end{bmatrix}_q = \frac{(q;q)_n}{(q;q)_x (q;q)_{n-x}}.$$

The definition of a general basic hypergeometric series (q-series) that was given in the second edition of this book was the one used by Bailey (1935) and Slater (1966):

$$
\begin{aligned}
{}_A\phi_B(a_1, \ldots, a_A; b_1, \ldots, b_B; q, z) &= {}_A\phi_B \begin{bmatrix} a_1, \ldots, a_A; q, z \\ b_1, \ldots, b_B \end{bmatrix} \\
&= \sum_{j=0}^{\infty} \frac{(a_1;q)_j \ldots (a_A;q)_j z^j}{(b_1;q)_j \ldots (b_B;q)_j (q;q)_j}.
\end{aligned}
$$

This is no longer in general use.

The publication of the book by Gasper and Rahman (1990) (G/R) on basic hypergeometric functions has led to the universal adoption of a new notation for generalized basic hypergeometric series (q-series) in the mathematics and physics literature:

$$
\begin{aligned}
&{}_A\phi_B(a_1, \ldots, a_A; b_1, \ldots, b_B; q, z) \\
&\quad = \sum_{j=0}^{\infty} \frac{(a_1; q)_j \ldots (a_A; q)_j z^j}{(b_1; q)_j \ldots (b_B; q)_j (q; q)_j} \left[(-1)^j q^{\binom{j}{2}}\right]^{B-A+1}.
\end{aligned} \tag{1.173}
$$

The only difference between the two definitions is the additional factor

$$[(-1)^j q^{j(j-1)/2}]^{B-A+1};$$

there is no difference when $A = B + 1$. The very considerable advantage conferred by the use of the additional factor when $A \neq B + 1$ is that limiting forms of G/R q-series as parameters tend to zero are themselves q-series.

As $q \to 1$,

$$
\begin{aligned}
\frac{(q^a; q)_j}{(1-q)^j} &= \left(\frac{1-q^a}{1-q}\right)\left(\frac{1-q^{a+1}}{1-q}\right)\cdots\left(\frac{1-q^{a+j-1}}{1-q}\right) \\
&= (1 + q + \cdots + q^{a-1})(1 + q + \cdots + q^a)\cdots(1 + q + \cdots + q^{a+j-2}) \\
&\to a(a+1)\cdots(a+j-1) = (a)_j,
\end{aligned}
$$

where $(a)_j$ is Pochhammer's symbol. It follows that as $q \to 1$ a generalized basic hypergeometric series tends to a generalized hypergeometric series:

$$
\begin{aligned}
&\lim_{q\to 1} {}_A\phi_B(q^{a_1}, \ldots, q^{a_A}; q^{b_1}, \ldots, q^{b_B}; q, (1-q)^{B+1-A} z) \\
&\quad = {}_AF_B[a_1, \ldots, a_A; b_1, \ldots, b_B; z].
\end{aligned}
$$

Heine's theorem

$$ {}_1\phi_0(a; -; q, z) = {}_1\phi_0(a;\ ; q, z) = \prod_{j=0}^{\infty} \frac{1 - aq^j z}{1 - q^j z} \tag{1.174}$$

follows from the relationship

$$(1 - z)\,{}_1\phi_0(a;\ ; q, z) = (1 - az)\,{}_1\phi_0(a;\ ; q, qz).$$

When $a = q^{-k}$ and k is a positive integer, Heine's theorem gives the following q-series analog of the binomial theorem:

$$\prod_{j=0}^{k-1} (1 - q^{j-k} z) = {}_1\phi_0(q^{-k};\ ; q, z). \tag{1.175}$$

Another consequence of Heine's theorem is

$$ {}_1\phi_0(a;\ ;q,z)\,{}_1\phi_0(b;\ ;q,az) = {}_1\phi_0(ab;\ ;q,z). \tag{1.176}$$

Letting $a \to 0$ gives

$$ {}_0\phi_0(-;-;q,z) = {}_0\phi_0(\ ;\ ;q,z) = \prod_{j=0}^{\infty}(1-q^j z)^{-1}; \tag{1.177}$$

that is,

$$\begin{aligned} 1 &+ \frac{z}{1-q} + \frac{z^2}{(1-q)(1-q^2)} + \cdots + \frac{z^j}{(1-q)\cdots(1-q^j)} + \cdots \\ &= (1-z)^{-1}(1-qz)^{-1}(1-q^2z)^{-1}\cdots. \end{aligned} \tag{1.178}$$

If z is replaced by $-z/a$ and $a \to \infty$, we obtain

$$\begin{aligned} 1 &+ \frac{z}{1-q} + \frac{qz^2}{(1-q)(1-q^2)} + \cdots + \frac{q^{j(j-1)/2}z^j}{(1-q)\cdots(1-q^j)} + \cdots \\ &= (1+z)(1+qz)(1+q^2z)\cdots. \end{aligned} \tag{1.179}$$

The general bilateral basic hypergeometric series with base q, $0 < q < 1$, r numerator parameters, and s denominator parameters is

$$\begin{aligned} {}_r\psi_s(a_1,\ldots,a_r;b_1,\ldots,b_s;q,z) &= {}_r\psi_s\left[\begin{matrix} a_1,\ldots,a_r;q,z \\ b_1,\ldots,b_s \end{matrix}\right] \\ &= \sum_{j=-\infty}^{\infty} \frac{(a_1;q)_j\ldots(a_r;q)_j z^j}{(b_1;q)_j\ldots(b_s;q)_j}(-1)^{(s-r)j}\, q^{(s-r)j(j-1)/2}\, z^j \\ &= \sum_{j=-\infty}^{\infty} v_j\, z^j, \end{aligned} \tag{1.180}$$

where v_{j+1}/v_j is a rational function of q^j (Gasper and Rahman, 1990). It is assumed that each term in (1.180) is well defined; this is achieved when $q \neq 0$, $z \neq 0$, and $b_j \neq q^n$, where $n \in \mathbb{Z}^+$, $j = 0, 1, \ldots, s$.

1.2 PROBABILITY AND STATISTICAL PRELIMINARIES

1.2.1 Calculus of Probabilities

A *σ-field* is a collection $\mathcal{F}$ of subsets of a set Ω that contains the empty set (ϕ) as a member and is closed under countable unions and complements.

A *probability measure* P on a σ*-field* $\mathcal{F}$ of subsets of Ω is a function from $\mathcal{F}$ to the unit interval [0, 1] such that $P(\Omega) = 1$ and the probability measure of a countable union of disjoint sets $\{E_i\}$ is equal to $\sum P(E_i)$.

A *probability space* is a triple $(\Omega, \mathcal{F}, P)$, where Ω is a set, $\mathcal{F}$ is a σ*-field*, and P is a *probability measure* on $\mathcal{F}$.

For $\Pr(E)$ to be a probability measure, we require the following *probability axioms* to be satisfied:

1. $0 \le \Pr(E) \le 1$.
2. $\Pr(\Omega) = 1$.
3. If the events E_i are *mutually exclusive*, then $\Pr(\bigcup_i E_i) = \sum_i \Pr(E_i)$.

Probabilities defined in this way accord with the intuitive notion that the probability of an event E is the proportion of times that E might be expected to occur in repeated independent observations under specified conditions and that the probability of E therefore takes some value in the (closed) interval [0, 1]. The probability of an impossibility is taken to be zero, while the sum of the probabilities of all possibilities is deemed to be unity (the probability of a certainty). Given two events that cannot occur simultaneously, then the probability that one or other of them will occur is equal to the sum of their separate probabilities. If all the outcomes are equally likely, then

$$\Pr(A) = \frac{n(A)}{n(\Omega)} = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega}.$$

The compound event "either E_1 or E_2 or both" is called the *logical sum*, or *union*, of E_1 and E_2 and is written symbolically as $E_1 + E_2$ or $E_1 \bigcup E_2$. (The two names and symbols refer to the same concept.)

The compound event "both E_1 and E_2" is called the *logical product*, or *intersection*, of E_1 and E_2 and is written symbolically as $E_1 E_2$ or $E_1 \bigcap E_2$. (Again the two names and symbols refer to the same concept.)

If $\Pr(E_1 \cap E_2) = 0$, the events E_1 and E_2 are *mutually exclusive*.

These definitions can be extended to combinations of any number of events. Thus $E_1 + E_2 + \cdots + E_k$ or $\bigcup_{j=1}^{k} E_j$ means "at least one of $E_1, E_2, \ldots, E_k$," while $E_1 E_2 \cdots E_k$ or $\bigcap_{j=1}^{k} E_j$ means "every one of $E_1, E_2, \ldots, E_k$." By a natural extension we can form such compound events as $(E_1 \cup E_2) \cap E_3$, meaning "both E_3 and at least one of E_1 and E_2." By a further extension we can form compounds of enumerable infinities of events $\bigcup_{j=1}^{\infty} E_j$ and $\bigcap_{j=1}^{\infty} E_j$.

The following theorems hold:

1. $\Pr(\phi) = 0$ where ϕ is the empty set.
2. The event "negation of E" is called the *complement* of E and is often denoted by $\overline{E}$. We have $\Pr(E \cup \overline{E}) = 1$.

3. For any events E_1 and E_2,

$$\Pr(E_1) = \Pr(E_1 \cap E_2) + \Pr(E_1 \cap \overline{E}_2). \tag{1.181}$$

4. *De Morgan's laws* state that

$$\begin{aligned} \Pr(\overline{E}) &= 1 - \Pr(E), \\ \Pr(\overline{E_1 \cup E_2}) &= \Pr(\overline{E_1} \cap \overline{E_2}), \\ \Pr(\overline{E_1 \cap E_2}) &= \Pr(\overline{E_1} \cup \overline{E_2}). \end{aligned} \tag{1.182}$$

5. If $E \subset A$, then $\Pr(E) \leq \Pr(A)$.

6. For any events $E_1, E_2, \ldots, E_n$,

$$\Pr\left(\bigcup_{j=1}^{n} E_j\right) \leq \sum_{j=1}^{n} \Pr(E_j). \tag{1.183}$$

7. An important formula connecting probabilities of different but related events is

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2). \tag{1.184}$$

8. The following extension of this formula is known as *Boole's formula*:

$$\begin{aligned} \Pr\left(\bigcup_{j=1}^{n} E_j\right) &= \sum_{j=1}^{n} \Pr(E_j) - \sum\sum \Pr(E_{j_1} \cap E_{j_2}) \\ &\quad + \sum\sum\sum \Pr(E_{j_1} \cap E_{j_2} \cap E_{j_3}) - \cdots \\ &\quad + (-1)^{n-1} \Pr\left(\bigcap_{j=1}^{n} E_j\right), \end{aligned} \tag{1.185}$$

where a summation sign repeated m times means summation over all integers $j_1, j_2, \ldots, j_m$ subject to $1 \leq j_i \leq n,\ j_1 < j_2 < \cdots < j_m$. [The *inclusion–exclusion principle* is closely related to Boole's formula (see Section 10.2); it is important in the derivation of matching and occupancy distributions.]

9. The absolute values of the terms in (1.185) are nonincreasing. Boole's formula therefore enables bounds to be obtained for $\Pr(\bigcup_{j=1}^{n} E_j)$ by stopping at any two consecutive sets of terms. For example,

$$\sum_{j=1}^{n} \Pr(E_j) - \sum\sum \Pr(E_{j_1} \cap E_{j_2}) \leq \Pr\left(\bigcup_{j=1}^{n} E_j\right) \leq \sum_{j=1}^{n} \Pr(E_j). \tag{1.186}$$

10. If every pair of the events $E_1, E_2, \ldots, E_n$ is mutually exclusive, then Boole's formula becomes

$$\Pr\left(\bigcup_{j=1}^{n} E_j\right) = \sum_{j=1}^{n} \Pr(E_j). \tag{1.187}$$

[The mutually exclusive events $E_1, E_2, \ldots, E_n$ are said to be *exhaustive* if $\sum_{j=1}^{n} \Pr(E_j) = 1$.]

11. The *conditional probability* of event E_1 given that E_2 has occurred is denoted by $\Pr(E_1|E_2)$ and is given by

$$\Pr(E_1|E_2) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)}, \tag{1.188}$$

where $\Pr(E_2) > 0$; therefore

$$\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2|E_1) = \Pr(E_2)\Pr(E_1|E_2). \tag{1.189}$$

12. More generally,

$$\Pr\left(\bigcap_{j=1}^{n} E_j\right) = \Pr(E_1)\Pr(E_2|E_1)\Pr(E_3|E_1 \cap E_2)\cdots\Pr\left(E_n \middle| \bigcap_{j=1}^{n-1} E_j\right). \tag{1.190}$$

13. If $\Pr(E_2|E_1) = \Pr(E_2)$, then E_2 is said to be *independent* of the event E_1 and (1.189) becomes

$$\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2). \tag{1.191}$$

14. We say that n events are *mutually independent* if, for every subset $\{E_{j_1}, E_{j_2}, \ldots, E_{j_k}\}$, $k \leq n$,

$$\Pr(E_{j_1} \cap E_{j_2} \cap \cdots \cap E_{j_k}) = \prod_{i=1}^{k} \Pr(E_{j_i}). \tag{1.192}$$

15. If E_j is independent of $\bigcap_{i=1}^{j-1} E_i$ for all $j \leq n$ (this is certainly true if the n events are mutually independent), then (1.190) simplifies to

$$\Pr(E_1 \cap E_2 \cap \cdots \cap E_n) = \prod_{i=1}^{n} \Pr(E_j). \tag{1.193}$$

16. The *theorem of total probability* states that, if $E_1, E_2, \ldots, E_n$ are mutually exclusive and exhaustive, then

$$\Pr(A) = \sum_{j=1}^{n} \Pr(A|E_j)\Pr(E_j). \tag{1.194}$$

17. *Bayes's theorem* is an important consequence of (1.188) and (1.194) and is central to modern Bayesian methods of inference; see the next section.

1.2.2 Bayes's Theorem

Bayesian methods of inference involve the systematic formulation and use of Bayes's theorem. These approaches are distinguished from other statistical approaches in that, prior to obtaining the data, the statistician formulates *degrees of belief* concerning the possible models that may give rise to the data. These degrees of belief are regarded as probabilities.

Suppose that $\{M_1, M_2, \ldots, M_k\}$ is a mutually exclusive and exhaustive set of possible probability models for the experimental situation of interest, and suppose that $\{D_1, D_2, \ldots, D_r\}$ is the set of possible outcomes when the experiment is carried out. Also let

1. $\Pr(M_i)$, $i = 1, \ldots, k$, be the probability that the correct model is M_i prior to learning the outcome of the experiment;
2. $\Pr(D_j)$, $j = 1, \ldots, r$, be the probability that the result of the experiment is the outcome D_j;
3. $\Pr(D_j|M_i)$ be the probability that model M_i will produce the outcome D_j; and
4. $\Pr(M_i|D_j)$ be the probability that the model M_i is the correct model given that the experiment has had the outcome D_j.

Then, by the definition of conditional probability,

$$\Pr(M_i|D_j)\Pr(D_j) = \Pr(M_i \cap D_j) = \Pr(D_j|M_i)\Pr(M_i),$$

and by the theorem of total probability,

$$\Pr(D_j) = \sum_i \Pr(D_j|M_i)\Pr(M_i);$$

together these lead to the *discrete form of Bayes's theorem*

$$\Pr(M_i|D_j) = \frac{\Pr(D_j|M_i)\Pr(M_i)}{\sum_i \Pr(D_j|M_i)\Pr(M_i)}, \tag{1.195}$$

where

$\Pr(M_i)$ is termed the *prior probability* of the model M_i,

$\Pr(D_j|M_i)$ is termed the *likelihood of the outcome* D_j under the model M_i, and

$\Pr(M_i|D_j)$ is termed the *posterior probability* of the model M_i given that the outcome D_j has occurred.

It follows that

$$\frac{\Pr(M_i|D_j)}{\Pr(\overline{M_i}|D_j)} = \frac{\Pr(D_j|M_i)}{\Pr(D_j|\overline{M_i})}\frac{\Pr(M_i)}{\Pr(\overline{M_i})}. \tag{1.196}$$

Because the ratio $\Pr(A)/[1 - \Pr(A)]$ is called the *odds on A*, the discrete form of Bayes's theorem is sometimes rephrased as "posterior odds are equal to the likelihood ratio times the prior odds."

Suppose now that the models do not form an enumerable set but instead are indexed by a parameter θ. Let $p(\theta)$ be the prior probability density of the parameter θ, let $p(x|\theta)$ be the likelihood that the experiment will yield an observed value x for the random variable X given the value of the parameter θ, and let $p(\theta|x)$ be the posterior probability density of θ given that the experiment has yielded the observation x. Then

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\displaystyle\int_{\Theta} p(x|\theta)p(\theta)\,d\theta}, \tag{1.197}$$

where Θ is the set of all possible values of the parameter θ; that is,

$$\int_{\Theta} p(\theta)\,d\theta = 1.$$

This is the *continuous form of Bayes's theorem*; it is sometimes summarized as "posterior density is proportional to likelihood times prior density."

The resolution of the problem of assigning a prior distribution to the parameter θ by the use of Bayes's postulate caused controversy. *Bayes's postulate* is, in brief, the assumption that, if there is no information to the contrary, then all prior probabilities are to be regarded as equal. This is known as the adoption of a *vague (diffuse, uninformative) prior*, in contradistinction to an *informative prior* that takes into account positive empirical or theoretical information concerning the distribution of θ.

Smith (1984) and Lindley (1990) have written helpful expository articles on Bayesian inference. Books that are written from a Bayesian standpoint include the seminal work by Box and Tiao (1973) and those by O'Hagan (1994) and Congdon (2003). Much of the literature on Bayesian methods is in edited volumes, for example, the Oxford University Press Series, *Bayesian Statistics,* edited by Bernado et al. (1992, 1996, 1999).

1.2.3 Random Variables

A *random variable* X is a mapping from a sample space into the real numbers, with the property that for every outcome there is an associated probability $\Pr[X \leq x]$ which exists for all real values of x. Random variables (rv's) will be denoted throughout this work by uppercase letters. Realized values of a rv will be denoted by the corresponding lowercase letter.

The *cumulative distribution function* (cdf) of X, often just called the *distribution function* (DF), is defined as $\Pr[X \leq x]$ and regarded as a function of x; it is customarily denoted by $F_X(x)$.

Clearly $F_X(x)$ is a nondecreasing function of x and $0 \leq F_X(x) \leq 1$. If $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to+\infty} F_X(x) = 1$, then the distribution is *proper*. We shall be concerned only with proper distributions.

The study of distributions is essentially a study of cdf's. In all cases in these volumes the cdf belongs to one of two classes, discrete or continuous, or it can be constructed by mixing elements from the two classes.

For *discrete distributions* $F_X(x)$ is a step function with only an enumerable number of steps. If the height of the step at x_j is p_j, then

$$\Pr[X = x_j] = p_j.$$

We call p_j a *probability mass function* (pmf), and we say that its *support* is the set $\{x_j\}$. If the distribution is proper, $\sum_j p_j = 1$. Random variables belonging to this class are called *discrete random variables*. Most of the discrete distributions of interest are defined either on the nonnegative unit lattice $x = 0, 1, \ldots$ or on $1, 2, \ldots$. A discrete distribution is said to be *logconvex* when $p_x p_{x+2}/p_{x+1}^2 > 1$. It is *logconcave* when $p_x p_{x+2}/p_{x+1}^2 < 1$.

For *continuous distributions* $F_X(x)$ is absolutely continuous and can be expressed as an integral,

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx. \tag{1.198}$$

Any function $f_X(x)$ for which (1.198) holds for every x is a probability density function (pdf) of X. Random variables in this class are called *continuous random variables*.

When the subscripts for $f_X(x)$ and $F_X(x)$ are well understood, they are often dropped, provided that this does not cause confusion.

The above concepts can be extended to the *joint distribution* of a finite number of rv's $X_1, X_2, \ldots, X_n$. The *joint cumulative distribution function* is

$$\begin{aligned}\Pr\left[\bigcap_{j=1}^{n}(X_j \leq x_j)\right] &= F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)\\ &= F(x_1, x_2, \ldots, x_n).\end{aligned} \tag{1.199}$$

If $\Pr\left[\bigcap_{j=1}^{n}(X_j = x_j)\right]$ is zero except for an enumerable number of sets of values $\{x_{1_i}, x_{2_i}, \ldots, x_{n_i}\}$ and

$$\sum_i \Pr\left[\bigcap_{j=1}^{n}(X_j = x_{j_i})\right] = 1,$$

then we have a *discrete joint distribution*. For such distributions

$$\sum_i \Pr\left[\bigcap_{j=1}^{n-1}(X_j = x_j) \cap (X_n = x_{n_i})\right] = \Pr\left[\bigcap_{j=1}^{n-1}(X_j = x_j)\right], \tag{1.200}$$

where the summation is over all values of x_{n_i} for which the probability is not zero.

If $F(x_1, x_2, \ldots, x_n)$ is absolutely continuous, then

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n}\int_{-\infty}^{x_{n-1}} \cdots \int_{-\infty}^{x_1} f(x_1, x_2, \ldots, x_n)\, dx_1 \cdots dx_n, \tag{1.201}$$

where $f(x_1, x_2, \ldots, x_n)$ [or strictly $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$] is the *joint probability density function* of $X_1, X_2, \ldots, X_n$. For a continuous joint distribution

$$\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_n = f(x_1, x_2, \ldots, x_{n-1}). \tag{1.202}$$

By repeated summation or integration, it is possible in principle to obtain the joint distribution of any subset of $X_1, X_2, \ldots, X_n$; in particular the distributions of each separate X_j can be found. These are called *marginal distributions*.

The *conditional joint distribution* of $X_1, X_2, \ldots, X_r$, given X_{r+1}, $X_{r+2}, \ldots, X_n$ [i.e., the joint distribution of the subset of the first r rv's in the case where particular values have been given to the remaining $(n-r)$ variables], is defined as

$$\Pr\left[\bigcap_{j=1}^{r}(X_j = x_j) \middle| \bigcap_{i=r+1}^{n}(X_i = x_i)\right] = \frac{\Pr\left[\bigcap_{j=1}^{n}(X_j = x_j)\right]}{\Pr\left[\bigcap_{j=r+1}^{n}(X_j = x_j)\right]} \tag{1.203}$$

(provided that $\Pr[\bigcap_{j=r+1}^{n}(X_j = x_j)] > 0$) for discrete distributions and by the pdf

$$f(x_1, x_2, \ldots, x_r | x_{r+1} \ldots, x_n) = \begin{cases} \dfrac{f(x_1, x_2, \ldots, x_n)}{f(x_{r+1} \ldots, x_n)}, & f(x_{r+1}, \ldots, x_n) > 0, \\ 0, & f(x_{r+1}, \ldots, x_n) = 0 \end{cases} \tag{1.204}$$

for continuous distributions. (The subscripts for F and f have been omitted for convenience in three of the above equations.)

Usually a distribution depends on one (or more) *parameters*, say θ. When we want to emphasize the dependence of the distribution on the value of θ, we write

$$F_X(x) = F_X(x;\theta) = F_X(x|\theta) \quad \text{and} \quad f_X(x) = f_X(x;\theta) = f_X(x|\theta). \quad (1.205)$$

1.2.4 Survival Concepts

Here we consider only lifetimes on the nonnegative integers with $\Pr[T = t] = p_t$, $t = 0, 1, \ldots$. We use T, not X, and t, not x, in this section to emphasize that time is assumed to be discrete.

Often $p_0 = 0$; the case $p_0 \neq 0$ corresponds to a nonzero probability of a death at birth (e.g., an egg that fails to hatch) or to a proportion p_0 of dud items.

Six representations characterize such distributions:

1. The pmf is

$$\Pr[T = t] = p_t, \qquad t = 0, 1, 2, \ldots, \quad (1.206)$$

2. The *survival function (survivor function)* is

$$S_0 = 1, \qquad S_t = 1 - \Pr(T < t) = \sum_{j \geq t} p_j, \qquad t = 1, 2, \ldots. \quad (1.207)$$

This is a nonincreasing step function that is left continuous since

$$\lim_{\epsilon \to 0} [S_{t-\epsilon} - S_t] = 0, \qquad \epsilon > 0, \qquad t \geq 0.$$

3. The *hazard function (failure rate, FR)* is

$$h_t = \frac{p_t}{\sum_{j \geq t} p_j} = \frac{S_t - S_{t+1}}{S_t}; \quad (1.208)$$

it is the probability that an item has survived to time t, given that it has survived to at least time t, that is, the amount of risk associated with an item at time t.

4. By analogy with the continuous case, the function

$$\Lambda_t = -\ln S_t \quad (1.209)$$

is called the *cumulative hazard function*.

5. In the discrete case summing the hazard function gives

$$H_t = \sum_{j=0}^{t} h_j. \quad (1.210)$$

This will be called the *accumulated hazard function*; it is a more tractable function for discrete data; see Kemp (2004). In general, $H_t \neq \Lambda_t$.

6. Following Kalbfleisch and Prentice (1980), Lawless (1982), and Leemis (1995), the *mean residual life function* is

$$L_t = E[(T - t)|T \geq t], \qquad t \geq 0. \tag{1.211}$$

Each of these six discrete lifetime functions can be stated uniquely in terms of each of the other functions. For example,

$$\begin{aligned} L_t &= \frac{\sum_{j \geq t} j p_j}{\sum_{j \geq t} p_j} - t = \frac{\sum_{j > t} S_j}{S_t} = \sum_{j \geq t} \prod_{x=t}^{j} (1 - h_x) \\ &= \sum_{j > t} e^{\Lambda_t - \Lambda_j} \quad = \sum_{j \geq t} \prod_{x=t}^{j} (1 - H_x + H_{x-1}) \end{aligned}$$

and

$$p_t = \left(1 - \frac{L_t}{1 + L_{t+1}}\right) \prod_{j=0}^{t-1} \left(\frac{L_j}{1 + L_{j+1}}\right), \qquad h_t = 1 - \frac{L_t}{1 + L_{t+1}},$$

$$S_t = \frac{S_{t-1} L_{t-1}}{1 + L_t} = \prod_{j=0}^{t-1} \left(\frac{L_j}{1 + L_{j+1}}\right), \qquad \Lambda_t = -\sum_{j=0}^{t-1} \ln \left(\frac{L_j}{1 + L_{j+1}}\right),$$

$$H_t = t + 1 - \sum_{j=0}^{t} \left(\frac{L_j}{1 + L_{j+1}}\right).$$

The above definitions lead by analogy with the continuous case to the following classes of discrete lifetime distributions:

IFR/DFR A discrete distribution with infinite support has a monotonically nondecreasing failure rate with time (IFR) or a monotonically nonincreasing failure rate with time (DFR) according as

$$\frac{p_{t+1}}{p_t} \gtrless \frac{p_{t+2}}{p_{t+1}} \tag{1.212}$$

(Gupta, Gupta, and Tripathi, 1997).

IFRA/DFRA A discrete lifetime distribution has an increasing or decreasing failure rate on average (IFRA or DFRA) according as

$$\frac{H_t}{t+1} \gtrless \frac{H_{t-1}}{t}, \qquad t \geq 1, \tag{1.213}$$

where H_t is the accumulated hazard function.

NBU/NWU A lifetime distribution is new better than used (NBU) or new worse than used (NWU) according as the conditional survival probability at time x for an item that has survived to time t is less (or greater) than the survival probability at time x for a new item, that is, according as

$$\frac{S_{t+x}}{S_t} \lesseqgtr S_x. \tag{1.214}$$

NBUE/NWUE A discrete lifetime distribution is new better than used in expectation (NBUE) or new worse than used in expectation (NWUE) according as

$$\frac{\sum_{j=0}^{\infty} S_{t+j}}{S_t} \lesseqgtr \sum_{j=0}^{\infty} S_j. \tag{1.215}$$

IMRL/DMRL An increasing mean residual life (IMRL) or a decreasing mean residual life (DMRL) is determined by

$$\begin{aligned} L_t - L_{t+1} &= \sum_{j>t} \left(\frac{S_j}{S_t} - \frac{S_{j+1}}{S_{t+1}} \right) = \sum_{j>t} \left[(h_j - h_t) \prod_{x=t+1}^{j-1} (1 - h_x) \right] \\ &\gtrless 0, \end{aligned} \tag{1.216}$$

that is, according as $h_j \gtrless h_t,\ j > t$.

The above definitions enable the interrelationships between these classes of discrete lifetime distributions to be stated as

$$\begin{aligned} \text{IFR/DFR} &\Rightarrow \text{IFRA/DFRA} \\ &\Rightarrow \text{NBU/NWU} \ \Rightarrow \text{NBUE/NWUE} \\ &\Rightarrow \text{DMRL/IMRL} \end{aligned}$$

See Kemp (2004) and the references therein for examples and proofs.

1.2.5 Expected Values

The *expected value* of a mathematical function $g(X_1, X_2, \ldots, X_n)$ of $X_1, X_2, \ldots, X_n$ is defined as

$$E[g(X_1, X_2, \ldots, X_n)] = \sum_i g(x_{1_i}, x_{2_i}, \ldots, x_{n_i}) \Pr\left[\bigcap_{j=1}^{n} (X_j = x_{j_i})\right] \tag{1.217}$$

for discrete distributions and as

$$E[g(X_1, X_2, \ldots, X_n)] \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n \quad (1.218)$$

for continuous distributions.

In particular, when $n = 1$,

$$E[g(X)] = \sum_{x} g(x) \Pr[X = x] \qquad \text{or} \qquad \int_{-\infty}^{\infty} g(x) f(x)\, dx. \quad (1.219)$$

If K is constant, then

$$E[K] = K,$$

$$E[Kg(X)] = KE[g(X)].$$

Also

$$E[g_1(X_1) + g_2(X_2)] = E[g_1(X_1)] + E[g_2(X_2)]. \quad (1.220)$$

More generally,

$$E\left[\sum_{j=1}^{M} K_j g_j(X_1, X_2, \ldots, X_n)\right] = \sum_{j=1}^{M} K_j E[g_j(X_1, X_2, \ldots, X_n)]. \quad (1.221)$$

These results apply to both discrete and continuous rv's. Conditional expected values are defined similarly, and formulas like (1.217) and (1.218) are valid for them.

The continuous rv's X_1, X_2 are said to be *independent*, if, for all real x_1, x_2, the events $(X_1 \le x_1)$, $(X_2 \le x_2)$ are independent.

The set $\{X_1, X_2, \ldots, X_n\}$ is a *mutually independent* set of discrete rv's if, for any combination of values $x_1, x_2, \ldots, x_n$ assumed by the rv's $X_1, X_2, \ldots, X_n$, $\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \ldots \Pr[X_n = x_n]$. In this case

$$E\left[\prod_{j=1}^{k} g_j(X_j)\right] = \prod_{j=1}^{k} E[g_j(X_j)]. \quad (1.222)$$

The *Shannon entropy* for a discrete rv is

$$H(X) = E\left[\log_2\left(\frac{1}{p_x}\right)\right] = -\sum_{\forall x} p_x \log_2(p_x). \quad (1.223)$$

For a continuous rv it is

$$H(X) = -\int_{\forall x} f(x) \log_2[f(x)]\, dx. \tag{1.224}$$

A stochastic process $\{X_n | n \geq 1\}$ with $E[|X_n|] < \infty$ for all n is a *martingale* process if

$$E[X_{n+1}|X_1, X_2, \ldots, X_n] = X_n. \tag{1.225}$$

It is a submartingale if $E[X_{n+1}|X_1, X_2, \ldots, X_n] \geq X_n$; it is a supermartingale if $E[X_{n+1}|X_1, X_2, \ldots, X_n] \leq X_n$.

1.2.6 Inequalities

1. *Cauchy–Schwartz Inequality* If X and Y are rv's such that $E[X^2]$ and $E[Y^2]$ exist, then

$$(E[XY])^2 \leq E[X^2]E[Y^2].$$

2. *Jensen's Inequality* If $E[X]$ exists and if $f(x)$ is a convex function, then

$$E[f(X)] \geq f(E[X]).$$

3. *Chebyshev Inequality* If $c > 0$ is a real number and if X is a rv such that $E[(X-c)^2]$ is finite, then

$$\Pr[|X - c|] \geq \epsilon] \leq \frac{1}{\epsilon^2} E[(X-c)^2]$$

for every $\epsilon > 0$.

4. *Bienaymé–Chebyshev Inequality* If $a > 0$ is a real number and if $E[|X|^r]$ is finite, then

$$\Pr[|X|] \geq a] \leq \frac{E[|X|^r]}{a^r}.$$

5. *Markov's Inequality* If X is a rv that takes only nonnegative values, then for all $a > 0$

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

Patel, Kapardia, and Owen (1976) have catalogued further inequalities, with references. Their Sections 2.1 and 2.2 give moment and Chebyshev-type inequalities; their Section 10.14 contains combinatorial inequalities.

1.2.7 Moments and Moment Generating Functions

Uncorrected Moments The expected value of X^r for r any real number is termed the rth *uncorrected (crude) moment* (alternatively the rth *moment about zero*):

$$\mu'_r(X) = \mu'_r = E[X^r]. \tag{1.226}$$

Unless otherwise stated, we will restrict consideration to integer values of r. The *(uncorrected) moment generating function* (mgf), if it exists (i.e., is finite), is

$$M_X(t) = E[e^{tX}] = 1 + \sum_{r \geq 1} \frac{\mu'_r t^r}{r!} \tag{1.227}$$

[when $M_X(t)$ exists for some interval $|t| < T$, where $T > 0$, then μ'_r is the coefficient of $t^r/r!$ in the Taylor expansion of $M_X(t)$]. If $\varphi(t)$ is the characteristic function of X (see Section 1.2.10), then $M(t) = \varphi(-it)$.

The first uncorrected moment μ'_1 is called the *mean* and is often written as μ.

The uncorrected moments can also be obtained from the cdf $F_X(x)$. If X is a continuous rv, then

$$E[X^r] = \int_0^\infty r x^{r-1}[1 - F_X(x) + (-1)^r F_X(-x)]\,dx. \tag{1.228}$$

If X is discrete, taking values $0, 1, \ldots, n$, where n is finite or infinite, then

$$E[X^r] = \sum_{x=0}^{n-1} [(x+1)^r - x^r][1 - F(x)]. \tag{1.229}$$

From the definition of the mgf,

$$M_{X+c}(t) = e^{ct} M_X(t). \tag{1.230}$$

Moreover, if X_1 and X_2 are independent rv's, then

$$M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t) \qquad \text{and} \qquad M_{X_1-X_2}(t) = M_{X_1}(t) M_{X_2}(-t). \tag{1.231}$$

If $X_1, X_2, \ldots, X_n$ are mutually independent rv's, then

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t). \tag{1.232}$$

Moments about the Mean The rth moment about a constant a is $E[(X-a)^r]$. When $a = \mu$, we have the rth *moment about the mean* (also called the rth *central* *moment* or the rth *corrected moment*),

$$\mu_r(X) = \mu_r = E[(X - \mu)^r] = E[(X - E[X])^r]. \qquad (1.233)$$

The *central moment generating function*, if it exists, is

$$E[e^{(X-\mu)t}] = e^{-\mu t} M_X(t) = 1 + \sum_{r \geq 1} \frac{\mu_r t^r}{r!}. \qquad (1.234)$$

The first central moment μ_1 is always zero. The second central moment μ_2 is called the *variance* of X [written as $\mathrm{Var}(X)$]. The positive square root of this quantity is called the *standard deviation* and is often denoted by the symbols $\sigma(X)$, or σ when no confusion is likely to arise. We have $\mu_2 \equiv \sigma^2$. The *coefficient of variation* (sometimes abbreviated to C. of V. or CV) is σ/μ.

Median The *median* of a distribution is the value of the variate that divides the total frequency into equal halves. For a continuous distribution this is unique. For a discrete distribution with $2N + 1$ elements, the median is the value of the $(N + 1)$th element; when there are $2N$ elements, there is ambiguity, and it is usual to define the median as the average of the Nth and $(N + 1)$th elements.

Mode If $f_X(x)$ is a continuous and twice-differentiable pdf, then x is a *mode* if $df_X(x)/dx = 0$ and $d^2 f_X(x)/dx^2 < 0$. A discrete distribution has a mode at $X = x$ if

$$\Pr[X = x - c_1 - 1] < \Pr[X = x - c_1] \leq \cdots \leq \Pr[X = x]$$

and

$$\Pr[X = x] \geq \cdots \geq \Pr[X = x + c_2] > \Pr[X = x + c_2 + 1],$$

where $0 \leq c_1$, $0 \leq c_2$.

A distribution with only one mode is said to be *unimodal*; otherwise it is *multimodal*. A distribution with support $x \geq 0$ and a peak in frequency at $X = 0$ is sometimes said to have a *half mode* at $X = 0$ and to be *sesquimodal*. Abouammoh and Mashour (1981) have given necessary and sufficient conditions for a discrete distribution to be unimodal. Olshen and Savage (1970) have introduced the concept of α-unimodality for continuous distributions. For *discrete α-unimodality* (for discrete distributions), see Abouammoh (1987) and Steutel (1988).

Shape Commonly used indices of the shape of a distribution are the *moment ratios*. The most important are

$$\alpha_3(X) = \sqrt{\beta_1(X)} = \mu_3(\mu_2)^{-3/2} \quad \text{(an index of } \textit{skewness}\text{)}, \qquad (1.235)$$

$$\alpha_4(X) = \beta_2(X) = \mu_4(\mu_2)^{-2} \quad \text{(an index of } \textit{kurtosis}\text{)}, \qquad (1.236)$$

and more generally

$$\alpha_r(X) = \mu_r(\mu_2)^{-r/2}. \tag{1.237}$$

The α and β notations are both in use. Note that these moment ratios have the same value for any linear function $A + BX$ with $B > 0$. When $B < 0$, the absolute values are not altered, but ratios of odd order have their signs reversed.

Moments about the Mean from Uncorrected Moments It is often convenient to calculate the central moments μ_r from the uncorrected moments and, less often, vice versa. Formulas for this involve the binomial coefficients:

$$\mu_r = E[(X - E[X])^r] = \sum_{j=0}^{r}(-1)^j \binom{r}{j} \mu'_{r-j}\mu^j; \tag{1.238}$$

we refer the reader to Stuart and Ord (1987, p. 73) for further relevant formulas. In particular

$$\begin{aligned}
\mu_2 &= \mu'_2 - \mu^2, \\
\mu_3 &= \mu'_3 - 3\mu'_2\mu + 2\mu^3, \\
\mu_4 &= \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4.
\end{aligned} \tag{1.239}$$

For the inverse calculation

$$\begin{aligned}
\mu'_2 &= \mu_2 + \mu^2, \\
\mu'_3 &= \mu_3 + 3\mu_2\mu + \mu^3, \\
\mu'_4 &= \mu_4 + 4\mu_3\mu + 6\mu_2\mu^2 + \mu^4.
\end{aligned} \tag{1.240}$$

The characterization of a distribution via its moment properties has been studied by several authors; see Johnson and Kotz (1990a,b) for a discussion of the methods that have been used and for new results.

Absolute Moments Besides the uncorrected and the central moments there are *absolute moments*, defined as the expected values of the absolute values (moduli) of various functions of X. Thus the rth *absolute moment about zero* of X is

$$\nu'_r(X) = E[|X|^r], \tag{1.241}$$

while the rth *absolute central moment* is

$$\nu_r(X) = E[|X - E[X]|^r]. \tag{1.242}$$

If r is even, $\nu'_r = \mu'_r$ and $\nu_r = \mu_r$, but not if r is odd. Whereas $\mu_1 = 0$, in general $\nu_1 > 0$. We call ν_1 the *mean deviation* of X.

Factorial Moments When studying discrete distributions, it is often advantageous to use the *factorial moments*. Those most commonly used are the descending factorial moments. The rth *descending factorial moment* of X is the expected value of $X!/(X-r)!$:

$$\mu'_{[r]} = E\left[\frac{X!}{(X-r)!}\right]. \tag{1.243}$$

Readers are WARNED that there are other notations in use for the descending factorial moments. In the first edition of this book $\mu_{(r)}$ was used. Patel, Kapardia, and Owen (1976) use $\mu'_{(r)}$. The rth *ascending factorial moment* of X is $E[(X + r - 1)!/(X - 1)!]$.

Since $X!/(X-r)! = \sum_{j=0}^{r} s(r, j)X^j$, where $s(r, j)$ is the Stirling number of the first kind (see Section 1.1.3), we find that

$$\mu'_{[r]} = \sum_{j=0}^{r} s(r, j)\mu'_j. \tag{1.244}$$

Thus

$$\begin{aligned}
\mu'_{[1]} &= \mu, \\
\mu'_{[2]} &= \mu'_2 - \mu, \\
\mu'_{[3]} &= \mu'_3 - 3\mu'_2 + 2\mu, \\
\mu'_{[4]} &= \mu'_4 - 6\mu'_3 + 11\mu'_2 - 6\mu, \\
\mu'_{[5]} &= \mu'_5 - 10\mu'_4 + 35\mu'_3 - 50\mu'_2 + 24\mu, \\
\mu'_{[6]} &= \mu'_6 - 15\mu'_5 + 85\mu'_4 - 225\mu'_3 + 274\mu'_2 - 120\mu.
\end{aligned} \tag{1.245}$$

Similarly

$$X^r = \sum_{j=0}^{r} \frac{S(r, j)X!}{(X-r)!},$$

where $S(r, j)$ are the Stirling numbers of the second kind (see Section 1.1.3), and so

$$\mu'_r = \sum_{j=0}^{r} S(r, j)\mu'_{[j]}. \tag{1.246}$$

Hence

$$\begin{aligned}
\mu &= \mu'_{[1]}, \\
\mu'_2 &= \mu'_{[2]} + \mu, \\
\mu'_3 &= \mu'_{[3]} + 3\mu'_{[2]} + \mu, \\
\mu'_4 &= \mu'_{[4]} + 6\mu'_{[3]} + 7\mu'_{[2]} + \mu, \\
\mu'_5 &= \mu'_{[5]} + 10\mu'_{[4]} + 25\mu'_{[3]} + 15\mu'_{[2]} + \mu, \\
\mu'_6 &= \mu'_{[6]} + 15\mu'_{[5]} + 65\mu'_{[4]} + 90\mu'_{[3]} + 31\mu'_{[2]} + \mu.
\end{aligned} \tag{1.247}$$

The (descending) *factorial moment generating function* (fmgf), if it exists, is

$$E[(1+t)^X] = 1 + \sum_{r\geq 1} \frac{\mu'_{[r]} t^r}{r!}. \tag{1.248}$$

The relationship between the fmgf and the probability generating function enables the probabilities of a discrete distribution to be expressed in terms of its factorial moments; see Section 1.2.11.

Finite difference methods for obtaining the moments of a discrete distribution were discussed in a series of papers and letters in *The American Statistician* in the early 1980s; see, in particular, Johnson and Kotz (1981), Chan (1982), and Khatri (1983).

1.2.8 Cumulants and Cumulant Generating Functions

Cumulants The logarithm of the uncorrected moment generating function of X is the *cumulant generating function* (cgf) of X. If the mgf exists, then so does the cgf. The coefficient of $t^r/r!$ in the Taylor expansion of the cgf is the rth *cumulant* of X and is denoted by the symbol $\kappa_r(X)$ or, when no confusion is likely to arise, by κ_r:

$$K_X(t) = \ln M_X(t) = \sum_{r\geq 1} \frac{\kappa_r t^r}{r!} \tag{1.249}$$

(note that there is no term in t^0 in this equation).

We have

$$K_{X+a}(t) = at + K_X(t). \tag{1.250}$$

Hence for $r \geq 2$ the coefficients of $t^r/r!$ in $K_{X+a}(t)$ and $K_X(t)$ are the same; that is, the cumulants for $r \geq 2$ are not affected by the addition of a constant to X. For this reason the cumulants have also been called *seminvariants* or *half invariants*. Putting $a = -\mu$ shows that, for $r \geq 2$, the cumulants κ_r are functions of the central moments. In fact,

$$\begin{aligned} &\kappa_1 = \mu, \qquad \kappa_2 = \mu_2, \qquad \kappa_3 = \mu_3, \\ &\kappa_4 = \mu_4 - 3\mu_2^2, \qquad \kappa_5 = \mu_5 - 10\mu_3\mu_2. \end{aligned} \tag{1.251}$$

Note that the first three moments are equal to the first three cumulants.

Smith (1995) gave the following formulas connecting the uncorrected moments μ'_r to the cumulants for discrete distributions:

$$\mu'_r = \sum_{i=0}^{r-1} \binom{r-1}{i} \kappa_{r-i} \mu'_i \tag{1.252}$$

$$\kappa_r = \mu'_r - \sum_{i=1}^{r-1} \binom{r-1}{i} \kappa_{r-i} \mu'_i. \tag{1.253}$$

He gave analogous formulas for multivariate discrete distributions; see also Balakrishnan, Johnson, and Kotz (1998).

Let $X_1, X_2, \ldots, X_n$ be independent rv's and let $X = \sum_1^n X_j$; then, if the relevant functions exist,

$$K_X(t) = \sum_{j=1}^{n} K_{X_j}(t). \tag{1.254}$$

It follows from this equation that

$$\kappa_r \left(\sum_{j=1}^{n} X_j \right) = \sum_{j=1}^{n} \kappa_r(X_j) \quad \text{for all } r; \tag{1.255}$$

that is, the cumulant of a sum equals the sum of the cumulants, which makes the name "cumulant" appropriate.

Factorial Cumulants The logarithm of the (descending) fmgf is called the *factorial cumulant generating function* (fcgf). The coefficient of $t^r/r!$ in the Taylor expansion of this function is the rth *factorial cumulant* $\kappa_{[r]}$:

$$\ln G(1+t) = \sum_{r \geq 1} \frac{\kappa_{[r]} t^r}{r!}. \tag{1.256}$$

Formulas connecting $\{\kappa_r\}$ and $\{\kappa_{[r]}\}$ parallel those connecting $\{\mu'_r\}$ and $\{\mu'_{[r]}\}$:

$$\begin{aligned}
\kappa_1 &= \kappa_{[1]} = \mu, \\
\kappa_2 &= \kappa_{[2]} + \mu, \\
\kappa_3 &= \kappa_{[3]} + 3\kappa_{[2]} + \mu, \\
\kappa_4 &= \kappa_{[4]} + 6\kappa_{[3]} + 7\kappa_{[2]} + \mu, \\
&\vdots
\end{aligned} \tag{1.257}$$

Douglas (1980) has given a very full account of the relationships between the various types of moments and cumulants; see also Stuart and Ord (1987).

A sampling distribution arises as the distribution of some function of observations taken over all possible samples from a particular distribution according to a specified sampling scheme. The moment properties of a sampling distribution can be expressed in terms of symmetric functions of the observations, known as k-statistics; these were introduced by Fisher (1929). The expected value of the univariate k-statistic of order r is the rth cumulant; see Stuart and Ord (1987, Chapter 12).

1.2.9 Joint Moments and Cumulants

Moments of joint distributions, that is, quantities like $E\left[\prod_{j=1}^{n} X_j^{a_j}\right]$, are called *product moments (about zero)* and are denoted by $\mu'_{a_1 a_2 \ldots a_n}$. Quantities like

$$E\left[\prod_{j=1}^{n}(X_j - E[X_j])^{a_j}\right] = \mu_{a_1 a_2 \ldots a_n} \tag{1.258}$$

are called *central product moments* (sometimes *central mixed moments*).

The central product moment

$$\mu_{11} = E[(X_j - E[X_j])(X_{j'} - E[X_{j'}])] \tag{1.259}$$

is called the *covariance* of X_j and $X_{j'}$ and is denoted by $\mathrm{Cov}(X_j, X_{j'})$. The *correlation* between X_j and $X_{j'}$ is defined as

$$\rho(X_j X_{j'}) = \rho_{jj'} = \frac{\mathrm{Cov}(X_j, X_{j'})}{[\mathrm{Var}(X_j)\mathrm{Var}(X_{j'})]^{1/2}}. \tag{1.260}$$

[This is also sometimes written as $\mathrm{Corr}(X_j X_{j'})$.] It can be shown that $-1 \leq \rho_{jj'} \leq 1$. If X_j and $X_{j'}$ are mutually independent, then $\mathrm{Cov}(X_j X_{j'}) = 0 = \rho_{jj'}$; the converse is not necessarily true.

The *joint moment generating function* of $X_1, X_2, \ldots, X_n$ is defined as a function of n generating variables $t_1, t_2, \ldots, t_n$:

$$M_{X_1,\ldots,X_n}(t_1, t_2, \ldots, t_n) = M(t_1, t_2, \ldots, t_n) = E\left[\exp\sum_{j=1}^{n} t_j X_j\right]. \tag{1.261}$$

The *joint central moment generating function* is

$$E\left[\exp\left(\sum_{j=1}^{n} t_j(X_j - E[X_j])\right)\right] = \exp\left[-\sum_{j=1}^{n} t_j E[X_j]\right] M(t_1, t_2, \ldots, t_n). \tag{1.262}$$

The *joint cumulant generating function* is $\ln M_{X_1,\ldots,X_n}(t_1, t_2, \ldots, t_n)$. Use of these generating functions is similar to that for the single-variable functions.

The *regression function* of a rv X on m other random variables $X_1, X_2, \ldots, X_m$ is defined as

$$E[X|X_1, X_2, \ldots, X_m]; \tag{1.263}$$

it is an important tool for the prediction of X from $X_1, X_2, \ldots, X_m$. If (1.263) is a linear function of $X_1, X_2, \ldots, X_m$, then the regression is called *linear* (or *multiple linear*). The variance of the conditional distribution of X given $X_1, X_2, \ldots, X_m$ is called the *scedasticity*. If $\mathrm{Var}(X|X_1, X_2, \ldots, X_m)$ does not depend on $X_1, X_2, \ldots, X_m$, then the conditional distribution is said to be *homoscedastic*.

Given that X_1 and X_2 are random variables, their joint distribution is determined by the distribution of X_1 together with the conditional distribution of X_2 given X_1. There has been much research during the past three decades on characterizations based on regression properties; see, for instance, Korwar (1975) and Papageorgiou (1985). Kotz and Johnson (1990) have provided a good review concerning characterizations for discrete distributions; see also Prasaka Rao (1992).

1.2.10 Characteristic Functions

The *characteristic function* (cf) of a continuous distribution is defined as

$$\varphi(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx},\ dF(x), \tag{1.264}$$

where $i = \sqrt{-1}$ and t is real. It is a complex-valued function. For a discrete distribution on the nonnegative integers, it is defined as

$$\varphi(t) = E[e^{itX}] = \sum_{j=0}^{\infty} e^{ijt}\,\Pr[X = j]. \tag{1.265}$$

The cf has great theoretical importance, particularly for continuous distributions. It is uniquely determined by the cdf and exists for all distributions. It satisfies (1) $\varphi(0) = 1$, (2) $|\varphi(t)| \le 1$, and (3) $\varphi(-t) = \overline{\varphi(t)}$, where the overline denotes the complex conjugate.

If the distribution with cdf $F(x)$ has finite moments μ'_r up to order n, then

$$\mu'_r = i^r \varphi^{(r)}(0), \qquad 1 \le r \le n, \tag{1.266}$$

where $\varphi^{(r)}(0)$ is the rth derivative of $\varphi(t)$ evaluated at $t = 0$ and $i^2 = -1$.

The cf uniquely determines the pdf of a continuous distribution; we have

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt. \tag{1.267}$$

Gauss (1900) called this *Ein Schönes Theorem der Wahrscheinlichkeitsrechnung*.

The corresponding inversion formula for discrete distributions on the nonnegative integers is

$$\Pr[X = x] = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx}\varphi(t)\,dt. \tag{1.268}$$

Lukacs (1970) gave further inversion formulas for continuous and for discrete distributions.

If X_1 and X_2 are independent rv's with cf's $\varphi_1(t)$ and $\varphi_2(t)$, respectively, then the cf of their sum $X_1 + X_2$ is the product of their cf's $\varphi_1(t)\varphi_2(t)$. Moreover the cf of their difference $X_1 - X_2$ is $\varphi_1(t)\varphi_2(-t)$.

Under very general conditions the cf for a limiting distribution is the limiting cf. If $\lim_{j\to\infty} \varphi_{X_j}(t) = \varphi_X(t)$, where $\varphi_X(t)$ is the cf of a rv with cdf F_X, then $\lim_{j\to\infty} F_{X_j}(x) = F_X(x)$.

A cf $\varphi(t)$ is said to be *infinitely divisible* if

$$\varphi(t) = [\varphi_n(t)]^n$$

for all positive integer n, where $\varphi_n(t)$ is itself a cf.

A cf $\varphi(t)$ is said to be *decomposable* if there are two nondegenerate cf's $\varphi_1(t)$ and $\varphi_2(t)$ such that

$$\varphi(t) = \varphi_1(t)\varphi_2(t).$$

A cf is said to be *stable* if $\varphi(a_1 t)e^{itb_1}\varphi(a_2 t)e^{itb_2} = \varphi(a_3 t)e^{itb_3}$, where $a_i > 0$ for $i = 1, 2, 3$.

Seminal references concerning cf's are Lukacs (1970, 1983) and Laha (1982).

1.2.11 Probability Generating Functions

Consider a nonnegative discrete rv X with nonzero probabilities only at nonnegative integer values. Let

$$p_j = \Pr[X = j], \qquad j = 0, 1, \ldots. \tag{1.269}$$

If the distribution is proper, then $\sum_{j=0} p_j = 1$, and hence $\sum_{j=0} p_j z^j$ converges for $|z| \le 1$. (This is also true when the distribution is not proper since then $0 < \sum_{j=0} p_j < 1$. However, we will be concerned only with proper distributions.)

The *probability generating function* (pgf) of the distribution with probability mass function (1.269) (or equivalently of the rv X) is defined as

$$G(z) = \sum_{j=0}^{\infty} p_j z^j = E[z^X]. \tag{1.270}$$

Although it would be logical to use the notation $G_X(z)$ for the pgf of X, we will in general suppress the suffix when it is clearly understood.

Probability generating functions have many properties:

1. The pgf is closely related to the cf; we have

$$\varphi_X(t) = E[e^{itX}] = G(e^{it}). \tag{1.271}$$

2. The pgf is defined by the probabilities; the uniqueness of a power series expansion implies that the pgf in turn defines the probabilities. We find that

$$p_j = \left[\frac{1}{j!} \frac{d^j G(z)}{dz^j} \right]_{z=0}, \qquad j = 0, 1, \ldots. \tag{1.272}$$

3. The rth moment, if it exists, is

$$\mu'_r = \sum_{j=0}^{\infty} j^r p_j = \left[\frac{d^r G(e^t)}{dt^r} \right]_{t=0}, \qquad r = 1, 2, \ldots. \tag{1.273}$$

4. The factorial moment generating function (fmgf) (if it exists) is given by

$$E[(1+t)^X] = G(1+t) = 1 + \sum_{r \geq 1} \frac{\mu'_{[r]} t^r}{r!}. \tag{1.274}$$

When the pgf is known, therefore, successive differentiation of the pgf enables the (descending) factorial moments to be obtained in a straightforward manner:

$$\mu'_{[r]} = \sum_{j=r}^{\infty} \frac{j!}{(j-r)!} p_j = \left[\frac{d^r G(z)}{dz^r} \right]_{z=1} = \left[\frac{d^r G(1+t)}{dt^r} \right]_{t=0}. \tag{1.275}$$

(The moments can be derived from the factorial moments as

$$\begin{aligned} \mu &= \mu'_{[1]}, \\ \mu_2 &= \mu'_{[2]} + \mu - \mu^2, \\ &\vdots \end{aligned}$$

see Section 1.2.7.)

5. The following relationships hold between the probabilities and the factorial moments in the case of a discrete distribution:

$$\Pr[X = x] = \sum_{j \geq x} (-1)^{x+j} \binom{j}{x} \frac{\mu'_{[j]}}{j!} = \sum_{r \geq 0} (-1)^r \frac{\mu'_{[x+r]}}{x! r!} \tag{1.276}$$

(Fréchet, 1940, 1943) and

$$\sum_{i \geq x} \Pr[X = i] = \sum_{j \geq x} (-1)^{x+j} \binom{j-1}{x-1} \frac{\mu'_{[j]}}{j!} \tag{1.277}$$

(Laurent, 1965).

6. If X_1 and X_2 are two independent rv's with pgf's $G_1(z)$ and $G_2(z)$, then the distribution of their sum $X = X_1 + X_2$ has the pgf

$$G(z) = E[z^X] = E[z^{X_1}]E[z^{X_2}] = G_1(z)G_2(z). \tag{1.278}$$

This is called the *convolution* of the two distributions. Let A, B, and C be the names of the distributions of X_1, X_2, and X; then we write $C \sim A^*B$.

7. More generally, let $X_1, X_2, \ldots, X_n$ be mutually independent rv's with pgf's $G_1(z), G_2(z), \ldots, G_n(z)$, respectively. Then the pgf of $X = \sum_{i=1}^{n} X_j$ is

$$G_X(z) = \prod_{j=1}^{n} G_j(z). \tag{1.279}$$

8. The pgf for the difference of two independent discrete rv's with pgf's $G_1(z)$ and $G_2(z)$ is

$$G_{X_i - X_j} = G_i(z)G_j(z^{-1}), \tag{1.280}$$

where the definition of a pgf is extended to encompass negative values of the variable.

9. The *joint probability generating function* of n discrete variables X_1, X_2, $\ldots, X_n$ is

$$G(z_1, z_2, \ldots, z_n) = E\left[\prod_{j=1}^{n} z_j^{X_j}\right], \tag{1.281}$$

where

$$\Pr\left[\bigcap_{j=1}^{n}(X_j = a_j)\right] = P_{a_1,a_2,\ldots,a_n}, \qquad a_j = 0, 1, 2, \ldots.$$

Let $r = \sum_{j=1}^{n} r_j$. Then the factorial moments of the distribution are given by

$$\mu'_{[r_1,r_2,\ldots,r_n]} = \left[\frac{\partial^r G(z_1, z_2, \ldots, z_n)}{\partial z_1^{r_1}\, \partial z_2^{r_2} \cdots \partial z_n^{r_n}}\right]_{z_1=z_2=\cdots=z_n=1}. \tag{1.282}$$

10. Relationships between probability generating functions and other generating functions are as follows:

Table 1.1 Relationships between Generating Functions

Probability generating function	$G(z)$
Characteristic function	$\varphi(t) = G(e^{it})$
Moment generating function	$M(t) = G(e^t)$
Central moment generating function	$e^{-\mu t}M(t) = e^{-\mu t}G(e^t)$
Factorial moment generating function	$G(1+t)$
Cumulant generating function	$K(t) = \ln G(e^t)$
Factorial cumulant generating function	$\ln G(1+t)$

An historical account of the use of pgf's in discrete distribution theory has been given by Seal (1949b).

1.2.12 Order Statistics

Let $X_1, X_2, \ldots, X_n$ be independent continuous rv's; then the jth *order statistic* $X_{j:n} \equiv X_{(j)}$, $j = 1, 2, \ldots, n$, is defined to be equal to the jth smallest of these. Here, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are called the *order statistics* corresponding to $X_1, X_2, \ldots, X_n$. Evidently

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

In particular,

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\} \qquad \text{and} \qquad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}.$$

For a continuous distribution the probability of a *tie* (two equal values) is zero, and therefore the definition is unambiguous. Ties may, however, occur given a discrete distribution; unambiguity can be achieved here also, provided that we interpret "jth smallest value" to mean "not more than $j-1$ smaller values *and* not more than $n-j$ larger values."

The difference, $w = X_{(n)} - X_{(1)}$, is called the *range*. If n is odd, the "middle" value $X_{(n+1)/2}$ is called the *median*; see Section 1.2.7. When n is even, the median is not uniquely defined; often the jth order statistic observations are grouped, and reference is made to the *median class*. The closely related concepts of *hinges* and *fences* play a central role in exploratory data analysis (EDA); we refer the reader to Tukey (1977) and Emerson and Hoaglin (1983).

In the continuous case the pdf for the jth order statistic is

$$\begin{aligned} f_{X_{(j)}}(x) &= n\binom{n-1}{j-1}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x) \\ &= \frac{n!}{(j-1)!(n-j)!}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x); \end{aligned} \tag{1.283}$$

the cdf for the jth order statistic is

$$F_{X_{(j)}}(x) = \Pr[X_{(j)} \leq x] = \Pr[j \text{ or more observations } \leq x]$$
$$= \sum_{k=j}^{n} \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}. \tag{1.284}$$

For the first order statistic, the pdf and the cdf are, respectively,

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x) \qquad \text{and} \qquad F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n.$$

For the nth order statistic they are

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x) \qquad \text{and} \qquad F_{X_{(n)}}(x) = [F(x)]^n.$$

In the discrete case, equation (1.284) holds and the pmf is

$$f_{X_{(j)}}(x) = F_{X_{(j)}}(x) - F_{X_{(j)}}(x - 1).$$

There has been much work on characterizations of distributions using properties of their order statistics; see, for instance, Arnold and Meeden (1975), Shah and Kabe (1981), Hwang and Lin (1984), Khan and Ali (1987), and Lin (1987). David's (1981) book provided, at the time, an encyclopedic coverage of properties, statistical techniques, characterizations, and applications relating to order statistics from both continuous and discrete distributions. Harter's (1988) paper gave definitions and examined the history and the importance of order statistics. Balakrishnan (1986) extended previous work on recurrence relations for single and product moments of order statistics in both the continuous and the discrete case.

Arnold, Balakrishnan, and Nagaraja (1992) have written a good introduction to the subject, including a chapter on discrete order statistics. Nagaraja's (1990) book also gives a good account of work on order statistics for discrete distributions, particularly concerning characterizations of the geometric distribution, and extreme order statistics. Estimation methods based on order statistics are discussed by Balakrishnan and Cohen (1991).

Nagaraja's (1992) survey of results on order statistics for random samples from discrete distributions includes discussion and rejoinder. It reviews finite sample theory, characterization results, and asymptotic results and discusses applications to testing and selection. The emphasis is on order statistics from the discrete uniform, geometric, binomial, and multinomial distributions.

1.2.13 Truncation and Censoring

If values of the rv's $X_1, X_2, \ldots, X_n$ in a given region $\overline{R}$ are excluded, then the joint cdf of the variables is

$$F(x_1, x_2, \ldots, x_n | R) = \Pr\left[\bigcap_{j=1}^{n} (X_j \leq x_j) | (X_1, \ldots, X_n) \subset R\right]$$
$$= \frac{\Pr\left[\bigcap_{j=1}^{n} (X_j \leq x_j) \cap (X_1, \ldots, X_n) \subset R\right]}{\Pr[(X_1, X_2, \ldots, X_n) \subset R]}, \qquad (1.285)$$

where R is the complement of $\overline{R}$ and comprises all the points that are not truncated. The distribution given by (1.285) is called a *truncated distribution.* All the quantities on the right-hand side of the equation can be calculated from the (unconditional) joint cdf $F(x_1, x_2, \ldots, x_n)$.

We shall usually be concerned with truncated distributions of single variables, for which R is a finite or infinite interval. If R is a finite interval with end points a and b inside the range of values taken by X, the distribution is *doubly truncated* (or *left-and-right truncated*, or *truncated below and above*); a and b are the *truncation points*.

If R consists of all values greater than a, then the distribution is said to be *truncated from below*, or *left truncated*; if R consists of all values less than b, then the distribution is said to be *truncated from above*, or *right truncated.* (The same terms are also used when R includes values equal to a or to b, as the case may be.)

If X' is a rv having a distribution formed by doubly truncating the distribution of a continuous random variable X, then the pdf of X', in terms of the pdf and cdf of X, is

$$f_{X'}(x') = \frac{f_X(x')}{F_X(b) - F_X(a)}, \qquad a \leq x' \leq b. \qquad (1.286)$$

A distinction needs to be made between truncation of a distribution and truncation of a sample. Truncation of a distribution occurs when a range of possible variate values either is ignored or is impossible to observe.

Truncation of a sample is commonly called *censoring*. Sometimes censoring is with respect to a fixed variate value; for instance, in a survival study it may be impossible within a limited time span to ascertain the length of survival of all the patients. When the existence of observations outside a certain range is known but their exact value is unknown, the form of censoring is known as *type I censoring*.

When a predetermined number of order statistics are omitted from a sample, the form of censoring is known as *type II*. If the ℓ smallest values $X_{(1)}, \ldots, X_{(\ell)}$ are omitted, it is *censoring from below*, or *left censoring*; when the m largest values are omitted, it is *censoring from above*, or *right censoring*. If both sets of order statistics are omitted, we have *double censoring*.

The term “truncation” is used in a different sense in sequential analysis, where it refers to the imposition of a cutoff point leading to cessation of the sequential sampling process before a decision has been reached.

1.2.14 Mixture Distributions

A *mixture of distributions* is a superimposition of distributions with different functional forms or different parameters, in specified proportions.

Suppose that

$$\{F_j(x_1, x_2, \ldots, x_n)\}, \qquad j = 0, 1, 2, \ldots, m,$$

represents a set of different (proper) cdf's, where m is finite or infinite. Suppose also that $a_j \geq 0$, $\sum_{j=0}^{m} a_j = 1$. Then

$$F(x_1, x_2, \ldots, x_n) = \sum_{j=0}^{m} a_j F_j(x_1, x_2, \ldots, x_n) \tag{1.287}$$

is a proper cdf. This mixture of the distributions $\{F_j\}$ is *finite* or *infinite* according as m is finite or infinite.

For many of the mixture distributions in this book, the distributions to be mixed all have cdf's of the same functional form but are dependent on some parameter Θ. If Θ itself has a discrete distribution with pmf $\Pr[\Theta = \theta_j] = p_j$, $j = 0, 1, \ldots,$ then the resultant mixture has the cdf

$$\sum_{j \geq 0} p_j F(x_1, x_2, \ldots, x_n | \theta_j).$$

If Θ has a continuous distribution with cdf $H_\Theta(\theta)$, then the resultant mixture has cdf

$$\int F(x_1, \ldots, x_n | \theta)\, dH_\Theta(\theta),$$

where integration is over all values of θ. In either case (discrete or continuous distribution of Θ) we have

$$F(x_1, x_2, \ldots, x_n) = E_\Theta[F_j(x_1, x_2, \ldots, x_n | \theta)], \tag{1.288}$$

where the expectation is with respect to Θ. We call $F(x_1, x_2, \ldots, x_n)$ the *mixture distribution* and say that the distribution of Θ is the *mixing distribution*.

More generally the distribution of $X_1, X_2, \ldots, X_n$ may depend on several parameters $\Theta_1, \ldots, \Theta_k, \Theta_{k+1}, \ldots, \Theta_m$, where $\Theta_1, \ldots, \Theta_k$ vary and $\Theta_{k+1}, \ldots,$ Θ_m are constant. The mixture distribution then has the cdf

$$F(x_1, x_2, \ldots, x_n | \theta_{k+1}, \ldots, \theta_m) = E_{\Theta_1, \ldots, \Theta_k}[F(x_1, x_2, \ldots, x_n | \theta_1, \ldots, \theta_m)]. \tag{1.289}$$

Note that the parameters $\theta_1, \ldots, \theta_k$ do not appear in the mixture distribution because they have been summed out (for a discrete mixture) or integrated out (for a continuous distribution). The parameters $\Theta_{k+1}, \ldots, \Theta_m$ have not been eliminated in this way.

Mixtures of discrete distributions are dealt with in depth in Chapter 8.

1.2.15 Variance of a Function

Given the moments of a rv X, suppose that we wish to obtain the moments of a mathematical function of X, that is, of $Y = h(X)$.

If exact expressions can be obtained and are convenient to use, this should of course be done. However, in some cases it may be necessary to use approximate methods. One approximate method is to expand $h(X)$ as a Taylor series about $E[X]$:

$$Y = h(E[X]) + (X - E[X])h'(E[X]) + (X - E[X])^2 \frac{h''(E[X])}{2!} + \cdots. \tag{1.290}$$

Then, taking expected values of both sides of (1.290),

$$E[Y] = h(E[X]) + \mathrm{Var}(X)\frac{h''(E[X])}{2} + R \tag{1.291}$$

$$\approx h(E|X]) + \mathrm{Var}(X)\frac{h''(E[X])}{2}. \tag{1.292}$$

Also

$$\{Y - h(E[X])\}^2 \approx \{(X - E[X])h'(E[X])\}^2,$$

whence

$$\mathrm{Var}(Y) \approx \{h'(E[X])\}^2 \mathrm{Var}(X). \tag{1.293}$$

This method of approximation has been used widely under a number of different names, for example, the *delta method*, the *method of statistical differentials*, and the *propagation of error*. The method assumes that the expected value of the remainder term R in (1.291) is small and that the higher order central moments do not become large; otherwise the outcome may be very unreliable. The method will usually be more reliable for small values of $\mathrm{Var}(X)$.

Equation (1.293) can be made the basis for an approximate *variance-stabilizing transformation*. If $\mathrm{Var}(X)$ is a function $g(E[X])$ of $E[X]$, then $\mathrm{Var}(Y)$ might be expected to be more nearly constant if $[h'(E[X])]^2 g(E[X])$ is a constant.

This will be so if

$$h(X) \propto \int^X \frac{dt}{[g(t)]^{1/2}}. \tag{1.294}$$

This suggests the use of $Y = h(X)$, with $h(X)$ satisfying (1.294), as a variance-stabilizing transformation. Often such transformations are also effective as *normalizing transformations* in that the distribution of Y is nearer to normality than that of X; see Johnson et al. (1994, Chapter 12).

When $Y = X^a$, the method gives

$$\begin{aligned} E[X^a] &\approx \mu^a\left[1 + \frac{a(a-1)\sigma^2}{2\mu^2}\right], \\ \mathrm{Var}(X^a) &\approx \mu^{2a-2}a^2\sigma^2, \end{aligned} \tag{1.295}$$

where $\mu = E[X]$ and $\sigma^2 = \mathrm{Var}(X)$. The coefficient of variation of X^a is therefore very approximately $|a|(\sigma/\mu)$.

There are exact methods for obtaining the moments of a product of two rv's. For the quotient of two nonnegative rv's the delta method gives

$$\begin{aligned} E\left[\frac{X_1}{X_2}\right] &\approx \frac{\xi_1}{\xi_2}\left[1 + \frac{\mathrm{Var}(X_2)}{\xi_2^2} - \frac{\mathrm{Cov}(X_1, X_2)}{\xi_1\xi_2}\right], \\ \mathrm{Var}\left(\frac{X_1}{X_2}\right) &\approx \frac{\xi_1^2}{\xi_2^2}\left[\frac{\mathrm{Var}(X_1)}{\xi_1^2} - \frac{2\,\mathrm{Cov}(X_1, X_2)}{\xi_1\xi_2} + \frac{\mathrm{Var}(X_2)}{\xi_2^2}\right], \end{aligned} \tag{1.296}$$

where ξ_1 and ξ_2 are the expected values of X_1 and X_2, respectively.

For further discussion and use of the delta method see Stuart and Ord (1987, Chapter 10).

1.2.16 Estimation

Since the publication of the first edition of this book an immense amount of research has been devoted to statistical estimation, both to theoretical developments and to practical aspects of inferential procedures. Computer-intensive methods are now widely used. The books by Cox and Hinkley (1974) and Barnett (1999) are valuable for their lucid discussions of the many approaches to inference. Desmond and Godambe (1998) and Doksum (1998) give good introductory accounts with references.

Readers of this book will occasionally meet references to results from goodness of fit, hypothesis testing, and decision theory. Kocherlakota and Kocherlakota (1986) concentrated on goodness of fit tests for discrete distributions. The many well-written books on these topics also include D'Agostino and Stephens (1986), Rayner and Best (1989) (*goodness of fit*), DeGroot (1970), Cox and Hinkley (1974), Lehmann (1986) (*hypothesis testing*), and Berger (1985) (*Bayesian decision theory*).

Bayesian methods of inference continue to receive special attention and are widely applied. Many aspects of Bayesian statistics are discussed in the volumes edited by Bernado et al. (1988, 1992, 1996, 1999). Maritz and Lwin (1989) were concerned with empirical Bayes methods. O'Hagan (1994) and Congdon (2003) have useful bibliographies.

In this preliminary chapter we sketch only briefly some of the basic concepts and methods in classical estimation theory.

Parameters and Statistics A cdf that depends on the values of a finite number of quantities $\theta_1, \theta_2, \ldots, \theta_m$ (called *parameters*) is written

$$F(X_1, X_2, \ldots, X_n | \theta_1, \theta_2, \ldots, \theta_m).$$

Often we want to estimate the values of these parameters. This is done using functions of the random variables $T_j \equiv T_j(X_1, X_2, \ldots, X_n)$ called *statistics*. When a statistic T_j is used to estimate a parameter θ_j, it is called an *estimator* of θ_j. An *estimate* is a realized value of an estimator for a particular sample of data.

Properties of Estimators

1. A statistic T_j is said to be an *unbiased estimator* of the parameter θ_j if $E[T_j] = \theta_j$. If $E[T_j] \neq \theta_j$, the estimator is *biased*. The bias is $B(T_j) = E[T_j - \theta_j]$.
2. All distributions with finite means and variances possess unbiased estimators of their means and variances, namely,

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} \qquad \text{and} \qquad s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}, \tag{1.297}$$

   respectively, where n is the sample size.
3. If T_j and T_j^* are both unbiased estimators of the same parameter θ_j, then any weighted average $wT_j + (1-w)T_j^*$ is also an unbiased estimator of θ_j. An estimator is said to be *asymptotically unbiased* if $\lim_{n\to\infty} E[T_j] = \theta_j$.
4. The *relative efficiency* of two unbiased estimators is measured by the inverse ratio of their variances, that is, $\mathrm{Var}(T_j^*)/\mathrm{Var}(T_j)$ measures the efficiency of T_j relative to T_j^*. Comparisons of the efficiencies of biased estimators are often made on the basis of their mean-squared errors. The *mean-squared error* is defined to be

$$E[(T_j - \theta_j)^2] = \mathrm{Var}(T_j) + (E[T_j] - \theta_j)^2. \tag{1.298}$$

5. If a measure of overall efficiency is required when several parameters $\theta_1, \theta_2, \ldots, \theta_m$ are being estimated by the unbiased estimators $T_1, T_2, \ldots, T_m$, respectively, then the *generalized variance* may be used. This is a determinant in which the element in the jth row and jth column is

$$\mathrm{Cov}(T_j, T_{j'} | \theta_1, \theta_2, \ldots, \theta_m).$$

   (Comparisons of the generalized variances of biased estimators are only meaningful if the biases are small enough to be neglected.)
6. A *consistent estimator* is one for which

$$\lim_{n\to\infty} \Pr\left[|T_j - \theta_j| \geq c\right] = 0 \tag{1.299}$$

for all positive c. If T_j is unbiased, then it will also be consistent provided that

$$\lim_{n \to \infty} \mathrm{Var}(T_j) = 0.$$

Consistency is an asymptotic property.

7. A *minimum-variance unbiased estimator* (MVUE) T_j of θ_j is an unbiased estimator of θ_j with a variance that is not greater than that of any other unbiased estimator of θ_j. If T_j is an unbiased estimator of θ_j, then the Cramér–Rao theorem states that the variance of T_j satisfies the *Cramér–Rao inequality*

$$\mathrm{Var}(T_j) \geq \frac{1}{nE[\{\partial \ln f(x)/\partial \theta_j\}^2]}. \tag{1.300}$$

An MVUE may or may not, however, attain the Cramér–Rao lower bound.

8. The *efficiency of an unbiased estimator* is the ratio of its variance to the Cramér–Rao lower bound. An estimator is called an *efficient estimator* if this ratio is unity; it is said to be an *asymptotically efficient estimator* if this ratio tends to unity as the sample size becomes large.

9. A *sufficient estimator* is one that summarizes from the sample of observations all possible information concerning the parameter; that is, no other statistic formed from the observations provides any more information. Such a statistic will exist if and only if the likelihood (see below) can be factorized into two parts, one depending only on the statistic and the parameters and the other depending only on the sample observations. If an unbiased estimator has a variance equal to the Cramér–Rao lower bound, then it must be a sufficient estimator.

10. A family of distributions dependent on a vector of parameters Θ is said to be *complete* if $E_\Theta[h(T)] = 0$ for all values of the parameters implies that $\Pr[h(T) = 0] = 1$ for all Θ, where $h(T)$ is a function of the observations and $E_\Theta[\cdot]$ denotes expectation with respect to the distribution with parameters Θ.

Stuart and Ord (1987) have given a careful and very full account of estimation principles as well as details concerning the major types of estimation procedures.

Estimation Methods

1. The method of *maximum likelihood* is widely advocated. If observed values of $X_1, X_2, \ldots, X_n$ are $x_1, x_2, \ldots, x_n$, then their likelihood is

$$L(x_1, x_2, \ldots, x_n) = \Pr\left[\bigcap_{j=1}^{n} (X_j = x_j | \theta_1, \theta_2, \ldots, \theta_m)\right] \tag{1.301}$$

for discrete distributions and

$$L(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n | \theta_1, \theta_2, \ldots, \theta_m) \tag{1.302}$$

for continuous distributions. In either case the values $\hat{\theta}_1 = T_1$, $\hat{\theta}_2 = T_2, \ldots$, $\hat{\theta}_m = T_m$ that maximize the likelihood are called *maximum-likelihood estimators* (MLEs). (Note that the $\hat{\theta}_j$'s are random variables.) If $X_1, X_2, \ldots, X_n$ are mutually independent and have identical distributions, then under rather general conditions

(i) $\lim_{n\to\infty} E[\hat{\theta}_j | \theta_1, \theta_2, \ldots, \theta_m] = \theta_j$, $j = 1, 2, \ldots, m$, and

(ii) asymptotic estimates of the variances and covariances of the $\hat{\theta}_j$'s are given by the corresponding elements in the inverse of the information matrix evaluated at the maximum-likelihood values.

A MLE may or may not be unique, may or may not be unbiased, and need not be consistent. Nevertheless, MLEs possess certain attractive properties. Under certain mild regularity conditions they are asymptotically MVUEs and are also asymptotically normally distributed. The maximum-likelihood estimation method yields sufficient estimators whenever they exist. Also, if $\hat{\theta}$ is a MLE of θ and if $h(\cdot)$ is a function with a single-valued inverse, then the MLE of $h(\theta)$ is $h(\hat{\theta})$.

Maximizing the likelihood can usually be achieved by solving the equations

$$\frac{\partial L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2, \ldots, \theta_m)}{\partial \theta_i} = 0, \qquad i = 1, 2, \ldots, m; \tag{1.303}$$

these equations are called the *maximum-likelihood equations*. They are often intractable and require iteration (e.g., by the Newton–Raphson method) for their solution. The almost universal accessibility to cheap computing power has led to the development of a number of computer routines for maximum-likelihood estimation. Also the log-likelihood is negative, and so maximizing the likelihood is equivalent to minimizing the absolute value of the log-likelihood; this can be achieved by means of a computer optimization routine. The leading computer packages supply maximum-likelihood and suitable function optimization routines.

Reparameterization, where feasible, so that the new parameters are orthogonal, has been advocated by a number of authors; see, for example, Cox and Reid (1987), Ross (1990), and Willmot (1988b).

2. The *method of moments* usually requires less onerous calculations than maximum-likelihood estimation, although the method cannot be guaranteed to give explicit estimators. The method is based on equating the first k *uncorrected sample moments* about zero, $m'_r = n^{-1} \sum_{j=1}^{n} x_j^r$, $r = 1, \ldots, k$, to the corresponding theoretical expressions for μ'_r, where k is the number of unknown parameters. If preferred, the first k *central sample moments* $m_r = n^{-1} \sum_{j=1}^{n} (x_j - \overline{x})^r$, $r = 1, \ldots, k$, can be equated to the corresponding expressions for the central moments μ_r, or the first k factorial sample moments $m'_{[r]}$, $r = 1, \ldots, k$, can be set equal to $\mu'_{[r]}$; the three procedures give identical estimators. The equation obtained by equating the *sample mean* to the theoretical mean is called the *first-moment equation*.

The higher sample moments generally have large variances, however. This has led to the use of methods based on the quantiles (for continuous distributions) and on the mean and lowest observed frequencies (for discrete distributions). An alternative approach is to solve equations that are approximations to the maximum-likelihood equations. This approach has been discussed by A. W. Kemp (1986) for discrete distributions.

When the method of moments or a similar method leads to explicit estimators, they can be used to provide initial estimates for maximum-likelihood estimation.

The desirable properties of MLEs have led to the development of a number of variants of maximum-likelihood estimation, especially for situations where a family of distributions is to be fitted to data, and hence there are a large number of parameters to be estimated:

Generalized MLEs exist and have near-optimal properties in cases where MLEs do not exist. In most other cases, though, the two methods give identical results (see, e.g., Weiss (1983).

Modified maximum-likelihood estimation is used particularly for censored data. In many situations where maximum-likelihood estimation requires iteration, modified maximum-likelihood estimation gives explicit estimators (see, e.g., Tiku (1989).

Penalized maximum-likelihood estimation is used in curve estimation. It involves a sacrifice of efficiency in order to achieve smooth fits (see, e.g., Silverman (1985).

Partial maximum-likelihood estimation was introduced by Cox (1975) for analyzing regression models involving explanatory variables as a way to reduce the number of nuisance parameters (see, e.g., Kay (1985).

Conditional, marginal, and *profile* likelihood procedures are methods somewhat similar to partial maximum-likelihood estimation; they are collectively described as *pseudolikelihood* methods. They have probabilistic interpretations [see, e.g., Kalbfleisch (1986) and Barndorff-Neilsen (1991)].

Quasi-likelihood estimation is a nonlinear weighted least-squares method for generalized linear models. It is based on families of linear-exponential distributions. Only second-moment assumptions are used, and hence linear exponentiality does not necessarily hold. The method yields equations similar to (1.303) and can be useful for situations with overdispersion (see, e.g., McCullagh (1991).

Interval Estimates Often interval estimates of a parameter θ are wanted. These have the property

$$\Pr(\theta \in \{a \le \theta \le b\}) = 1 - \alpha,$$

where a and b are functions of a statistic (or statistics) derived from the sample data. The confidence probability (confidence level) is the probability that the (random) interval covers the true (fixed) parameter θ.

Confidence intervals are often symmetric but need not be so. There are various ways of determining a and b; inverting a significance test may be a convenient method. Intervals based on minimal sufficient statistics, where they exist, are generally preferred.

For discrete distributions, when a particular value of α is specified, it is often difficult (or maybe impossible) to construct an interval where the confidence probability is exactly $1-\alpha$; This is due to the discrete nature of the statistic(s) on which the interval is based (see, e.g., Agresti and Coull, 1998). References to ingenious methods for obtaining "exact confidence intervals" for discrete distributions are given in later chapters of this book.

Robinson's (1982) introductory article gives useful references. Hahn and Meeker (1991) discuss various kinds of confidence intervals, prediction intervals, and tolerance intervals as well as their applications. They devote a chapter each to the binomial and Poisson distributions.

1.2.17 General Comments on the Computer Generation of Discrete Random Variables

The methods of generating rv's that are discussed in this section depend on an infinite sequence of random numbers $\{U_i\}$ uniformly distributed on [0, 1]. A suitable sequence is customarily generated by a computer, using a *pseudo-random-number algorithm*, that is, an algorithm that generates a deterministic stream of numbers that appears to have the same relevant statistical properties as a sequence of truly random numbers. Linear congruential generators (particularly multiplicative generators), shift registers, generalized feedback shift register generators, lagged-Fibonacci, inversive congruential generators, and nonlinear congruential generators are described in detail by Gentle (1998); see also Lewis and Orav (1989) and L'Ecuyer (1990). The *period of a generator* is the length of the sequence of numbers that it produces before it starts to repeat itself. Research is partly driven by the need in parallel processing for generators with longer periods than many currently in use.

A number of very fast general methods for generating discrete random variables have been developed. Such methods are distribution nonspecific, in the sense that they require tables relating to the actual values of the probability mass function, for example, a table of the cdfs, rather than knowledge of the structural properties of the distribution. They can, in principle, be applied to any univariate discrete distribution and are usually the method of choice when large numbers of rv's are required from a particular distribution with constant parameters. The following methods are particularly suitable for discrete distributions:

(i) The *inversion method* can be used for the generation of both continuous and discrete distributions. A uniform [0, 1] variate is generated and is transformed into a variate from the *target distribution* (the distribution of interest) by the use of a monotone transformation of the uniform cdf to the target cdf; this procedure is called *inversion* of the (target) cdf.

(ii) The *table look-up method* is an adaptation of the inversion procedure that is particularly suitable for a discrete distribution and is very widely used. A set-up routine is required in which the cumulative probabilities for the target distribution are calculated correctly and are stored in computer memory. The cdf for a discrete distribution is of course a step function, with step jumps occurring at successive variate values and step heights equal to successive probabilities. A uniform [0, 1] variate is generated, and the appropriate step height interval within which it lies is sought using a search procedure. The variate value corresponding to this step jump is then "returned" (i.e., made available) as a variate from the target distribution. The use of one of the many sophisticated search procedures that are now available can make this a very fast method.

(iii) Walker's (1974, 1977) *alias method* is based on the following theorem:

> Every discrete distribution with probabilities $p_0, p_1, \cdots, p_{K-1}$ can be expressed as an equiprobable mixture of K two-point distributions.

First the probabilities for the target distribution must be calculated. Next a set-up procedure for constructing the K equiprobable mixtures is required; the information concerning these mixtures can be put into two arrays of size K by an ingenious method described in the books referenced below. One uniform variate on [0, 1] is then used to choose a component in the equiprobable mixture, while a second uniform variate on [0, 1] decides which of the two points for that component should be returned as a target variable. Once the set-up algorithm has been implemented (this may take a nontrivial amount of computer time), the generation of large numbers of variates from the target distribution is very rapid. For implementation details for these methods see, for example, Chen and Asau (1974) (for the indexed table look-up method) and Kronmal and Peterson (1979) (for the alias method).

(iv) If the order in which the variates are generated is immaterial, then C. D. Kemp and A. W. Kemp's (1987) *frequency table method* provides an even faster approach in the fixed-parameter situation. The method generates a sample of values in the form of a frequency table and is useful, for example, for studying the properties of estimators; the method does not attempt to provide a sequence of uncorrelated variate values.

Distributionally nonspecific methods are not, however, suitable when the parameters of a distribution change from call to call to the computer generator. Consequently many different *distribution-specific generation methods* have been devised. Some of these are mentioned in the appropriate chapter later in this book. Special attention has been received by the binomial and Poisson distributions because of their central role in discrete distribution theory. It should be emphasized that the practical implementation of computer generation algorithms is not straightforward; the use of thoroughly tested standard packages (e.g., NAG Libraries [Numerical Algorithms Group], IMSL Libraries [Visual Numerics]) is recommended.

Useful references concerning the computer generation of rv's are the books by Morgan (1984), Ripley (1987), Bratley, Fox, and Schrage (1987), and Dagpunar (1988); Boswell, Gore, Patil, and Taillie (1993) provided a helpful general survey article. Devroye (1986) gives an encyclopedic coverage of the mathematical methodology of nonuniform random-variate generation. Gentle (1998, 2002) covers the recent literature very thoroughly.

1.2.18 Computer Software

The advent of cheap and powerful computing facilities has transformed the statistical analysis of data. The computer languages Algol, APL, and Pascal have largely given way to Fortran 90, C, C++, and S-Plus and the open source variant R.

Many more computer packages are now available. The most flexible of these allow interaction between the user and the package and the incorporation of modules for specific tasks. It is not possible in a limited space to make recommendations. We list some of the major packages with their websites so that the reader can obtain further information.

GenStat	VSN International	www.vsn-intl.com
GLIM	NAG	www.nag.co.uk
IMSL	Visual Numerics	www.vni.com
Minitab	Minitab	www.minitab.com
MLwiN	Centre for Multilevel Modelling	multilevel.ioe.ac.uk
R	R Foundation	www.r-project.org
SAS	SAS Institute	www.sas.com
S-PLUS	Insightful	www.insightful.com
SPSS	SPSS	www.spss.com
Stata	Stata	www.stata.com
StatXact and LogXact	Cytel Software	www.cytel.com

CHAPTER 2

Families of Discrete Distributions

2.1 LATTICE DISTRIBUTIONS

In Section 1.2.3, the class of discrete distributions was defined as having a cumulative distribution function (cdf) that is a step function with only a denumerable number of steps. According to this definition, the class has considerable variety. For example, the distribution defined by

$$\Pr\left[X = \frac{r}{s}\right] = (e-1)^2(e^{r+s}-1)^{-1},$$

where r and s are relatively prime positive integers, is a discrete distribution. Its expected value is $e[1 - \ln(e-1)]$ and all positive moments are finite. However, it is not possible to write down the values of r/s of X in ascending order of magnitude, though it is possible to enumerate them according to the values of r and s.

Most of the discrete distributions used in statistics belong to a much narrower class, the *lattice distributions*. In these distributions the intervals between the values of any one random variable for which there are nonzero probabilities are all integral multiples of one quantity (which depends on the random variable). Points with these coordinates thus form a lattice. By an appropriate linear transformation it can be arranged that all variables take values that are integers. For most of the discrete distributions that we will discuss, the values taken by the random variable cannot be negative.

There are a number of ways of classifying nonnegative lattice distributions. Classification into broad classes helps us to understand the multitude of available distributions and the relationships between them. The mere number of distributions in a class is not, in itself, a measure of importance. What is important is the inclusion of a wide variety within a specific class, with emphasis on the existence of special properties that apply to all members of the class. This is helpful when

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

constructing models and for the derivation of methods of analysis by analogy with known techniques for closely related distributions.

2.2 POWER SERIES DISTRIBUTIONS

2.2.1 Generalized Power Series Distributions

The very broad class of power series distributions includes many of the most common distributions. Membership of the class confers a number of special properties.

A distribution is said to be a *power series distribution* (PSD) if its probability mass function can be written in the form

$$\Pr[X = x] = \frac{a_x \theta^x}{A(\theta)}, \qquad x = 0, 1, \ldots, \qquad \theta > 0, \tag{2.1}$$

where $a_j \geq 0$ and $A(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$. In (2.2) θ is the *power parameter* of the distribution and $A(\cdot)$ is the *series function*.

Continuous and discrete frequency functions of the form

$$f(x) = \frac{a_x \eta(x, \theta)}{A(\theta)},$$

where the a_x depend only on x and not on θ, $\eta(x, \theta) = e^{-\alpha x} = \theta^x$, and $A(\theta)$ is the sum or integral of $a_x \eta(x, \theta)$ over the sample space, were studied by Tweedie (1947, 1965), who called such distributions "Laplacian"; a more modern term is "linear exponential." A PSD is a *discrete linear exponential distribution*; see Lehmann (1986) for inference properties of the exponential family of distributions.

The term "power series distribution" is generally credited to Noack (1950). Kosambi (1949) and Noack showed that many important discrete distributions belong to this class. They also investigated its moment and cumulant properties. The definition (2.1) was extended to multivariate distributions by Khatri (1959).

Patil (1961, 1962a) allowed the set of values that the variate can take to be any nonempty enumerable set S of nonnegative integers. Patil called this extended class *generalized power series distributions* (GPSDs). Estimation and other properties of GPSDs have been explored further in Patil (1962a,b,c, 1964a).

Among distributions of major importance belonging to this class are the binomial (Chapter 3), Poisson (Chapter 4), negative binomial (Chapter 5), and logarithmic (Chapter 7) distributions and their related multivariate distributions. Furthermore, if a GPSD is truncated, then the truncated version is also a GPSD. Also the sum of n mutually independent rv's, each having the same GPSD, has a distribution of the same class, with series function $[\eta(\theta)]^n$. The pgf for (2.1) is

$$G(z) = \frac{\eta(\theta z)}{\eta(\theta)}. \tag{2.2}$$

Kosambi (1949) and Noack (1950) obtained the following results:

For the binomial distribution	$\eta(\theta) = (1+\theta)^n$, n a positive integer;
For the Poisson distribution	$\eta(\theta) = e^\theta$;
For the negative binomial distribution	$\eta(\theta) = (1-\theta)^{-k}$, $k > 0$;
For the logarithmic distribution	$\eta(\theta) = -\ln(1-\theta)$.

The moment generating function (mgf) for a PSD is

$$G(e^t) = \frac{\eta(\theta e^t)}{\eta(\theta)}, \tag{2.3}$$

and hence the mean and variance are

$$E[X] = \mu = \mu(\theta) = \theta \frac{d}{d\theta}[\ln \eta(\theta)], \tag{2.4}$$

$$\mathrm{Var}(X) = \mu_2 = \theta \frac{d\mu}{d\theta} = \theta^2 \frac{d^2}{d\theta^2}[\ln \eta(\theta)] + \mu; \tag{2.5}$$

see Kosambi (1949), who showed that $\mu = \mu_2$ characterizes the Poisson distribution among PSDs; see also Patil (1962a). The higher moments are

$$\begin{aligned} E[X^{r+1}] &= \mu'_{r+1} = \theta \frac{d}{d\theta}[\mu'_r] + \mu\mu'_r, \\ E[(X-\mu)^{r+1}] &= \mu_{r+1} = \theta \frac{d}{d\theta}[\mu_r] + r\mu_2\mu_{r-1} \end{aligned} \tag{2.6}$$

(Craig, 1934; Noack, 1950).

The factorial moment generating function (fmgf) is

$$G(1+t) = \frac{\eta(\theta + \theta t)}{\eta(\theta)}, \tag{2.7}$$

and so the rth factorial moment is

$$\mu'_{[r]} = \frac{\theta^r}{\eta(\theta)} \frac{d^r}{d\theta^r}[\eta(\theta)] \tag{2.8}$$

and

$$\mu'_{[r+1]} = (\mu - r)\mu'_{[r]} + \theta \frac{d}{d\theta}[\mu'_{[r]}] \tag{2.9}$$

(Patil, 1961).

From (2.3) and (2.7) the cumulant generating function (cgf) and factorial cumulant generating function (fcgf) are

$$\ln G(e^t) = \ln \left[\frac{\eta(\theta e^t)}{\eta(\theta)} \right] \tag{2.10}$$

and

$$\ln G(1+t) = \ln\left[\frac{\eta(\theta+\theta t)}{\eta(\theta)}\right]. \tag{2.11}$$

The cumulants satisfy the equation

$$\kappa_{r+1} = \theta\frac{d\kappa_r}{d\theta} \tag{2.12}$$

and the factorial cumulants satisfy

$$\kappa_{[r+1]} = \theta\frac{d\kappa_{[r]}}{d\theta} - r\kappa_{[r]} \tag{2.13}$$

(Khatri, 1959).

From these recurrence relations it can be seen that, if $\kappa_1 = \kappa'_{[1]} = \mu$ is known *as a function of* θ, then all the cumulants (and so all the moments) are determined from this one function. Alternatively, the variance or a higher cumulant might be given as a function of θ (Tweedie and Veevers, 1968). An even more remarkable result was obtained by Khatri (1959). According to this, knowledge of the first two moments (or, equivalently, the first two factorial moments, cumulants, or factorial cumulants) as functions of a parameter ω is sufficient to determine the whole distribution, given that it is a PSD.

For suppose that $\kappa_1 = y_1(\omega)$ and $\kappa_2 = y_2(\omega)$. Then, from (2.12)

$$y_2(\omega) = \theta\frac{dy_1}{d\theta} = \theta\frac{dy_1}{d\omega}\frac{d\omega}{d\theta} \tag{2.14}$$

and from (2.4)

$$y_1(\omega) = \theta\frac{d[\ln\eta(\theta)]}{d\theta};$$

that is,

$$\frac{d\ln\theta}{d\omega} = \frac{1}{y_2(\omega)}\frac{dy_1}{d\omega}, \tag{2.15}$$

$$\frac{d[\ln\eta(\theta)]}{d\omega} = \frac{y_1(\omega)}{y_2(\omega)}\frac{dy_1}{d\omega}. \tag{2.16}$$

Apart from multiplicative constants this pair of equations determines θ and $\eta(\theta)$ as functions of ω. Khatri pointed out that no other pair of consecutive cumulants possesses this property.

Integrals for the tail probabilities have been obtained by Joshi (1974, 1975), who showed that given a family of PSDs with support $\{0, 1, 2, \ldots, n\}$ or $\{0, 1, 2, \ldots\}$ and power parameter θ, where $0 < \theta < \rho$, there exists a family of absolutely continuous distributions with support $(0, \rho)$ such that the lower tail probabilities of the PSDs are equal to the upper tail probabilities of the family of continuous distributions. Three well-known examples are the relationships of

1. Poisson to gamma tail probabilities
2. Binomial to beta-of-the-second-kind tail probabilities
3. Negative-binomial to beta tail probabilities.

Consider now the estimation properties of PSDs.

Suppose that $x_1, x_2, \ldots, x_N$ are N independent observations from the same PSD, defined by (2.1). The likelihood function is then

$$\prod_{j=1}^{N} \left(\frac{a_{x_j} \theta^{x_j}}{\eta(\theta)} \right) = \theta^T [\eta(\theta)]^{-N} \prod_{j=1}^{N} a_{x_j}, \tag{2.17}$$

where $T = \sum_{j=1}^{N} x_j$, and the maximum-likelihood estimator (MLE) $\hat{\theta}$ of θ satisfies the first-moment equation

$$\hat{\theta}^{-1} \sum_{j=1}^{N} x_j - \frac{N\eta'(\hat{\theta})}{\eta(\hat{\theta})} = 0. \tag{2.18}$$

The MLE is thus a function of $T = \sum_{j=1}^{N} x_j$ and does not depend on the x_j's in any other way.

The conditional distribution of $X_1, \ldots, X_N$, given $\sum_{j=1}^{N} X_j$, does not depend on θ; that is, $T = \sum_{j=1}^{N} x_j$ is sufficient for θ. It is also complete, since it has a GPSD and the equation

$$\sum_{j=0}^{\infty} A_j [\eta(T)\theta]^j = 0 \quad \text{for all } \theta \tag{2.19}$$

implies that $\eta(T) = 0$. Results for the asymptotic variance and for the bias in $\hat{\theta}$ were obtained by Patil (1962a,b).

A minimum variance unbiased estimator (MVUE) exists if and only if (iff) the PSD has support $\{a_1 + a_2 x\}$, where $x = 0, 1, \ldots$ and a_1 and a_2 are nonnegative integers. It follows that while the logarithmic and left-truncated Poisson distributions are minimum-variance unbiased estimable, the binomial and right-truncated Poisson are not; see Roy and Mitra (1957) and Tate and Goen (1958). When it exists, the MVUE of θ is

$$\theta^\star = \begin{cases} \dfrac{b\left(\sum_{j=1}^{N} x_j - 1\right)}{b\left(\sum_{j=1}^{N} x_j\right)} & \text{if} \quad \sum_{j=1}^{N} x_j > 0, \\ 0 & \text{if} \quad \sum_{j=1}^{N} x_j = 0, \end{cases} \tag{2.20}$$

where $b(k)$ is the coefficient of θ^k in the expansion of $[\eta(\theta)]^N$ (Roy and Mitra, 1957). Properties associated with MVUEs for PSDs have been obtained by Patil (1963b) and Patil and Joshi (1970). Estimators based on ratios and moments were studied in Patil (1962c). Abdul-Razak and Patil (1986) have investigated Bayesian inference for PSDs. Eideh and Ahmed (1989) have investigated tests, based on the Kullback–Leibler information measure, for a one-parameter PSD.

Generalized power series distributions with two parameters, that is, with

$$a_x = a_x(\lambda) \qquad \text{and} \qquad \eta(\theta) = \eta(\theta, \lambda) = \sum_x a_x(\lambda)\theta^x,$$

were studied in Patil (1964a) and were found to possess many of the properties of one-parameter PSDs; see also Khatri (1959) and Douglas (1980).

2.2.2 Modified Power Series Distributions

Modified power series distributions (MPSDs) form an extension of the class of GPSDs. They were created by R. C. Gupta (1974) by replacing θ^x in (2.1) by $[u(\theta)]^x$; that is, their pmf's have the form

$$\Pr[X = x] = \frac{a_x[u(\theta)]^x}{\eta(\theta)}, \tag{2.21}$$

where the support of X is $0, 1, 2, \ldots$ or a subset thereof, $a_x \geq 0$, and $u(\theta)$ and $\eta(\theta)$ are positive, finite, and differentiable. Like GPSDs, MPSDs are linear exponential.

The series function is now

$$\eta(\theta) = \sum_x a_x[u(\theta)]^x. \tag{2.22}$$

Differentiating with respect to θ gives

$$\frac{d\eta(\theta)}{d\theta} = \sum_x x a_x[u(\theta)]^{x-1}\frac{du(\theta)}{d\theta}, \tag{2.23}$$

and so

$$\begin{aligned} E[X] = \mu = \mu(\theta) &= \sum_x \frac{x a_x[u(\theta)]^x}{\eta(\theta)} \\ &= \frac{u(\theta)}{\eta(\theta)}\frac{\eta'(\theta)}{u'(\theta)}. \end{aligned} \tag{2.24}$$

In a similar manner, differentiating

$$\mu_r' = \sum_x x^r \frac{a_x[u(\theta)]^x}{\eta(\theta)}$$

with respect to θ yields

$$\mu'_{r+1} = \frac{u(\theta)}{u'(\theta)} \frac{d\mu'_r}{d\theta} + \mu\mu'_r. \tag{2.25}$$

Also

$$\mu_{r+1} = \frac{u(\theta)}{u'(\theta)} \frac{d\mu_r}{d\theta} + r\mu_2\mu_{r-1}, \tag{2.26}$$

and

$$\mu'_{[r+1]} = \frac{u(\theta)}{u'(\theta)} \frac{d\mu'_{[r]}}{d\theta} + (\mu - r)\mu'_{[r]}. \tag{2.27}$$

In particular

$$\mathrm{Var}(X) = \frac{u(\theta)}{u'(\theta)} \frac{d\mu}{d\theta}. \tag{2.28}$$

R. C. Gupta (1974) also obtained a relationship between the cumulants and the moments of a MPSD. Further moment properties have been obtained by R. C. Gupta (1975b, 1984), P. L. Gupta (1982), Gupta and Singh (1981), and Kumar and Consul (1979). Tripathi, Gupta, and Gupta (1986) have studied the incomplete moments of MPSD. Length-biased MPSDs are studied in Gupta and Tripathi (1992).

The class of MPSDs reduces to the class of GPSDs when $u(\theta) = \theta$; the above formulas become the corresponding ones in Section 2.2.1.

Steyn's (1980, 1984) *two-parameter* and *multiparameter power series distributions* (see Section 11.2.17) are important extensions of MPSDs.

When $u(\theta)$ is invertible (e.g., by a Lagrangian expansion; see Section 2.5) and θ can be expressed as $\theta = \psi(u(\theta))$, the pgf can be written as

$$G(z) = \frac{\eta(\psi(u(\theta))z)}{\eta(\psi(u(\theta)))}. \tag{2.29}$$

Maximum-likelihood estimation for MPSDs has been researched by Gupta (1975a), who showed that the maximum-likelihood estimate of θ is given by

$$\mu(\hat{\theta}) = \overline{x}. \tag{2.30}$$

He obtained general expressions for the bias and asymptotic variance of $\hat{\theta}$ and showed that it is unbiased only in the case of the Poisson distribution.

Minimum-variance unbiased estimation has been studied when the support of the distribution is known by Jain and Gupta (1973), Gupta (1977a), Jani (1978b), Jani and Shah (1979a,b), and Kumar and Consul (1980). It has been studied when the support is unknown by Jani (1977, 1978a,b), Patel and Jani (1977), and Kumar and Consul (1980). The MVUE for the probability mass function has been investigated in Gupta and Singh (1982). Famoye and Consul (1989) have studied confidence intervals for MPSDs.

A number of characterizations have been obtained. Jani (1978b) has shown that a discrete distribution is a MPSD iff its cumulants satisfy the recurrence relation

$$\kappa_{r+1} = \frac{u(\theta)}{u'(\theta)} \frac{d\kappa_r}{d\theta}. \tag{2.31}$$

An integral expression for the tail probabilities of a MPSD in terms of absolutely continuous distributions has been derived by Jani and Shah (1979a). Gupta and Tripathi (1985) have given a number of other references to work on MPSDs.

Consul (1990b) has studied the subclass of MPSDs for which $u(\theta) = \mu$ (the mean of the distribution); the subclass contains several well–known distributions, such as the binomial, Poisson, negative binomial, and Lagrangian Poisson distributions. When $\mu = \theta$, (2.28) becomes

$$\operatorname{Var}(X) = \mu_2 = \frac{u(\theta)}{u'(\theta)} = \frac{u(\mu)}{u'(\mu)}; \tag{2.32}$$

(2.25)–(2.27) and (2.31) simplify on setting $u(\mu)/u'(\mu) = \mu_2$.

Consul has given several characterizations for this subclass. From (2.24) and (2.32),

$$\eta(\mu) = \exp\left(\int \frac{\mu\, d\mu}{\mu_2}\right) \qquad \text{and} \qquad u(\mu) = \exp\left(\int \frac{d\mu}{\mu_2}\right). \tag{2.33}$$

Hence, for specified support, the variance μ_2 as a function of the mean completely determines the distribution within this subclass [the constants induced by the integrations in (2.33) can be shown not to affect the resultant distribution].

Kosambi (1949) and Patil (1962a) proved that equality of the mean and variance within the family of PSDs characterizes the Poisson distribution; Gokhale (1980) strengthened the mean–variance result by proving that, within PSDs, $\mu_2 = m(1 - mc)$ iff X has a negative binomial, Poisson, or binomial distribution according to whether c is negative, zero, or $1/n$, where n is a positive integer.

Consul (1990b) considered characterizations of members of his subclass where the variance is a particular cubic function of the mean. The question "Do there exist some new families of discrete probability distributions [in the subclass of MPSD for which the mean is the power parameter] such that the variance equals a fourth-degree function of the mean m?" was posed by Consul (1990b); it remains as yet unanswered.

Deformations of a modified power series variable Y such that

$$\Pr[X = x] = \begin{cases} \Pr[Y = c] + \alpha \Pr[Y = c + 1], & x = c, \\ (1 - \alpha) \Pr[Y = c + 1], & x = c + 1, \\ \Pr[Y = x], & x \neq c, c + 1, \end{cases}$$

were investigated by Murat and Szynal (2003). These authors also considered deformations to factorial series distributions; for factorial series distributions, see Section 2.7.

A number of interesting MPSDs that are *not* GPSDs are, however, Lagrangian distributions. These include the Lagrangian Poisson, the Lagrangian negative binomial, and the Lagrangian logarithmic distributions; see Section 7.2. The two classes, MPSD and Lagrangian, clearly overlap. Consul (1981) has helped to clarify the situation by introducing a yet broader class of Lagrangian distributions; see Section 2.5.

2.3 DIFFERENCE-EQUATION SYSTEMS

2.3.1 Katz and Extended Katz Families

Pearson (1895) used the equation

$$\frac{f_r - f_{r-1}}{f_r} = \frac{r-a}{b_0 + b_1 r + b_2 r^2}, \qquad r \geq 1, \tag{2.34}$$

as a starting point for obtaining (by a limiting process) the differential equation defining the *Pearson system of continuous distributions*; see Chapter 12. Here a, b_0, b_1, and b_2 are parameters. Pearson apparently did not pursue the development of a *discrete* analog of his continuous system.

The difference equation (2.34) was used by Carver (1919) for smoothing actuarial data. Although Carver (1923) obtained expressions for the parameters in terms of the moments, he did not attempt a thorough examination of the discrete distributions arising from (2.34).

A detailed study of the simpler relationship

$$\frac{p_{x+1}}{p_x} = \frac{\alpha + \beta x}{1+x}, \qquad x = 0, 1, \ldots, \tag{2.35}$$

was undertaken by Katz in his (1945) thesis, in abstracts (Katz, 1946, 1948), and in a longer form in Katz (1965). This system of distributions is known as the *Katz family*.

The restrictions on the parameters are $\alpha > 0$, $\beta < 1$ ($\beta \geq 1$ does not yield a valid distribution). If $\alpha + \beta n < 0$, then p_{n+j} is understood to be equal to zero for all $j > 0$.

Katz showed that there are three possibilities:

1. When $\beta < 0$, the outcome is the binomial distribution with $\alpha = nq/p$, $\beta = -q/p$, that is, $n = -\alpha/\beta$, $p = \beta/(\beta - 1)$.
2. When $\beta = 0$, the outcome is the Poisson distribution with $\alpha = \theta$.
3. When $0 < \beta < 1$, the outcome is the negative binomial distribution with $\alpha = kP/(P+1)$, $\beta = P/(P+1)$, that is, $k = \alpha/\beta$, $P = \beta/(1-\beta)$, using the parameterization of (5.1).

He suggested the reparameterization

$$\xi = \frac{\alpha}{1-\beta} \quad (=\mu) \qquad \text{and} \qquad \eta = \frac{\beta}{1-\beta} \quad \left(=\frac{\sigma^2-\mu}{\mu}\right)$$

and gave the maximum-likelihood equations for the distributions in terms of these new parameters.

From (2.35),

$$(x+1)^{r+1} p_{x+1} = (\alpha + \beta x)(x+1)^r p_x; \tag{2.36}$$

summing both sides with respect to x gives

$$\mu'_{r+1} = \sum_{j=0}^{r} \binom{r}{j} (\alpha \mu'_j + \beta \mu'_{j+1}), \tag{2.37}$$

whence

$$\mu = \frac{\alpha}{1-\beta}, \qquad \mu'_2 = \frac{\alpha + (\alpha+\beta)\mu}{1-\beta}, \tag{2.38}$$

that is,

$$\mu_2 = \frac{\alpha}{(1-\beta)^2}. \tag{2.39}$$

Furthermore

$$\mu_3 = \mu_2(2c-1), \qquad \mu_4 = 3\mu_2^2 + \mu_2(6c^2 - 6c + 1), \tag{2.40}$$

where $c = \mu_2/\mu = (1-\beta)^{-1}$.

A major motivation for the Katz (1965) paper was the problem of discriminating between these three distributions when a given set of data is known to come from one or other of them. Katz found that for these three distributions $Z = (s^2 - \overline{x})/\overline{x}$ is very approximately normally distributed with mean $c-1$ and variance $2/N$, where N is the sample size. Tests for $H_0 : \mu_2 = \mu$ against $H_1 : \mu_2 < \mu$ (or alternatively $H_1 : \mu_2 > \mu$) are tests for a Poisson null hypothesis against a binomial alternative hypothesis (or a negative binomial alternative hypothesis).

The pgf for the Katz family is

$$G(z) = \left(\frac{1-\beta z}{1-\beta}\right)^{-\alpha/\beta}. \tag{2.41}$$

The fmgf is

$$G(1+t) = \left(\frac{1-\beta-\beta t}{1-\beta}\right)^{-\alpha/\beta} \tag{2.42}$$

and the rth factorial moment is

$$\mu'_{[r]} = \left(\frac{\alpha}{\beta}\right)\cdots\left(\frac{\alpha}{\beta}+r-1\right)\left(\frac{\beta}{1-\beta}\right)^r$$
$$= \left(\frac{\alpha + r\beta - \beta}{1-\beta}\right)\mu'_{[r-1]} \tag{2.43}$$

for $r \geq 1$, with $\mu'_{[0]} = 1$.

Ottestad (1939) had previously used the ratio $\mu'_{[r+1]}/\mu'_{[r]}$ as the basis for a graphical discrimination test; the slope of the sample estimate of $\mu'_{[r+1]}/\mu'_{[r]}$ plotted against r can be expected to be positive, zero, or negative for the negative binomial, Poisson, and binomial distributions, respectively.

Guldberg's (1931) approach to deciding whether any of the three distributions is appropriate was to show that

$$\frac{(x+1)p_{x+1}}{p_x} + \frac{(\mu-\mu_2)x}{\mu_2} = \alpha = \frac{\mu^2}{\mu_2};$$

that is,

$$T_x\mu_2 + (\mu-\mu_2)x = \mu^2, \tag{2.44}$$

where $T_x = (x+1)p_{x+1}/p_x$. The sample values

$$\frac{s^2(x+1)f_{x+1}/f_x + (\overline{x}-s^2)x}{\overline{x}^2} \tag{2.45}$$

(where f_x is the frequency of the observation x in the sample) should therefore be approximately equal to unity.

The mean deviation for the three distributions in the Katz family can be obtained by summing (2.36), with $r = 0$, over all values of x not less than m, where $m = [\mu]$ (the largest integer not greater than μ). From (2.38) and (2.39) we find $\alpha + \beta\mu = \mu_2(1-\beta)$; this gives

$$E[|X-\mu|] = 2p_m\left(\frac{\alpha+\beta m}{1-\beta}\right) \approx 2p_m\mu_2, \tag{2.46}$$

with an error of $2p_m(m-\mu)\beta/(1-\beta)$. The formula is exact if μ is an integer and also when $\beta = 0$; see Bardwell (1960) and Kamat (1965).

Consider now the *extended Katz family* of Gurland and Tripathi (1975) and Tripathi and Gurland (1977). The relationship between the probabilities now has the form

$$\frac{p_{x+1}}{p_x} = \frac{\alpha+\beta x}{\gamma+x}, \qquad \alpha > 0, \qquad \beta < 1, \qquad \gamma > 0, \tag{2.47}$$

yielding an extended hypergeometric distribution with pgf

$$G(z) = \frac{{}_2F_1[\alpha/\beta, 1; \gamma; \beta z]}{{}_2F_1[\alpha/\beta, 1; \gamma; \beta]}; \tag{2.48}$$

see Section 6.11.1. Yousry and Srivastava (1987) also studied the relationship (2.47); they called the outcome a *hyper-negative binomial* distribution.

When $\beta \to 0$ in (2.47),

$$\frac{p_{x+1}}{p_x} = \frac{\alpha}{\gamma + x}, \tag{2.49}$$

and the pgf becomes

$$G(z) = \frac{{}_1F_1[1; \gamma; \alpha z]}{{}_1F_1[1; \gamma; \alpha]}; \tag{2.50}$$

this is the hyper-Poisson distribution of Bardwell and Crow (1964) and Crow and Bardwell (1965); see Section 4.12.4. An extended form of this distribution with

$$\frac{p_{x+1}}{p_x} = \frac{\alpha(\rho + x)}{(1 + x)(\gamma + x)} \tag{2.51}$$

and pgf

$$G(z) = \frac{{}_1F_1[\rho; \gamma; \alpha z]}{{}_1F_1[\rho; \gamma; \alpha]} \tag{2.52}$$

was studied by Gurland and Tripathi (1975) and Tripathi and Gurland (1977, 1979); see Section 4.12.4.

Tripathi and Gurland have designated these four families of distributions K [for Katz (2.35)], EK [for extended Katz (2.47)], CB [for Crow and Bardwell (2.49)], and ECB [for extended Crow and Bardwell (2.51)]. All four are special forms of Kemp's generalized hypergeometric probability family of distributions; see Section 2.4.1. Tripathi and Gurland have examined various types of estimators, especially minimum χ^2 estimators. They have been particularly concerned with the problem of selecting an appropriate member from a combined set of distributions (e.g., K, EK, ECB) when fitting data for which a specific model is unclear.

2.3.2 Sundt and Jewell Family

Sundt and Jewell (1981) and Willmot (1988a) have studied the *Sundt-and-Jewell family* of distributions. Their interest lies in actuarial applications, so many of their papers are in actuarial journals. The family is a modified form of the Katz family. The recurrence relationship for the probabilities is

$$\frac{p_{x+1}}{p_x} = \frac{a+b+ax}{1+x}, \qquad x = 1, 2, 3, \ldots; \tag{2.53}$$

taking $\alpha = a + b$, $\beta = a$ shows this is the same as the Katz relationship (2.35) for all x except for the exclusion of $x = 0$.

The pgf for the Sundt-and-Jewell family has the form

$$\begin{aligned} G(z) &= c + (1-c)\left(\frac{1-az}{1-a}\right)^{-(a+b)/a} \\ &= c + (1-c)H(z), \end{aligned}$$

where $H(z)$ is the pgf for a Katz distribution. A necessary restriction on c is $H(0)/[H(0) - 1] \le c < 1$. The Sundt-and-Jewell family therefore contains the modified binomial, modified Poisson, and modified negative binomial distributions (Section 8.2.3) as well as the binomial, Poisson, and negative binomial distributions themselves (Chapters 3, 4, and 5, respectively). It also includes the logarithmic distribution (Chapter 7), the zero-modified logarithmic distribution (Sections 7.1.10 and 8.2.3), and Engen's extended negative binomial distribution (Section 5.12.2).

Klugman, Panjer, and Willmot (1998) have adopted the following terminology:

1. Their $(a, b, 0)$ class has the recurrence relationship $p_x/p_{x-1} = a + b/x$ for $x \ge 1$, that is,

   $$\frac{p_{x+1}}{p_x} = \frac{a+b+ax}{1+x} \quad \text{for } x \ge 0.$$

   This is the Katz family with $\alpha = a + b$, $\beta = a$ and so contains the binomial, Poisson, and negative binomial distributions.
2. **(i)** The $(a, b, 1)$ zero-truncated class has $p_0 = 0$, $p_x/p_{x-1} = a + b/x$ for $x \ge 2$. This class contains the zero-truncated binomial, zero-truncated Poisson, and zero-truncated negative binomial distributions as well as the logarithmic and Engen extended negative binomial distribution.

   (ii) The $(a, b, 1)$ zero-modified class is obtained from the $(a, b, 1)$ zero-truncated class by modifying it by the addition of an arbitrary amount of probability at zero. It contains the zero-modified binomial, zero-modified Poisson, zero-modified negative binomial, zero-modified logarithmic, and zero-modified extended negative binomial distributions.

2.3.3 Ord's Family

Ord's (1967a,b,c, 1972) difference-equation family—*Ord's family*—comprises all the distributions that satisfy

$$\Delta p_{x-1} = p_x - p_{x-1} = \frac{(a-x)p_x}{(a+b_0)+(b_1-1)x+b_2x(x-1)}, \tag{2.54}$$

where $p_x = \Pr[X = x]$ and x takes some range of integer values $\{T\}$.

Taking $b_0 = b_2 = 0$ in (2.54) gives

$$\frac{p_{x-1}}{p_x} = \frac{bx}{a+(b-1)x},$$

that is, the Katz family; Ord labeled the distributions III(B), III(P), and III(N) for the binomial, Poisson, and negative binomial, respectively. He noted furthermore that $b_2 = 0$, $b_0 \neq 0$ leads to a system of "hyper" distributions, in particular to the hyper-Poisson distribution; see Section 4.12.4.

When $b_2 \neq 0$, the nature of the roots of the quadratic in the denominator of (2.54) is central to Ord's classification system. He set

$$\kappa = \frac{(b_1-b_2-1)^2}{4(a+b_0)b_2} \tag{2.55}$$

and also used $I = \mu_2/\mu$; for $0 \le \kappa \le 1$ the roots are imaginary. Table 2.1 summarizes the types of distributions available in Ord's family.

Ord (1967b) published a (β_1, β_2) chart, where β_1 and β_2 are the usual indices of skewness and kurtosis, analogous to the (β_1, β_2) chart for Pearson's continuous

Table 2.1 Summary of Types of Distributions Derived by Ord (1967a)

Type	Name	Criteria	Comments
I(a)	Hypergeometric	$I < 1, \kappa > 1$	Chapter 6
I(b)	Negative hyper geometric, i.e., beta–binomial	$\kappa < 0$	Chapter 6
VI	Beta–Pascal	$I > 1, \kappa > 1$	Chapter 6
II(a)		$I < 1, \kappa = 1$	Symmetric form of I(a)
II(a)		$I < 1, \kappa = 0$	Symmetric form of I(b)
III(B)	Binomial	$I < 1, \kappa \to \infty$	Chapter 3
III(P)	Poisson	$I = 1, \kappa \to \infty$	Chapter 4
III(N)	Negative binomial	$I > 1, \kappa \to \infty$	Chapter 4
VII	Discrete Student's t	$0 < \kappa < 1$	Chapter 11

system. However, he later considered this to be not very useful in the discrete case [see Ord (1985)].

He turned his attention to the measures $I = \mu_2/\mu$ and $S = \mu_3/\mu_2$. A diagram of the (S, I) regions was published by Ord (1967c, 1972); note that I is necessarily positive. For the Katz family $S = 2I - 1$, with $0 < I < 1$, $I = 1$, and $1 < I$ for the binomial, Poisson, and negative binomial distributions, respectively. Also $S < 2I - 1$ gives the beta-binomial, and $S > 2I - 1$ yields the hypergeometric or the beta–Pascal according to whether, in addition, $S < 1$ or $S > 1$. The parameters of the distributions in the family are determined by the first three moments; see Ord (1967c). The (S, I) diagram can therefore be used as an aid for selecting an appropriate model for a given data set.

Ord's preferred selection method (Ord, 1967b) uses a plot of $u_r = rf_r/f_{r-1}$ against r, where f_r is a sample frequency; some degree of smoothing of the u_r may be helpful. The exact relationships that are satisfied by rp_r/p_{r-1} are summarized in Tables 2.2 and 2.3. Although sample ratios rf_r/f_{r-1} cannot be expected to satisfy these relationships exactly, Ord anticipated that sample plots will give a fair indication of an appropriate type of distribution.

Properties of members of the family were reviewed by Ord (1972). He included the following properties: modality (Janardan and Patil, 1970); mean deviation (Kamat, 1966; Ord, 1967c); mean difference (Katti, 1960; Ord, 1967c); incomplete moments (Guldberg, 1931; Kamat, 1965); and moment estimation and maximum-likelihood estimation (Carver, 1923; Ord, 1967b).

Table 2.2 Intercept and Slope of the Plot of xf_x/f_{x-1} against x for Certain Discrete Distributions

Publisher's Note:
Permission to reproduce this image online was not granted by the copyright holder. Readers are kindly requested to refer to the printed version of this article.

Table 2.3 Beginning and End Points of the Hypergeometric Distribution Curves Compared with the Binomial and Negative Binomial Lines

Publisher's Note:
Permission to reproduce this image online was not granted by the copyright holder. Readers are kindly requested to refer to the printed version of this article.

Bowman, Shenton, and Kastenbaum (1991) have studied an extension of Ord's family with

$$p_x = \left(1 + \frac{\alpha - x}{C_0 + C_1 y + C_2 y^2}\right) p_{x-1}, \tag{2.56}$$

where $y = x - \mu = x - E[X]$; the ratio of successive probabilities is here the ratio of two quadratic expressions in x. There are two possibilities: (i) distributions defined on the nonnegative integers (finite or infinite support) and (ii) distributions defined on the negative integers as well. There are four parameters, α, C_0, C_1, and C_2. The authors identified certain forms of these distributions and found it reasonably straightforward to estimate the four parameters using the first four moments (assuming that they exist) in case (i). For case (ii) slowly converging series may be encountered. Bowman, Shenton, and Kastenbaum investigated further extensions with recurrence relations such as

$$p_x = \left(1 + \frac{\alpha - x}{C_0 + C_1 y + C_2 y^2 + C_3 y^3 + C_4 y^4}\right) p_{x-1}, \tag{2.57}$$

with $x = \pm 1, \pm 2, \ldots,$ $y = x - \mu = x - E[X]$; these extensions require six or more parameters. Their *discrete normal distribution* received special attention. They did not consider applications to empirical data or estimation problems.

2.4 KEMP FAMILIES

2.4.1 Generalized Hypergeometric Probability Distributions

The distributions in Kemp's (1968a,b) very broad family of *generalized hypergeometric probability distributions* (GHPDs, or HP distributions for short) have many useful properties. Their pgf's have the form ${}_pF_q[\lambda z]/{}_pF_q[\lambda]$, where

$$\begin{aligned}
&{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda z] \\
&\quad = 1 + \frac{a_1 \cdots a_p(\lambda z)}{b_1 \cdots b_q} + \frac{a_1(a_1+1)\cdots a_p(a_p+1)(\lambda z)^2}{b_1(b_1+1)\cdots b_q(b_q+1)\,2!} + \cdots \\
&\quad = \sum_{j=0}^{\infty} \frac{(a_1)_j \cdots (a_p)_j (\lambda z)^j}{(b_1)_j \cdots (b_q)_j \; j!},
\end{aligned} \tag{2.58}$$

and $(u)_j = u(u+1)\cdots(u+j-1)$ (Pochhammer's symbol). This is known in the theory of special functions as a generalized hypergeometric function (see Section 1.1.8). Various models for HP distributions are discussed later in this section. These distributions are difference-equation distributions; when $\lambda \neq 1$, they are PSDs. They are also limiting forms as $q \to 1$ of *q-hypergeometric distributions*; see Section 2.9.

A second family of distributions, *generalized hypergeometric factorial moment distributions* (GHFDs, or HF distributions for short) have pgf's of the form ${}_pF_q[\lambda(z-1)]$; see Kemp (1968a) and Kemp and Kemp (1974). This family includes not only some of the most common distributions but also a number of important matching and occupancy distributions; see Sections 10.3 and 10.4.1. Their properties and modes of genesis are examined in Section 2.4.2.

The two families overlap. Their intersection includes some of the most common distributions, including the binomial, Poisson, negative binomial, and hypergeometric; such distributions have duplicate sets of properties, including very simple relationships for both their probabilities and their factorial moments.

Both families are contained within an even more general class of distributions whose pgf's have the form ${}_pF_q[\lambda z+\xi]/{}_pF_q[\lambda+\xi]$. These are *generalized hypergeometric recast distributions* (GHRDs, or HR distributions for short); see the end of Section 2.4.2. They arise from certain ascertainment models.

In Section 2.2.1 we saw that PSDs have pgf's of the form

$$G(z)=\frac{\eta(\lambda z)}{\eta(\lambda)}=\frac{\sum_{x\geq 0}\alpha_x\lambda^x z^x}{\sum_{x\geq 0}\alpha_x\lambda^x},\tag{2.59}$$

where λ is the power parameter and α_x is a function of x independent of λ. Suppose now that α_{x+1}/α_x is the ratio of two polynomials in x with all roots real:

$$\frac{\alpha_{x+1}}{\alpha_x}=\frac{(a_1+x)\ldots(a_p+x)}{(b_1+x)\ldots(b_q+x)(b_{q+1}+x)}\tag{2.60}$$

(the coefficients of the highest powers of x in these polynomials can be assumed to be unity without loss of generality concerning the form of the resultant distribution). Then

$$\frac{\Pr[X=x+1]}{\Pr[X=x]}=\frac{(a_1+x)\ldots(a_p+x)\lambda}{(b_1+x)\ldots(b_q+x)(b_{q+1}+x)}.\tag{2.61}$$

Assuming that X is a counting variable with $\Pr[X=0]\neq 0$, then when $\eta(\lambda)$ is a convergent series in λ, the support of X is $0,1,\ldots$, and when $\eta(\lambda)$ is a terminating series in λ, the support is $0,1,\ldots,n$. The corresponding pgf is

$$\begin{aligned}G(z)&=K\left(1+\frac{a_1\cdots a_p(\lambda z)}{b_1\cdots b_{q+1}}+\frac{a_1(a_1+1)\cdots a_p(a_p+1)(\lambda z)^2}{b_1(b_1+1)\cdots b_{q+1}(b_{q+1}+1)}+\cdots\right)\\&=\frac{{}_{p+1}F_{q+1}[1,a_1,\ldots,a_p;b_1,\ldots,b_{q+1};\lambda z]}{{}_{p+1}F_{q+1}[1,a_1,\ldots,a_p;b_1,\ldots,b_{q+1};\lambda]}.\end{aligned}\tag{2.62}$$

This is a particular HP distribution. Clearly the numerator parameters are commutable; so are the denominator ones. Suppose now that one of the denominator parameters in (2.60) is equal to unity, say $b_{q+1} = 1$; then (2.61) and (2.62) become

$$\frac{\Pr[X = x+1]}{\Pr[X = x]} = \frac{(a_1 + x)\dots(a_p + x)\lambda}{(b_1 + x)\dots(b_q + x)(x+1)} \tag{2.63}$$

and

$$\begin{aligned} G(z) &= K\left(1 + \frac{a_1 \cdots a_p(\lambda z)}{b_1 \cdots b_q 1!} + \frac{a_1(a_1+1)\cdots a_p(a_p+1)(\lambda z)^2}{b_1(b_1+1)\cdots b_q(b_q+1)2!} + \cdots\right) \\ &= \frac{{}_pF_q[a_1, \dots, a_p; b_1, \dots, b_q; \lambda z]}{{}_pF_q[a_1, \dots, a_p; b_1, \dots, b_q; \lambda]}; \end{aligned} \tag{2.64}$$

this is a general HP distribution [see Kemp (1968a,b)]. Note that (2.62) can be treated as a special case of (2.64), whereas it might be thought that the reverse holds.

The relationship (2.61) for successive probabilities seems to have appeared for the first time in a paper by Guldberg (1931), who concentrated on recurrence relationships for the moments for the special cases corresponding to the binomial, Poisson, Pascal, and hypergeometric distributions; see also Qvale (1932). Restating (2.61) in the form

$$\begin{aligned} &\frac{\Pr[X = x+1] - \Pr[X = x]}{\Pr[X = x]} \\ &\qquad = \frac{[(a_1 + x)\dots(a_p + x)\lambda] - [b_1 + x)\dots(b_{q+1} + x)]}{(b_1 + x)\dots(b_{q+1} + x)} \end{aligned} \tag{2.65}$$

emphasizes the relationship between HP distributions and difference-equation distributions.

The Katz family corresponds to (2.63) with $p = 1, q = 0$, while the extended Katz family corresponds to (2.61) with $p = 1, q = 0$; see Section 2.3.1. For Tripathi and Gurland's extended Bardwell–Crow family $p = 1, q = 1$ in (2.63); again see Section 2.3.1. Ord's difference-equation family (Section 2.3.3) has $p = 2, q = 1, \lambda = 1$ in (2.61), but note that Ord's family contains distributions for which the variable takes negative values; such distributions are no longer PSDs. Kemp and Kemp's (1956a) earlier "generalized" hypergeometric distributions with $p = 2, q = 1, \lambda = 1$ in (2.63) are GHPDs and are included within Ord's family; they utilize the Gaussian hypergeometric function of Section 1.1.6 (see Sections 6.1 and 6.2.4).

Kemp (1968a,b) showed that terminating distributions can arise for all nonnegative integer values of p and q, provided that at least one of the a_i is a negative integer and that the other parameters are chosen suitably. A shifted HP distribution (with support $m + 1, m + 2, \dots$) may arise if one of the b_j is a negative integer. The existence of nonterminating members of the family depends on the sign of $p - q - 1$; they exist when

1. $p - q - 1 < 0$, provided that $\lambda > 0$ and $a_i, b_j > 0\,\forall i, j$ (except that a pair of a_i, b_j may be negative provided that they lie in the interval between two consecutive nonpositive integers);
2. $p - q - 1 = 0$, provided that $0 < \lambda < 1$ and $a_i, b_j > 0\,\forall i, j$ (except as above);
3. $p - q - 1 = 0$ and $\lambda = 1$, provided that $\sum_{i=1}^{p} a_i < \sum_{j=1}^{q} b_j$ and $a_i, b_j > 0\ \forall\, i, j$ (except as above).

(Nonterminating members do not exist when $p > q + 1$.)

Table 2.4 is a shortened version of one in Dacey (1972). It gives a summary of probability distributions with pgf's expressible in terms of ${}_pF_q(\lambda z)$. As the table shows, a very large number of distributions belong to the family. The names of the distributions are taken from Patil and Joshi (1968). The notation of Patil et al. (1984) is used, and their dictionary should be consulted for the constraints on the parameter values.

The reversed form of a GHPD is also a GHPD. Truncating the first m probabilities of a GHPD gives a shifted GHPD; truncating all except the first m probabilities gives a new (terminating) GHPD.

The pgf satisfies the differential equation

$$\theta \prod_{j=1}^{q} (\theta + b_j - 1) G(z) = \lambda z \prod_{i=1}^{p} (\theta + a_i) G(z), \tag{2.66}$$

where θ is the differential operator $zd/dz \equiv zD$. Using

$$z^k D^k \equiv \theta(\theta - 1) \cdots (\theta - k + 1) \qquad \text{and} \qquad \theta^k \equiv \sum_{j=1}^{k} S(k, j) z^j D^j,$$

where $S(k, j)$ are Stirling numbers of the second kind, (2.66) can be transformed into a differential equation of the form

$$f_1(D)G(z) = \lambda f_2(D)G(z). \tag{2.67}$$

The fmgf satisfies

$$f_1(D)G(1 + t) = \lambda f_2(D)G(1 + t); \tag{2.68}$$

identifying the coefficient of $t^i/i!$ in (2.68) gives a recurrence formula for the factorial moments that involves at most $\max(p + 1, q + 2)$ of them.

The generating function for the moments about c satisfies

$$(D + c) \prod_{j=1}^{q} (D + c + b_j - 1)\, [e^{-ct} G(e^t)] = \lambda e^t \prod_{i=1}^{p} (D + c + a_i)\, [e^{-ct} G(e^t)]; \tag{2.69}$$

Table 2.4 Named Probability Distributions with pgf's Expressible in Terms of ${}_pF_q[\lambda z]$

Table 2.4 (*Continued*)

Publisher's Note:
Permission to reproduce this image online was not granted by the copyright holder. Readers are kindly requested to refer to the printed version of this article.

identifying the coefficient of $t^i/i!$ in this equation yields a recurrence formula for the moments about c. The uncorrected and the corrected moments are given by $c = 0$ and $c = \mu'_1 = \mu$, respectively. Kemp (1968a,b) investigated these and other moment properties, including ones for the cumulants, factorial cumulants, and incomplete moments. Over-, under-, and equidispersion aspects of GHPDs have been studied by Tripathi and Gurland (1979).

Three results concerning limiting forms of GHPDs are

$$\lim_{v\to\pm\infty} \frac{{}_{p+1}F_q[a_1, \ldots, a_p, u+v; b_1, \ldots, b_q; \lambda z/v]}{{}_{p+1}F_q[a_1, \ldots, a_p, u+v; b_1, \ldots, b_q; \lambda/v]} = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda]}, \tag{2.70}$$

$$\lim_{v\to\pm\infty} \frac{{}_pF_{q+1}[a_1, \ldots, a_p; b_1, \ldots, b_q, u+v; \lambda v z]}{{}_pF_{q+1}[a_1, \ldots, a_p; b_1, \ldots, b_q, u+v; \lambda v]} = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda]}, \tag{2.71}$$

$$\lim_{v\to\pm\infty} \frac{{}_{p+1}F_{q+1}[a_1, \ldots, a_p, v; b_1, \ldots, b_q, u+kv; \lambda k z]}{{}_{p+1}F_{q+1}[a_1, \ldots, a_p, v; b_1, \ldots, b_q, u+kv; \lambda k]} = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda]}. \tag{2.72}$$

Many limit theorems in the literature can be proved using these results. The Poisson is an ultimate limiting form for distributions with infinite support; the binomial is an ultimate limiting form for terminating ones.

Results concerning gamma-type and beta-type mixtures of GHPDs have been derived by Kemp (1968a) and Tripathi and Gurland (1979); they are described in Section 8.4.

Weighted distributions are distributions that are modified by the method of ascertainment; see Rao (1965), Patil, Rao, and Ratnaparkhi (1986) and

Section 3.12.4. Let the sampling chance (i.e., weighting factor) for a value x be γ_x. Then if the original distribution has pgf $\sum_x p_x z^x$, the ascertained (weighted) distribution has pgf

$$G(z) = \frac{\sum_x \gamma_x p_x z^x}{\sum_x \gamma_x p_x}.$$

If $\gamma_x = \gamma^x$, then for all PSDs the form of the ascertained distribution is the same as the original. For an HP distribution with pgf (2.64), the ascertained pgf is

$$G(z) = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \gamma\lambda z]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \gamma\lambda]}. \tag{2.73}$$

Suppose now that $\gamma_x = x$; then (2.64) is ascertained as

$$G(z) = z\frac{{}_pF_q[a_1+1, \ldots, a_p+1; b_1+1, \ldots, b_q+1; \lambda z]}{{}_pF_q[a_1+1, \ldots, a_p+1; b_1+1, \ldots, b_q+1; \lambda]}. \tag{2.74}$$

More generally, when $\gamma_x = x!/(x-r)!$, (2.64) becomes

$$G(z) = z^r\frac{{}_pF_q[a_1+r, \ldots, a_p+r; b_1+r, \ldots, b_q+r; \lambda z]}{{}_pF_q[a_1+r, \ldots, a_p+r; b_1+r, \ldots, b_q+r; \lambda]}. \tag{2.75}$$

Particular HP distributions can arise from nonequilibrium stochastic processes; for example, the Poisson process yields the Poisson distribution (Section 4.2) and the nonhomogeneous linear growth process yields the negative binomial distribution (Section 5.3).

Consider now the time-homogeneous Markov process with Kolmogorov differential equations

$$\begin{aligned}\frac{dp_0(t)}{dt} &= \beta_1 p_1(t) - \alpha_0 p_0(t),\\ \frac{dp_i(t)}{dt} &= \beta_{i+1}p_{i+1}(t) - (\beta_i + \alpha_i)p_i(t) + \alpha_{i-1}p_{i-1}(t),\end{aligned} \tag{2.76}$$

where $\alpha_i = P(i)/Q(i)$, $\beta_i = R(i)/S(i)$, and $P(i)$, $Q(i)$, $R(i)$, and $S(i)$ are polynomials in i with all real roots such that α_i does not become infinite and β_i does not become zero. Then the equilibrium solution, if it exists, is given by

$$\frac{p_i}{p_{i-1}} = \frac{\alpha_{i-1}}{\beta_i} = \frac{P(i-1)S(i)}{Q(i-1)R(i)}.$$

This can be rewritten as

$$\frac{p_i}{p_{i-1}} = \frac{(i+a_1-1)\cdots(i+a_p-1)\rho}{(i+b_1-1)\cdots(i+b_q-1)},$$

and hence the equilibrium pgf is

$$G(z) = \frac{{}_{p+1}F_q[1, a_1, \ldots, a_p; b_1, \ldots, b_q; \rho z]}{{}_{p+1}F_q[1, a_1, \ldots, a_p; b_1, \ldots, b_q; \rho]}. \tag{2.77}$$

This can be interpreted both in terms of a birth–death process and as a busy-period distribution; see Kemp (1968a) and Kemp and Newton (1990). Kapur (1978a,b) has studied the special case $Q(x) = S(x) = 1$ [see Srivastava and Kashyap (1982)].

Applications of HP distributions to topological features of drainage basins have been made by Dacey (1975).

2.4.2 Generalized Hypergeometric Factorial Moment Distributions

The HF distributions have pgf's of the form

$$G(z) = {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda(z-1)] \tag{2.78}$$

(Kemp, 1968a), and hence their factorial moments are generated by

$$G(1+t) = {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda t]; \tag{2.79}$$

thus

$$\mu'_{[r]} = \frac{(a_1+r-1)! \cdots (a_p+r-1)!(b_1-1)! \cdots (b_q-1)!\lambda^r}{(a_1-1)! \cdots (a_p-1)!(b_1+r-1)! \cdots (b_q+r-1)}. \tag{2.80}$$

Kemp and Kemp (1969b) commented that many common discrete distributions have pgf's of this form and gave a number of examples; see also Gurland (1958), Katti (1966), Kemp (1968a), Kemp and Kemp (1974), and Tripathi and Gurland (1979). In Section 2.4.1 it was possible to give exhaustive conditions under which (2.64) is a pgf. It is more difficult, however, to determine criteria for a set of values to be the factorial moments for a valid discrete distribution [see, e.g., Fréchet (1940)] and hence to state the conditions for (2.79) to be a pgf, though some progress has been made by Kemp (1968a) and Tripathi and Gurland (1979). Interest has centered on the identification of HF distributions rather than their complete enumeration.

Table 2.5 gives lists of terminating and nonterminating GHFDs. See Kemp and Kemp (1974) for constraints on the parameters. Note the overlap with Table 2.4.

Given that $G(1+t) = \sum_{r \geq 0} \mu'_{[r]} t^r / r!$, we have

$$G(z) = \sum_{j \geq 0} \sum_{x=0}^{j} \frac{\mu'_{[j]}(-1)^{j-x} z^x}{x!(j-x)!} = \sum_{x \geq 0} \sum_{k \geq 0} \frac{\mu'_{[x+k]}(-1)^k z^x}{x!k!}$$

Table 2.5 Named HF Distributions[a] with pgf's Expressible as ${}_pF_q[\lambda(z-1)]$

Name	Form	Reference
	Terminating	
Binomial	${}_1F_0[-n;\,;\,p(1-z)]$	
Hypergeometric	${}_2F_1[-n, -Np; -N; 1-z]$	
Negative hypergeometric	${}_2F_1[-n, a; a+b; 1-z]$	
Discrete rectangular	${}_2F_1[-n, 1; 2; 1-z]$	
Chung–Feller	${}_2F_1[-n; 1/2; 1/2; 1-z]$	
Pólya	${}_2F_1[-n, M/c; (M+N)/c; 1-z]$	
Matching distribution	${}_1F_1[-n; -n; z-1]$	
Gumbel's matching distribution	${}_1F_1[-n, -n; \lambda(z-1)]$	Fréchet, 1943
Laplace–Haag	${}_1F_1[-n\,; -N; M(z-1)]$	Fréchet, 1943
Anderson's matching distribution	${}_1F_{N-1}\left[\begin{matrix} -n & ;(-1)^N(z-1) \\ -n, \ldots, -n & \end{matrix}\right]$	Anderson, 1943
Stevens–Craig (Coupon collecting)	${}_nF_{n-1}\left[\begin{matrix} 1-k, \ldots, 1-k;\ 1-z \\ -k, \ldots, -k \end{matrix}\right]$	Patil et al., 1984
STERRED rectangular	${}_3F_2[-n, 1, 1; 2, 2; 1-z]$	Kemp and Kemp, 1969b
	Nonterminating	
Poisson	${}_0F_0[;\,;\,\lambda(z-1)]$	
Negative binomial	${}_1F_0[k;\,;\,p(z-1)]$	
Poisson$\wedge$beta	${}_1F_1[a; a+b; \lambda(z-1)]$	
Type H_2	${}_2F_1[k, a; a+b; \lambda(z-1)]$	Gurland, 1958; Katti, 1966
STERRED geometric	${}_2F_1[1, 1; 2; q(z-1)/(1-q)]$	Kemp and Kemp, 1969b

[a]Kemp and Kemp (1974) give parameter constraints.

(cf. Section 10.2). The probabilities for HF distributions are therefore

$$\Pr[X=x] = \sum_{k\geq 0}\left[\frac{\prod_{i=1}^{p}[(a_i+x+k-1)!/(a_i-1)!]}{\prod_{j=1}^{p}[(b_j+x+k-1)!/(b_j-1)!]}\cdot\frac{\lambda^{x+k}(-1)^k}{x!k!}\right]$$
$$= \frac{\lambda^x\prod_{i=1}^{p}\Gamma(a_i+x)/\Gamma(a_i)}{x!\prod_{j=1}^{q}\Gamma(b_j+x)/\Gamma(b_j)}$$
$$\times\ {}_pF_q[a_1+x, \ldots, a_p+x; b_1+x, \ldots, b_q+x; -\lambda]. \quad (2.81)$$

A simpler method of handling the probabilities is via a recursion formula. The fmgf satisfies the differential equation

$$\theta \prod_{j=1}^{q} (\theta + b_j - 1) G(1+t) = \lambda t \prod_{i=1}^{p} (\theta + a_i) G(1+t), \tag{2.82}$$

where θ is now the differential operator $td/dt \equiv tD$; θ satisfies the symbolic equations

$$t^k D^k \equiv \theta(\theta - 1) \dots (\theta - k + 1) \qquad \text{and} \qquad \theta^k \equiv \sum_{j=1}^{k} S(k, j) t^j D^j,$$

where $S(k, j)$ is a Stirling number of the second kind (see Section 1.1.3). The pgf therefore satisfies

$$\left[\theta \prod_{j=1}^{q} (\theta + b_j - 1)\right]_{t=z-1} G(z) = \lambda(z-1) \left[\prod_{i=1}^{p} (\theta + a_i)\right]_{t=z-1} G(z). \tag{2.83}$$

Restated in terms of the operator D, a factor of $z - 1$ can be removed, and identifying the coefficient of $z^i/i!$ in the resultant equation gives a recurrence formula involving at most $\max(p + 1, q + 2)$ of the probabilities.

Equation (2.83) has the form

$$h_1(z, \theta) G(z) = \lambda h_2(z, \theta) G(z). \tag{2.84}$$

The generating function for moments about c can be obtained by substituting e^t for z and $c + D$ for θ in (2.84); identification of the coefficient of $t^i/i!$ then leads to a recurrence formula for the moments about c, where $c = 0$ and $c = \mu_1' = \mu$ give the uncorrected and central moments, respectively. Examples of the use of these recurrence relations and results for the cumulants and factorial cumulants are in Kemp and Kemp (1974). Tripathi and Gurland (1979) have investigated over-, under-, and equidispersion aspects of GHFDs.

The distribution with pgf

$$G(z) = {}_pF_q[a_1, \dots, a_p; b_1, \dots, b_q; \lambda(z-1)] \tag{2.85}$$

is the limiting form as $v \to \pm\infty$ for those with pgf's

$$\begin{aligned}
G(z) &= {}_{p+1}F_q[a_1, \dots, a_p, u + v; b_1, \dots, b_q; \lambda v^{-1}(z-1)], \\
G(z) &= {}_pF_{q+1}[a_1, \dots, a_p; b_1, \dots, b_q, u + v; \lambda v(z-1)], \\
G(z) &= {}_{p+1}F_{q+1}[a_1, \dots, a_p, v; b_1, \dots, b_q, u + kv; \lambda k(z-1)].
\end{aligned} \tag{2.86}$$

As in the case of HP distributions, the Poisson and binomial are ultimate limiting forms for nonterminating and terminating HF distributions, respectively. For applications to matching and occupancy distributions, see Kemp (1978b).

Contagion models and time-dependent stochastic processes leading to GHFDs are discussed in Kemp (1968a).

Beta and gamma mixtures of GHFDs yield another GHFD, subject to existence conditions; see Section 8.4.

The relationship in inventory decision theory between a demand distribution and its STER distribution was expressed by Patil and Joshi (1968) in the form of an integral; see Section 11.2.13. Kemp and Kemp (1969b) showed that applying Patil and Joshi's result to a demand distribution with pgf (2.78) shifted to support $1, 2, \ldots$ yields another GHFD with support $0, 1, \ldots$ and altered parameters.

In the last section we saw that a weighted GHPD, with sampling chance (weighting factor) γ_x set equal to x, is a GHPD on $1, 2, \ldots$ with the original numerator and denominator parameters increased by unity. A parallel result holds for GHFDs; see Kemp (1974). Suppose, however, that the sampling chance has the form $\gamma_x = \gamma^x$. Then the distribution with pgf (2.78) is ascertained not as a GHFD but as a GHRD with pgf

$$G(z) = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda(\gamma z - 1)]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda(\gamma - 1)]} \tag{2.87}$$

(Kemp, 1974).

Probability generating functions for GHRDs can be stated in the general form

$$G(z) = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda z + \xi]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \lambda + \xi]} \tag{2.88}$$

(see Section 2.4.2); $\xi = 0$ gives a GHFD, and $\xi = -\lambda$ gives a GHPD.

Haight's accident distribution (Haight, 1965) with pgf

$$G(z) = \frac{{}_2F_1[N + b, b; b + 1; (Tz - T - a)/t]}{{}_2F_1[N + b, b; b + 1; -a/t]} \tag{2.89}$$

and Kemp's (1968c) *limited risk* ${}_cP_p$ *distribution* with pgf

$$G(z) = \frac{{}_1F_1[b; b + 1; \psi Tz - \psi T - \psi a]}{{}_1F_1[b; b + 1; -\psi a]} \tag{2.90}$$

both belong to the GHRD family.

2.5 DISTRIBUTIONS BASED ON LAGRANGIAN EXPANSIONS

Distributions belonging to the wide class of *Lagrangian distributions* are based on Lagrangian expansions. Given the transformation $z = ug(z)$, Lagrange

[1736–1813] gave two different expansions for functions of z in terms of w. These lead to two families of Lagrangian distributions, LD_1 and LD_2. They are considered in detail in Section 7.2.

Suppose that $g(z)$ is an analytic function of z such that $g(0) > 0$, $g(1) = 1$, and

$$\left[\left(\frac{\partial}{\partial z}\right)^{x-1} [g(z)]^x\right]_{z=0} \geq 0, \qquad x \geq 2; \tag{2.91}$$

it is not necessary for $g(z)$ to be a pgf as was at first thought, provided that it satisfies these conditions. The smallest nonzero root of the transformation $z = u\, g(z)$ yields a pgf with the following Lagrangian expansion in powers of u:

$$z = G(u) = \sum_{x=1}^{\infty} \frac{u^x}{x!} \left[\left(\frac{\partial}{\partial z}\right)^{x-1} [g(z)]^x\right]_{z=0}. \tag{2.92}$$

The corresponding probabilities are

$$\Pr[X = x] = \frac{1}{x!} \left[\left(\frac{\partial}{\partial z}\right)^{x-1} [g(z)]^x\right]_{z=0}, \qquad x \in \mathbb{Z}^+. \tag{2.93}$$

These distributions are called *basic Lagrangian distributions of the first kind* (BLD_1). They include the *Borel, Consul, Geeta, and Katz distributions*; see Sections 7.2.2, 7.2.3, and 7.2.4.

Distributions that are n-fold convolutions of BLD_1 are called *delta Lagrangian distributions of the first kind* (DLD_1). They are particular cases of GLD_1 with $f(z) = z^n$.

General Lagrangian distributions of the first kind (GLD_1) are formed from two functions, $g(z)$ and $f(z)$. Let $f(z)$ be another analytic function of z such that $0 \leq f(0) < 1$, $f(1) = 1$, and

$$\left[\frac{\partial^{x-1}}{\partial z^{x-1}}\left([g(z)]^x \frac{\partial f(z)}{\partial z}\right)\right]_{z=0} \geq 0, \qquad x \geq 1. \tag{2.94}$$

Then a GLD_1 has the pmf

$$\Pr[X = x] = \begin{cases} f(0), & x = 0, \\ \dfrac{1}{x!}\left[\dfrac{\partial^{x-1}}{\partial z^{x-1}}\left\{[g(z)]^x \dfrac{\partial f(z)}{\partial z}\right\}\right]_{z=0}, & x \in \mathbb{Z}^+. \end{cases} \tag{2.95}$$

If $g(z)$ and $f(z)$ are pgf's, then they satisfy the above conditions. However, they do not need to be pgf's. For instance, $f(z) = (1 - p + pz)^k$, $kp = 1$, $k \in \mathbb{R}^+$, is a suitable function, even when $k \notin \mathbb{Z}^+$.

The GLD_1 have been studied intensively. They include the *Lagrangian Poisson distribution* (generalized Poisson distribution), the *Lagrangian negative binomial distribution* (generalized negative binomial distribution), and the *Lagrangian logarithmic distribution*; these are discussed in Sections 7.2.6, 7.2.7, and 7.2.8.

The *general Lagrangian distributions of the second kind* (GLD_2) are obtained as follows. Suppose that $f(z)$ and $g(z)$ are two pgf's defined on the nonnegative integers and that $g(0) \neq 0$. Then the second Lagrangian expansion gives

$$f(z) = \left(1 - \frac{zg'(z)}{g(z)}\right) \sum_{x \in T} \frac{u^x}{x!} \left[D^x \ \{[g(z)]^x f(z)\}\right]_{z=0}, \tag{2.96}$$

where T is a subset of the nonnegative integers and $D = \partial/\partial z$. Suppose also that if $D^x\{[g(z)]^x f(z)\} \geq 0$ for all $x \in T$ and $g'(1) < 1$, then (Janardan and Rao, 1983)

$$\Pr[X = x] = \begin{cases} \dfrac{1 - g'(1)}{x!} \left[\ D^x\{[g(z)]^x f(z)\}\right]_{z=0} & \text{for } x \in T, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.97}$$

where $D^x = \partial^x/\partial z^x$ and $g'(1) = \partial g(z)/\partial z\big|_{z=1}$.

Janardan (1997) and Consul and Famoye (2001) have widened the class of Lagrangian distributions of the second kind by modifying the restriction that $f(z)$ and $g(z)$ are pgf's.

When $f(z) = z$, the pmf (2.97) becomes the pmf of a *basic Lagrangian distribution of the second kind* (BLD_2). Taking $f(z) = z^n$, it becomes the pmf of an n-fold (BLD_2) and belongs to the family of *delta Lagrangian distributions of the second kind* (DLD_2).

The Lagrangian distributions of the second kind are discussed in more detail in Section 7.2.9.

2.6 GOULD AND ABEL DISTRIBUTIONS

Gould series distributions are generated by expressing a suitable function as a series of Gould polynomials (see below). Suppose that $A(s; r)$ is a positive function of a parameter s that may also depend on a second parameter r; assume also that

$$A(s; r) = \sum_{x=0}^{\infty} a(x; r) s(s + rx - 1)_{x-1}, \tag{2.98}$$

where $(s, r) \in S_0 \times R_0$, $S_0 = \{s; |s| < s_0\}$, $R_0 = \{r; |r| < r_0\}$ and the coefficients $a(x; r)$, $x = 0, 1, 2, \ldots$, are independent of s. Then

$$\Pr[X = x; s, r] = p(x; s, r) = [A(s; r)]^{-1} a(x; r) s(s + rx - 1)_{x-1}, \tag{2.99}$$

$x = 0, 1, 2, \ldots,$ satisfies the properties of a pmf, provided that the parameter space $S_0 \times R_0$ is restricted to $S \times R$ in such a way that the terms of the expansion (2.98) are nonnegative; both $\{a(x; r) \geq 0,\ 0 \leq s < s_0,\ 0 \leq r < r_0\}$ and $\{(-1)^x a(x; r) \geq 0, -s_0 < s \leq 0, -r_0 < r \leq 0\}$ are possibilities. The polynomials

$$\begin{aligned} G_x(s; r) &= s(s + rx - 1)_{x-1} \\ &= s(s + rx - 1)(s + rx - 2) \cdots (s + rx - x + 1), \qquad x = 1, 2, \ldots, \\ G_0(s; r) &= 1 \end{aligned}$$

were introduced by Gould (1962) and were called *Gould polynomials* by Roman and Rota (1978).

The family was developed by Charalambides (1986a), who found that they occur in fluctuations of sums of interchangeable rv's. He showed that they have applications concerning the busy period in queueing processes and the time to emptiness in dam and storage processes. The family includes the generalized binomial and negative binomial distributions of Jain and Consul (1971) and Consul and Shenton (1972); these distributions have the pmf

$$\Pr[X = x] = \frac{s}{s + rx}\binom{s + rx}{x} p^x (1 - p)^{s+rx-x}, \qquad x = 0, 1, 2, \ldots, \tag{2.100}$$

where the parameters satisfy (i) s, r positive integers and $0 < p < \min\{1, 1/r\}$ and (ii) $s, r < 0$ and $1/r < p < 0$, respectively. Other members of the family are the *quasi–hypergeometric I* and *quasi–Pólya I* distributions of Consul (1974) [also Janardan (1975)] with pmf

$$\Pr[X = x] = \binom{n}{x} \frac{s(s + rx - 1)_{x-1}[m + r(n - x)]_{n-x}}{(s + m + rn)_n}, \qquad x = 0, 1, 2, \ldots, n,$$

where s, r, m are positive integers and negative real numbers, respectively, and the *quasi–hypergeometric II* and *quasi–Pólya II* distributions of Consul and Mittal (1975), with pmf

$$\Pr[X = x] = \binom{n}{x} \frac{s(s + rx - 1)_{x-1} m[m + r(n - 1)]_{n-x-1}}{(s + m)(s + m + rn - 1)_{n-1}}, \qquad x = 0, 1, 2, \ldots, n,$$

where again s, r, m are positive integers and negative real numbers, respectively.

Charalambides (1986a) obtained closed-form expressions for the pgf and the factorial moments of such distributions. He also showed that the generalized binomial distribution with pmf (2.100) is the only member of the Gould series family that is closed under convolution.

More recently Charalambides (1990) has explored the use of Abel series. An *Abel series distribution* arises when a positive function $A_b(\theta; \lambda)$ is expanded as

$$A_b(\theta;\lambda) = \sum_{x=0}^{\infty} a_b(x;\lambda)\theta(\theta+\lambda x)^{x-1}, \tag{2.101}$$

where $(\theta, \lambda) \in \Theta_0 \times \Lambda_0$, $\Theta_0 = \{\theta; |\theta| < \rho_1\}$, $\Lambda_0 = \{\lambda; |\lambda| < \rho_2\}$, and the coefficients $a_b(x;\lambda)$, $x = 0, 1, 2, \ldots$, are independent of the parameter θ. Suppose that the parameter space $\Theta_0 \times \Lambda_0$ is restricted to $\Theta \times \Lambda$ in such a way that the terms of (2.101) are all nonnegative; then the pmf

$$\Pr[X = x] = [A_b(\theta;\lambda)]^{-1} a_b(x;\lambda)\theta(\theta+\lambda x)^{x-1}, \qquad x = 0, 1, 2, \ldots, \tag{2.102}$$

is a valid pmf. The "generalized Poisson" distribution of Consul and Jain (1973a,b) is a member of the family; see Section 7.2.6. So are the *quasi-binomial I and II distributions*, with pmf's

$$\Pr[X = x] = (1+\theta)^{-n}\binom{n}{x}(1-\lambda x)^{n-x}\theta(\theta+\lambda x)^{x-1}$$

(Consul, 1974) and

$$\Pr[X = x] = \binom{n}{x}\frac{(1+\lambda n-\lambda x)^{n-x-1}\theta(\theta+\lambda x)^{x-1}}{(1+\theta)(1+\theta+\lambda n)^{n-1}}$$

(Consul and Mittal, 1975), respectively, where in both cases $x = 0, 1, 2, \ldots, n$ and $0 < \theta < \infty$, $0 \le \lambda < 1/n$.

Charalambides (1990) has given new modes of genesis for these distributions and has investigated their occurrence in insurance risk, queueing theory, and dam and storage processes. In his study of their properties he has obtained formulas for their factorial moments and has shown that, within the family of Abel distributions, Consul and Jain's "generalized Poisson" distribution is the only one that is closed under convolution. See also Consul (1990d) concerning properties and applications of quasi-binomial distributions.

Nandi and Dutta (1988) have developed a family of *Bell distributions* based on a wide generalization of the Bell numbers.

Janardan (1988, 1993) and Charalambides (1991) have investigated *generalized Eulerian distributions*. These have pgf's involving generalized Eulerian polynomials; their factorial moments can be stated in terms of noncentral Stirling numbers [for the properties of noncentral Stirling numbers see Charalambides (2002)].

2.7 FACTORIAL SERIES DISTRIBUTIONS

Whereas PSDs are based on expansions of the pgf as a Taylor series and MPSDs often relate to Lagrangian expansions, the family of *factorial series distributions* (FSDs) is based on expansions of the pgf as a factorial series using forward differences of a defining function $A(N)$; see Berg (1974, 1975, 1983a, 1987).

More specifically, let $A(N)$ be a real function of an integer parameter N and suppose that $A(N)$ can be expressed as a factorial series in N with nonnegative coefficients,

$$A(N) \equiv \sum_{x=0}^{N} \binom{N}{x} [\Delta^x A(N)]_{N=0} = \sum_{x=0}^{N} \frac{N!}{(N-x)!} a_x,$$

with $a_x \geq 0$ for $x = 0, 1, \ldots, N, a_x$ not involving N. Then the corresponding FSD has pgf

$$G(z) = \sum_{x=0}^{N} \frac{N! a_x z^x}{(N-x)!} \tag{2.103}$$

and pmf

$$\Pr[X = x] = \binom{N}{x} \frac{\Delta^x A(0)}{A(N)} = \frac{N! a_x}{(N-x)! A(N)}, \qquad x = 0, 1, \ldots, N, \tag{2.104}$$

where $\Delta^x A(0)$ is understood to mean $[\Delta^x A(N)]_{N=0}$.

Berg (1983a) commented that the class of FSDs is the discrete parameter analog of the class of PSDs; he pointed out that the two classes have certain properties in common.

Consideration of the rv $Z = N - X$ leads to the following expression for the factorial moments of Z:

$$E\left[\frac{Z!}{(Z-r)!}\right] = \frac{N!}{(N-r)!} \frac{A(N-r)}{A(N)},$$

and hence the factorial moments of X can be proved to be

$$E\left[\frac{X!}{(X-r)!}\right] = \mu'_{[r]} = \frac{N!}{(N-r)!} \frac{\nabla^r A(N)}{A(N)},$$

where ∇ denotes the backward-difference operator. This gives

$$\begin{aligned} \mu &= N \cdot \frac{A(N) - A(N-1)}{A(N)}, \\ \mu_2 &= N(N-1) \cdot \frac{A(N-2)}{A(N)} + \frac{NA(N-1)}{A(N)} - \left[\frac{NA(N-1)}{A(N)}\right]^2. \end{aligned} \tag{2.105}$$

The genesis of FSDs (Berg, 1983a) can be understood by considering a finite collection of exchangeable events, $E_1, \ldots, E_N$ (where the probability of the occurrence of r events out of N depends only on N and r and not on the order in which the events might occur). Let $\overline{S_{r,N}}$ denote the probability that $N - r$ specific events fail to occur. Then by the inclusion–exclusion principle (see Section 10.2), the probability that exactly Z events do not occur is

$$\Pr[Z = z] = \sum_{j=z}^{N} (-1)^{j-z} \binom{j}{z} \binom{N}{j} \overline{S_{N-j,N}}. \tag{2.106}$$

The probability that exactly x of the events occur is obtained by putting $z = N - x$ in (2.106). If $\overline{S_{r,N}}$ has the property of factorizing into two parts, one depending only on r and the other only on N, then the outcome is a FSD.

Examples of FSDs include the binomial and hypergeometric distributions (Berg, 1974) and the Stevens–Craig and matching distributions (Berg, 1978). Related distributions are the Waring, generalized Waring and Yule distributions, and Marlow's factorial distribution (Marlow, 1965; Berg, 1983a); see the index for references to these distributions elsewhere in this volume.

Inference for FSDs has been studied in a series of papers by Berg; see Berg (1983a). Subject to a restriction on N, a unique MVUE for N can be obtained (Berg, 1974); moment and maximum-likelihood estimation were also discussed and the problem raised as to when a maximum-likelihood solution exists. Estimation theory for FSDs was applied by Berg (1974) to certain types of capture–recapture sampling. The close relationship between estimation for a zero-truncated PSD and estimation for a FSD was investigated by Berg (1975). The concept of a moment distribution was applied to the family of FSDs by Berg (1978).

Generalized factorial series distributions are defined in Berg (1983b) as having pmf's of the form

$$\Pr[X = x] = \frac{(N-m)!}{(M-m-x)!} \frac{a_x}{A(N; n(x))}, \qquad x = 0, 1, \ldots, N-m, \tag{2.107}$$

where m is an integer such that $0 < m < N$. Berg discusses their relationship to *snowball sampling* and to the *Reed–Frost chain binomial model*. The extra parameter m has to do with the starting value of the process, while $n(x)$ specifies the stopping rule of the sampling process.

Deformations of FSDs were studied by Murat and Szynal (2003). For the deformations of MPSDs studied by Murat and Szynal, see Section 2.2.2.

2.8 DISTRIBUTIONS OF ORDER-*k*

Order-k distributions occur in the theory of runs and generalized runs (scans).

Let $X_1, X_2, \ldots$ be a sequence of binary trials each resulting in either a success with probability $p = \Pr[X_i = 1]$ or else with a failure with probability $q = 1 - p = \Pr[X_i = 0]$. Also let T_k be the waiting time until a sequence of k consecutive successes has occurred for the first time, that is,

$$\begin{aligned} T_k &= \min\ n \text{ such that } X_{n-k+1} = X_{n-k+2} = \cdots = X_n = 1 \\ &= \min\ n \text{ such that } \prod_{i=n-k+1}^{n} X_i = 1 \\ &= \min\ n \text{ such that } \sum_{i=n-k+1}^{n} X_i = k. \end{aligned}$$

When the underlying sequence of trials are independently and identically distributed (iid) Bernoulli, then the distribution of T_k is called the geometric distribution of order k (Philippou, Georghiou, and Philippou, 1983).

This distribution can be generalized in many ways. In the classical counting method, as soon as k consecutive successes have occurred, counting starts again from zero; this gives a type I waiting-time distribution. In type II counting, a succession of k or more successes is counted as a single run, irrespective of the total length of the run. In overlapping counting, that is, type III counting, an uninterrupted sequence of $n \geq k$ successes preceded and followed by a failure gives a count of $n - k + 1$ runs of length k. Balakrishnan and Koutras (2002) make a clear distinction between these different methods of counting.

Negative binomial distributions of order k are r-fold convolutions of geometric distributions of order k. For binomial distributions of order k, the assumption is made that there are a fixed number of trials and the number of runs of k successes that it contains is counted.

Other generalizations include the Poisson distribution of order k (obtained as a limiting form of the negative binomial distribution of order k) and compound Poisson distributions of order k. Consideration of Markovian dependence among the binary variables has led to further generalizations. Mixtures of order-k distributions have been investigated by Philippou (1989) and others. Ling (1990) has introduced multiparameter (extended) distributions of order k.

In Chapters 34 and 42 of Johnson, Kotz, and Balakrishnan (1997) multivariate distributions of order k are called multivariate distributions of order s. This is because these authors used the symbol k to denote the dimensionality of a multivariate distribution.

Univariate distributions of order k are studied in detail in Section 10.7.

2.9 q-SERIES DISTRIBUTIONS

Basic hypergeometric series (*q-series*) are neither fundamental nor simple. The term *basic* refers to the introduction of a base q, usually in the form of q-factorials or q-binomial coefficients (also called Gaussian binomial coefficients); Section 1.1.12 gives definitions. Their introduction nearly always increases the complexity of the series.

The q-series arise in many mathematical areas, including combinatorics, number theory, quantum physics, and statistics. Although many combinatorial uses, such as lattice path counting, can be given probabilistic interpretations [notably by Pólya (1970) and by Handa and Mohanty (1980)], the relevant papers occur mainly in the combinatorial literature. Here we are concerned primarily with statistical uses of q-series that are published in the mainstream statistical journals.

The Bailey–Slater (Bailey, 1935; Slater, 1966) notation for q-hypergeometric series was used in the second edition of this book. The influential book by Gasper and Rahman (1990) has, however, firmly established the use of the modification (1.173). It is now in universal use in statistical as well as mathematical publications and is used here in the third edition of this book. It is repeated below for convenience of the reader:

$$
\begin{aligned}
&{}_A\phi_B(a_1, \ldots, a_A; b_1, \ldots, b_B; q, z) \\
&\qquad = \sum_{j=0}^{\infty} \frac{(a_1; q)_j \ldots (a_A; q)_j z^j}{(b_1; q)_j \ldots (b_B; q)_j (q; q)_j} \left[(-1)^j q^{\binom{j}{2}}\right]^{B-A+1} \qquad (2.108)
\end{aligned}
$$

where $(a; q)_0 = 1$, $(a; q)_j = (1 - a) \cdots (1 - aq^{j-1})$, and $|q| < 1$. When $A = B + 1$, as is the case for the Gaussian q-hypergeometric function, the modification $[(-1)^j q^{\binom{j}{2}}]^{B-A+1}$ leaves the function unaltered.

A considerable advantage of the new definition is that

$$
\begin{aligned}
&\lim_{\epsilon \to 0} {}_r\phi_s\left(\frac{a_1}{\epsilon}, a_2, \ldots, a_r; c_1, \ldots, c_s; q, \epsilon z\right) \\
&\quad = {}_{r-1}\phi_s(a_2, \ldots, a_r; c_1, \ldots, c_s; q, a_1 z)
\end{aligned}
$$

and

$$
\begin{aligned}
&\lim_{\epsilon \to 0} {}_r\phi_s\left(a_1, \ldots, a_r; \frac{c_1}{\epsilon}, c_2, \ldots, c_s; q, \frac{z}{\epsilon}\right) \\
&\quad = {}_r\phi_{s-1}\left(a_2, \ldots, a_r; c_2, \ldots, c_s; q, \frac{z}{c_1}\right);
\end{aligned}
$$

these limiting results are analogous to those for generalized hypergeometric series.

A basic hypergeometric series ${}_A\Phi_B[\cdot]$ tends to a generalized hypergeometric series ${}_AF_B[\cdot]$ when $q \to 1$. So, when $q \to 1$, a distribution involving ${}_A\phi_B[\cdot]$ tends to one involving ${}_AF_B[\cdot]$; the ${}_A\phi_B[\cdot]$ distribution is said to be a q-analog of the ${}_A\phi_B[\cdot]$ distribution. Often, however, there is more than one q-analog of a classical discrete distribution.

Univariate q-series distributions are studied in detail in Section 10.8.

CHAPTER 3

Binomial Distribution

3.1 DEFINITION

The *binomial distribution* can be defined, using the binomial expansion

$$(q+p)^n = \sum_{x=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{x=0}^{n} \frac{n!}{k!(n-k!)} p^k q^{n-k},$$

as the distribution of a random variable X for which

$$\Pr[X=x] = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, \ldots, n, \tag{3.1}$$

where $q+p=1$, $p>0$, $q>0$, and n is a positive integer. Occasionally a more general form is used in which the variable X is transformed to $a+bX$, where a and b are real numbers with $b \neq 0$. When $n=1$, the distribution is known as the *Bernoulli distribution*.

The characteristic function (cf) of the binomial distribution is $(1-p+pe^{it})^n$, and the probability generating function (pgf) is

$$\begin{aligned} G(z) &= (1-p+pz)^n = (q+pz)^n \\ &= \frac{{}_1F_0[-n; -; -pz/q]}{{}_1F_0[-n; -; -p/q]} &(3.2) \\ &= {}_1F_0[-n; -; p(1-z)], \qquad 0<p<1. &(3.3) \end{aligned}$$

The mean and variance are

$$\mu = np \qquad \text{and} \qquad \mu_2 = npq. \tag{3.4}$$

The distribution is a power series distribution (PSD) with finite support. From (3.2) it is a generalized hypergeometric probability distribution (GHPD) and from (3.3) it is a generalized hypergeometric factorial moment distribution. It is a member of the exponential family of distributions (when n is known), and it is an Ord and also a Katz distribution; for more details see Section 3.4.

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

3.2 HISTORICAL REMARKS AND GENESIS

If n independent trials are made and in each there is probability p that the outcome E will occur, then the number of trials in which E occurs can be represented by a rv X having the binomial distribution with parameters n, p. This situation occurs when a sample of fixed size n is taken from an *infinite* population where each element in the population has an equal and independent probability p of possession of a specified attribute. The situation also arises when a sample of fixed size n is taken from a *finite* population where each element in the population has an equal and independent probability p of having a specified attribute and elements are sampled independently and sequentially with replacement.

The binomial distribution is one of the oldest to have been the subject of study. The distribution was derived by James Bernoulli (in his treatise *Ars Conjectandi*, published in 1713), for the case $p = r/(r+s)$, where r and s are positive integers. Earlier Pascal had considered the case $p = \frac{1}{2}$. In his *Essay*, published posthumously in 1764, Bayes removed the rational restriction on p by considering the position relative to a randomly rolled ball of a second ball randomly rolled n times. The early history of the distribution is discussed, inter alia, by Boyer (1950), Stigler (1986), Edwards (1987), and Hald (1990).

A remarkable new derivation as the solution of the simple birth-and-emigration process was given by McKendrick (1914). The distribution may also be regarded as the stationary distribution for the Ehrenfest model (Feller, 1957). Haight (1957) has shown that the M/M/1 queue with balking gives rise to the distribution, provided that the arrival rate of the customers when there are n customers in the queue is $\lambda = (N-n)N^{-1}(n+1)^{-1}$ for $n < N$ and zero for $n \geq N$ (N is the maximum queue size).

3.3 MOMENTS

The moment generating function (mgf) is $(q + pe^t)^n$ and the cumulant generating function (cgf) is $n \ln(q + pe^t)$. The factorial cumulant generating function (fcgf) is $n \ln(1 + pt)$, whence $\kappa_{[r]} = n(r-1)!p^r$. The factorial moments can be obtained straightforwardly from the factorial moment generating function (fmgf), which is $(1 + pt)^n$. We have

$$\mu'_{[r]} = \frac{n!p^r}{(n-r)!};$$

that is,

$$\begin{aligned}
\mu'_{[1]} &= \mu = np, \\
\mu'_{[2]} &= n(n-1)p^2, \\
\mu'_{[3]} &= n(n-1)(n-2)p^3, \\
&\vdots
\end{aligned} \tag{3.5}$$

From $\mu'_r = \sum_{j=0}^{r} S(r, j)\mu'_{[j]}$ (see Section 1.2.7), it follows that the rth moment about zero is

$$\mu'_r = E[X^r] = \sum_{j=0}^{r} \frac{S(r, j)n!p^r}{(n-r)!}. \tag{3.6}$$

In particular

$$\mu'_1 = np,$$
$$\mu'_2 = np + n(n-1)p^2,$$
$$\mu'_3 = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3,$$
$$\mu'_4 = np + 7n(n-1)p^2 + 6n(n-1)(n-2)p^3 + n(n-1)(n-2)(n-3)p^4.$$

Hence (or otherwise) the central moments can be obtained. The lower order central moments are

$$\begin{aligned} \mu_2 &= \sigma^2 = npq, \\ \mu_3 &= npq(q-p), \\ \mu_4 &= 3(npq)^2 + npq(1-6pq). \end{aligned} \tag{3.7}$$

The moment ratios $\sqrt{\beta_1}$ and β_2 are

$$\sqrt{\beta_1} = (q-p)(npq)^{-1/2}, \qquad \beta_2 = 3 + (1-6pq)(npq)^{-1}. \tag{3.8}$$

For a fixed value of p (and so of q) the (β_1, β_2) points fall on the straight line

$$\frac{\beta_2 - 3}{\beta_1} = \frac{1-6pq}{(q-p)^2} = 1 - \frac{2pq}{(q-p)^2}.$$

As $n \to \infty$, the points approach the limit (0, 3).

Note that the same straight line is obtained when p is replaced by q. The two distributions are mirror images of each other, so they have identical values of β_2 and the same absolute value of $\sqrt{\beta_1}$. The slope $(\beta_2 - 3)/\beta_1$ is always less than 1. The limit of the ratio as p approaches 0 or 1 is 1. For $p = q = 0.5$ the binomial distribution is symmetrical and $\beta_1 = 0$. For $n = 1$, the point (β_1, β_2) lies on the line $\beta_2 - \beta_1 - 1 = 0$. (Note that for *any* distribution $\beta_2 - \beta_1 - 1 \geq 0$.)

Romanovsky (1923) derived the following recursion formula for the central moments:

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right). \tag{3.9}$$

An analogous relation holds for moments about zero,

$$\mu'_{r+1} = pq\left[\left(\frac{n}{q}\right)\mu'_r + \frac{d\mu'_r}{dp}\right]. \tag{3.10}$$

Kendall (1943) used differentiation of the cf to derive the relationship

$$\mu_r = npq\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_j - p\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_{j+1}. \tag{3.11}$$

A simpler recursion formula holds for the cumulants

$$\kappa_{r+1} = pq\frac{\partial\kappa_r}{\partial p}, \qquad r \geq 1. \tag{3.12}$$

Formula (3.10) also holds for the incomplete moments, defined as

$$\mu'_{j,k} = \sum_{i=k}^{n} i^j\binom{n}{i}p^i q^{n-i}.$$

The mean deviation is

$$\nu_1 = E[|X - np|] = 2n\binom{n-1}{[np]}p^{[np]+1}q^{n-[np]}, \tag{3.13}$$

where $[\cdot]$ denotes the integer part [see Bertrand (1889), Frisch (1924), and Frame (1945)]. Diaconis and Zarbell (1991) have discussed the provenance and import of this formula and other equivalent formulas. They found that ν_1 is an increasing function of n but that ν_1/n is a decreasing function of n. Johnson's (1957) article led to a number of generalizations. Using Stirling's approximation for $n!$,

$$\nu_1 \approx \left(\frac{2npq}{\pi}\right)^{1/2}\left[1 + \frac{(np - [np])(nq - [nq])}{2npq} - \frac{1 - 2pq}{12npq}\right]; \tag{3.14}$$

this shows that the ratio of the mean deviation to the standard deviation approaches the limiting value $(2/\pi)^{1/2} \approx 0.798$ as $n \to \infty$.

Katti (1960) devised an ingenious method for obtaining the absolute moments of general order about m. All inverse moments of the binomial distribution [i.e., $E(X^{-r})$ with $r = 1, 2, \ldots$] are infinite because $\Pr[X = 0] > 0$. Inverse moments of the *positive binomial distribution* (formed by zero truncation) are discussed in Section 3.11.

Direct manipulation of the definition of the *inverse factorial moment*, as in Stancu (1968), yields

$$E[\{(X + r)^{(r)}\}^{-1}] = [(n + r)^{(r)}]^{-1}p^{-r}\left[1 - \sum_{y=0}^{r-1}\binom{n+r}{y}p^y q^{n+r-y}\right]. \tag{3.15}$$

Chao and Strawderman (1972) obtained a general result for $E[(X+a)^{-k}]$ in terms of a k-fold multiple integral of $t^{-1}E[t^{X+a-1}]$, and they applied this to the binomial distribution. A modification of their approach enabled Lepage (1978) to express the inverse ascending factorial moment

$$R_x(a,k) = E[\{(X+a)\cdots(X+a+k-1)\}^{-1}]$$

as a k-fold multiple integral of $E[t^{X+a-1}]$. Cressie et al. (1981) obtained $E[(X+a)^{-k}]$ both as a k-fold multiple integral of $E[e^{-t(X+a)}]$ and also from a single integral of $t^{k-1}E[e^{-t(X+a)}]$. Jones (1987) developed an analogous single-integral result for $R_x(a,k)$, namely,

$$R_x(a,k) = [\Gamma(k)]^{-1}\int_0^1 (1-t)^{k-1}E[t^{X+a-1}]\,dt,$$

and compared the different (though equivalent) expressions that are yielded by the different approaches for $R_x(a,k)$ for the binomial distribution.

3.4 PROPERTIES

The binomial distribution belongs to a number of families of distributions and hence possesses the properties of each of the families.

It is a distribution with finite support. As defined by (3.1), it consists of $n+1$ nonzero probabilities associated with the values $0, 1, 2, \ldots, n$ of the rv X. The ratio

$$\frac{\Pr[X=x+1]}{\Pr[X=x]} = \frac{(n-x)p}{(x+1)q}, \qquad x = 0, 1, \ldots, n-1, \tag{3.16}$$

shows that $\Pr[X=x]$ increases with x so long as $x < np-q = (n+1)p-1$ and decreases with x if $x > np-q$. The distribution is therefore unimodal with the mode occurring at $x = [(n+1)p]$, where $[\cdot]$ denotes the integer part. If $(n+1)p$ is an integer, then there are joint modes at $x = np+p$ and $x = np-q$. When $p < (n+1)^{-1}$, the mode occurs at the origin.

The median is given by the minimum value of k for which

$$\sum_{j=0}^{k}\binom{n}{k}p^k q^{n-k} > \tfrac{1}{2}.$$

Kaas and Buhrman (1980) showed that

$$|\text{mean} - \text{median}| \le \max\{p, 1-p\}.$$

Hamza (1995) has sharpened this to

$$|\text{mean} - \text{median}| < \ln 2$$

when $p < 1-\ln 2$ or $p > \ln 2$.

The distribution is a member of the exponential family of distributions with respect to $p/(1-p)$, since

$$\Pr[X = x] = \exp\left[x \ln\left(\frac{p}{1-p}\right) + \ln\binom{n}{x} + n\ln(1-p)\right].$$

Morris (1982, 1983) has shown that it is one of the six subclasses of the natural exponential family for which the variance is at most a quadratic function of the mean; he used this property to obtain unified results and to gain insight concerning limit laws. Unlike the other five subclasses, however, it is not infinitely divisible (no distribution with finite support can be infinitely divisible).

Because $\Pr[X = x]$ is of the form

$$\frac{b(x)\theta^x}{\eta(\theta)}, \qquad \theta > 0, \qquad x = 0, 1, \ldots, n,$$

where $\theta = p/(1-p)$ (Kosambi, 1949; Noack, 1950), the distribution belongs to the important family of PSDs (see Section 2.2). Patil has investigated these in depth [see, e.g., Patil (1986)]. He has shown that for the binomial distribution

$$\frac{\theta\eta^{(r+1)}(\theta)}{\eta^{(r)}(\theta)} = \mu - pr, \tag{3.17}$$

where $\eta(\theta) = \sum_x b(x)\theta^x$. Integral expressions for the tail probabilities of PSDs were obtained by Joshi (1974, 1975), who thereby demonstrated the duality between the binomial distribution and the beta distribution of the second kind. Indeed

$$\sum_{x=r}^{n} \binom{n}{x} p^x q^{n-x} = I_p(r, n-r+1) = \Pr\left[F \le \frac{\nu_2 p}{\nu_1 q}\right], \tag{3.18}$$

where F is a random variable that has an F distribution with parameters $\nu_1 = 2r$, $\nu_2 = 2(n-r+1)$; see Raiffa and Schlaifer (1961).

Berg (1974, 1983a) has explored the properties of the closely related family of factorial series distributions with

$$\Pr[X = x] = \frac{n^{(x)}c(x)}{x!h(n)},$$

to which the binomial distribution can be seen to belong by taking $c(x) = (p/q)^x$.

Expression (3.16) shows that the binomial distribution belongs to the discrete Pearson system (Katz, 1945, 1965; Ord, 1967b). Tripathi and Gurland (1977) have examined methods for selecting from those distributions having

$$\frac{\Pr[X = x+1]}{\Pr[X = x]} = \frac{A + Bx}{C + Dx + Ex^2}$$

a particular member such as the binomial.

Kemp (1968a,b) has shown that the binomial is a generalized hypergeometric distribution with pgf

$$G(z) = \frac{{}_1F_0[-n;\ ;pz/(p-1)]}{{}_1F_0[-n;\ ;p/(p-1)]}.$$

Moreover, because the ratio of successive factorial moments is $(n-r)p$, the distribution is also a generalized hypergeometric factorial moment distribution with pgf ${}_1F_0[-n;\ ;p(1-z)]$ (Kemp, 1968a; Kemp and Kemp, 1974). Its membership of those families enabled Kemp and Kemp to obtain differential equations and associated difference equations for the pgf and various mgf's, including the generating functions for the incomplete and the absolute moments.

The binomial has an increasing failure rate (Barlow and Proschan, 1965). The Mills ratio for a discrete distribution is defined as

$$\sum_{j \ge x} \Pr[X = j]/\Pr[X = x],$$

and therefore it is the reciprocal of the failure rate. Diaconis and Zarbell (1991) showed that the Mills ratio for the binomial distribution satisfies

$$\frac{x}{n} \le \frac{\sum_{j=x}^{n} \Pr[X = j]}{\Pr[X = x]} \le \frac{x(1-p)}{x-np},$$

provided that $x > np$. The binomial distribution is also a monotone likelihood-ratio distribution. The skewness of the distribution is positive if $p < 0.5$ and is negative if $p > 0.5$. The distribution is symmetrical iff $p = 0.5$.

Denoting $\Pr[X \le c]$ by $L_{n,c}(p)$, Uhlmann (1966) has shown that, for $n \ge 2$,

$$\begin{aligned} L_{n,c}\left(\frac{c}{n-1}\right) &> \frac{1}{2} > L_{n,c}\left(\frac{c+1}{n+1}\right) && \text{for} \quad 0 \le c < \frac{n-1}{2}, \\ L_{n,c}\left(\frac{c}{n-1}\right) &= \frac{1}{2} = L_{n,c}\left(\frac{c+1}{n+1}\right) && \text{for} \quad c = \frac{n-1}{2}, \\ L_{n,c}\left(\frac{c+1}{n+1}\right) &> \frac{1}{2} > L_{n,c}\left(\frac{c}{n-1}\right) && \text{for} \quad \frac{n-1}{2} < c \le n. \end{aligned} \tag{3.19}$$

The distribution of the *standardized binomial variable*

$$X' = \frac{X - np}{\sqrt{npq}}$$

tends to the unit-normal distribution as $n \to \infty$; that is, for any real numbers α, β (with $\alpha < \beta$)

$$\lim_{n\to\infty} \Pr[\alpha < X' < \beta] = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2}\,du. \tag{3.20}$$

This result is known as the *De Moivre–Laplace theorem*. It forms a starting point for a number of approximations in the calculation of binomial probabilities; these will be discussed in Section 3.6.

In Section 3.1 we saw that the pgf of X is $(q + pz)^n$. If X_1, X_2 are independent rv's having binomial distributions with parameters n_1, p and n_2, p, respectively, then the pgf of $X_1 + X_2$ is $(q + pz)^{n_1}(q + pz)^{n_2} = (q + pt)^{n_1+n_2}$. Hence $X_1 + X_2$ has a binomial distribution with parameters $n_1 + n_2$, p. This property is also apparent on interpreting $X_1 + X_2$ as the number of occurrences of an outcome E having constant probability p in each of $n_1 + n_2$ independent trials.

The distribution of X_1, conditional on $X_1 + X_2 = k$, is

$$\begin{aligned}\Pr[X_1 = x|k] &= \frac{\binom{n_1}{x} p^x q^{n_1-x} \binom{n_2}{k-x} p^{k-x} q^{n_2-k+x}}{\binom{n_1+n_2}{k} p^k q^{n_1+n_2-k}} \\ &= \frac{\binom{n_1}{x}\binom{n_2}{k-x}}{\binom{n_1+n_2}{k}}, \end{aligned} \tag{3.21}$$

where $\max(0, k - n_2) \le x \le \min(n_1, k)$. This is a *hypergeometric distribution*; see Chapter 6.

The distribution of the difference $X_1 - X_2$ is

$$\Pr[X_1 - X_2 = x] = \sum_{x_1} \binom{n_1}{x_1}\binom{n_2}{x_1 - x} p^{2x_1-x} q^{n_1+n_2-2x_1+x}, \tag{3.22}$$

where the summation is between the limits $\max(0, x) \le x_1 \le \min(n_1, n_2 + x)$.

When $p = q = 0.5$,

$$\Pr[X_1 - X_2 = x] = \binom{n_1+n_2}{n_2+x} 2^{-n_1-n_2}, \qquad -n_2 \le x \le n_1,$$

so that $X_1 - X_2$ has a binomial distribution of the more general form mentioned in Section 3.1.

From the De Moivre–Laplace theorem and the independence of X_1 and X_2, it follows that the distribution of the standardized difference

$$[X_1 - X_2 - p(n_1 - n_2)][pq(n_1 + n_2)]^{-1/2}$$

tends to the unit-normal distribution as $n_1 \to \infty$, $n_2 \to \infty$ (whatever the ratio n_1/n_2). A similar result also holds when X_1 and X_2 have binomial distributions with parameters n_1, p_1 and n_2, p_2 with $p_1 \neq p_2$; however, the conditional distribution of X_1, given $X_1 + X_2 = x$, is no longer hypergeometric. Its distribution has been studied by Stevens (1951) as well as by Hannan and Harkness (1963), who developed asymptotic normal approximations.

Springer (1979) has examined the distribution of products of discrete independent rv's; he used as an illustration the product of two binomial variables with parameters n_1, p_1 and n_2, p_2, where $n_1 = n_2 = 2$.

3.5 ORDER STATISTICS

As is the case for most discrete distributions, order statistics based on observed values of random variables with a common binomial distribution are not often used. Mention may be made, however, of discussions of binomial order statistics by Gupta (1965), Khatri (1962), and Siotani (1956); see also David (1981). Tables of the cumulative distribution of the smallest and largest order statistic and of the range [in random samples of sizes 1(1)20] are in Gupta (1960b), Siotani and Ozawa (1948), and Gupta and Panchapakesan (1974). These tables can be applied in selecting the largest binomial probability among a set of k, based on k independent series of trials. This problem has been considered by Somerville (1957) and by Sobel and Huyett (1957).

Gupta (1960b) and Gupta and Panchapakesan (1974) have tabulated the mean and variance of the smallest and largest order statistic. Balakrishnan (1986) has given general results for the moments of order statistics from discrete distributions, and he has discussed the use of his results in the case of the binomial distribution.

3.6 APPROXIMATIONS, BOUNDS, AND TRANSFORMATIONS

3.6.1 Approximations

The binomial distribution is of such importance in applied probability and statistics that it is frequently necessary to calculate probabilities based on this distribution. Although the calculation of sums of the form

$$\sum_x \binom{n}{x} p^x q^{n-x}$$

is straightforward, it can be tedious, especially when n and x are large and when there are a large number of terms in the summation. It is not surprising that a great deal of attention and ingenuity have been applied to constructing useful approximations for sums of this kind.

The *normal approximation* to the binomial distribution (based on the De Moivre–Laplace theorem)

$$\Pr[\alpha < (X - np)(npq)^{-1/2} < \beta] \approx \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2}\, du = \Phi(\beta) - \Phi(\alpha) \tag{3.23}$$

has been mentioned in Section 3.4. This is a relatively crude approximation, but it can be useful when n is large. Numerical comparisons have been published in a number of textbooks (e.g., Hald, 1952).

A marked improvement is obtained by the use of a *continuity correction*. The following normal approximation is used widely on account of its simplicity:

$$\Pr[X \le x] \approx \Phi\left(\frac{x + 0.5 - np}{(npq)^{1/2}}\right); \tag{3.24}$$

its accuracy for various values of n and p was assessed by Raff (1956) and by Peizer and Pratt (1968), who used the absolute and the relative error, respectively. Various rules of thumb for its use have been recommended in various standard textbooks. Two such rules of thumb are

1. use when $np(1-p) > 9$ and
2. use when $np > 9$ for $0 < p \le 0.5 \le q$.

Schader and Schmid (1989) carried out a numerical study of these two rules which showed that, judged by the absolute error, rule 1 guarantees increased accuracy at the cost of a larger minimum sample size. Their study also showed that for both rules the value of p strongly influences the error. For fixed n the maximum absolute error is minimized when $p = q = \frac{1}{2}$; it is reasonable to expect this since the normal distribution is symmetrical whereas the binomial distribution is symmetrical only when $p = \frac{1}{2}$. The maximum value that the absolute error can take (over all values of n and p) is $0.140(npq)^{-1/2}$; Schader and Schmid showed that under rule 1 it decreases from $0.0212(npq)^{-1/2}$ to $0.0007(npq)^{-1/2}$ as p increases from 0.01 to 0.5.

Decker and Fitzgibbon (1991) have given a table of inequalities of the form $n^c \ge k$, for different ranges of p and particular values of c and k, that yield specified degrees of error when (3.24) is employed.

Approximation (3.24) can be improved still further by replacing α and β on the right-hand side of (3.23) by

$$\frac{[\alpha\sqrt{npq} + np] - 0.5 - np}{\sqrt{npq}} \quad \text{and} \quad \frac{[\alpha\sqrt{npq} + np] + 0.5 - np}{\sqrt{npq}},$$

respectively, where $[\cdot]$ denotes the integer part. A very similar approximation was given by Laplace (1820).

For *individual* binomial probabilities, the normal approximation with continuity correction gives

$$\Pr[X = x] \approx (2\pi)^{-1/2} \int_{(x-0.5-np)/\sqrt{(npq)}}^{(x+0.5-np)/\sqrt{(npq)}} e^{-u^2/2}\, du. \tag{3.25}$$

A nearly equivalent approximation is

$$\Pr[X = x] \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-np)^2}{npq}\right]; \tag{3.26}$$

see Prohorov (1953) concerning its accuracy.

Peizer and Pratt (1968) and Pratt (1968) developed a normal approximation formula for $\sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j}$ in which the argument of $\Phi(\cdot)$ is

$$\frac{x + \frac{2}{3} - \left(n + \frac{1}{3}\right) p}{\left[\left(n + \frac{1}{6}\right) pq\right]^{1/2}} \\ \times \frac{1}{\delta_x} \left\{ 2 \left[\left(x + \frac{1}{2}\right) \ln \left(\frac{x + \frac{1}{2}}{np} \right) + \left(n - x - \frac{1}{2}\right) \ln \left(\frac{n - x - \frac{1}{2}}{nq} \right) \right] \right\}^{1/2} \tag{3.27}$$

where $\delta_x = \left(x + \frac{1}{2} - np\right) / \sqrt{npq}$. This gives good results that are even better when the multiplier $x + \frac{2}{3} - \left(n + \frac{1}{3}\right) p$ is increased by

$$\tfrac{1}{50} \left[(x+1)^{-1} q - (n-x)^{-1} p + (n+1)^{-1} \left(q - \tfrac{1}{2}\right) \right].$$

With this adjustment, the error is less than 0.1% for $\min(x+1, n-x) \geq 2$.

Cressie (1978) suggested a slightly simpler formula, but it is not as accurate as Peizer and Pratt's improved formula, and the gain in simplicity is slight. Samiuddin and Mallick (1970) used the argument

$$\frac{\left(n - x - \frac{1}{2}\right)\left(x + \frac{1}{2}\right)}{n} \left[\ln \left(\frac{x + \frac{1}{2}}{np} \right) - \ln \left(\frac{n - x - \frac{1}{2}}{nq} \right) \right],$$

which has some points of similarity with Peizer and Pratt's formula. This approximation is considerably simpler but not as accurate.

Borges (1970) found that

$$Y = (pq)^{-1/6} \left(n + \tfrac{1}{3}\right)^{1/2} \int_p^y [t(1-t)]^{-1/3}\, dt, \tag{3.28}$$

where $y = \left(x + \frac{1}{6}\right) / \left(n + \frac{1}{3}\right)$, is approximately unit normally distributed; tables for the necessary beta integral are in Gebhardt (1971). This was compared numerically with other approximations by Gebhardt (1969). Another normal approximation is that of Ghosh (1980).

C. D. Kemp (1986) obtained an approximation for the *modal probability* based on Stirling's expansion (Section 1.1.2) for the factorials in the pmf,

$$\Pr[X = m] \approx \frac{e}{\sqrt{2\pi}} \left(\frac{a}{bc} \right)^{1/2} \\ \exp \left[n \ln \left(\frac{a(1-p)}{c} \right) + m \ln \left(\frac{cp}{b(1-p)} \right) + \frac{1}{12} \left(\frac{1}{a} - \frac{1}{b} - \frac{1}{c} \right) \right. \\ \left. - \frac{1}{360} \left(\frac{1}{a^3} - \frac{1}{b^3} - \frac{1}{c^3} \right) + \frac{1}{1260} \left(\frac{1}{a^5} - \frac{1}{b^5} - \frac{1}{c^5} \right) \right], \tag{3.29}$$

where m is the mode, $a = n + 1$, $b = m + 1$, and $c = n - m + 1$. He reported that it gives at least eight-figure accuracy.

Littlewood (1969) made an exhaustive analysis of binomial sums. He obtained complicated asymptotic formulas for $\ln[\sum_{j=x}^{n}\binom{n}{j}p^j q^{n-j}]$ with uniform bounds of order $O(n^{-3/2})$ for each of the ranges

$$n\left(p+\frac{q}{24}\right) \le x \le n(1-n^{-1/5}p) \qquad \text{and} \qquad n(1-n^{-1/5}) \le x \le n$$

and also for

$$np \le x \le n\left(1-\tfrac{1}{2}q\right).$$

A number of approximations to binomial probabilities are based on the equation

$$\begin{aligned}
\Pr[X \ge x] &= \sum_{j=x}^{n}\binom{n}{j}p^j q^{n-j} \\
&= B(x, n-x+1)^{-1}\int_0^p t^{x-1}(1-t)^{n-x}\,dt \\
&= I_p(x, n-x+1) \qquad (3.30)
\end{aligned}$$

(this formula can be established by integration by parts). Approximation methods can be applied either to the integral or to the incomplete beta function ratio $I_p(x, n-x+1)$.

Bizley (1951) and Jowett (1963) pointed out that since there is an exact correspondence between sums of binomial probabilities and probability integrals for certain central F distributions (see Sections 3.4 and 3.8.3), approximations developed for the one distribution are applicable to the other, provided that the values of the parameters correspond appropriately.

The *Camp–Paulson* approximation (Johnson et al., 1995, Chapter 26) was developed with reference to the F distribution. When applied to the binomial distribution, it gives

$$\Pr[X \le x] \approx (2\pi)^{-1/2}\int_{-\infty}^{Y(3\sqrt{Z})^{-1}} e^{-u^2/2}\,du, \qquad (3.31)$$

where

$$\begin{aligned}
Y &= \left[\frac{(n-x)p}{(x+1)q}\right]^{1/3}\left(9-\frac{1}{n-x}\right)-9+\frac{1}{(x+1)}, \\
Z &= \left[\frac{(n-x)p}{(x+1)q}\right]^{1/3}\left(\frac{1}{n-x}\right)+\frac{1}{(x+1)}.
\end{aligned}$$

The maximum absolute error in this approximation cannot exceed $0.007(npq)^{-1/2}$.

A natural modification of the normal approximation that takes into account asymmetry is to use a Gram–Charlier expansion with one term in addition to the leading (normal) term. The maximum error is now $0.056(npq)^{-1/2}$. It varies with n and p in much the same way as for the normal approximations but is usually substantially smaller (about 50%). The relative advantage, however, depends on x as well as on n and p. For details, see Raff (1956).

If $n \to \infty$ and $p \to 0$ in such a way that $np = \theta$ remains finite and constant, then $\Pr[X = x] \to e^{-\theta}\theta^x/x!$; that is, the limiting form is the Poisson distribution (see Chapter 4). This is the basis for the *Poisson approximation* to the binomial distribution

$$\Pr[X \le x] \approx e^{-np} \sum_{j=0}^{x} \frac{(np)^j}{j!}, \tag{3.32}$$

which has been used widely in inspection sampling. The maximum error is practically independent of n and approaches zero as p approaches zero. Anderson and Samuels (1967) showed that this gives an underestimate if $x \ge np$ and an overestimate if $x \le np/(1 + n^{-1})$. Thus the Poisson approximation tends to overestimate tail probabilities at both tails of the distribution. The absolute error of approximation increases with x for $0 \le x \le (np + 0.5) - \sqrt{(np + 0.25)}$ and decreases with x for $(np + 0.5) + \sqrt{(np + 0.25)} \le x \le n$.

Rules of thumb for the use of (3.32) that have been recommended by various authors are summarized in Decker and Fitzgibbon (1991), together with their own findings regarding levels of accuracy. For practical work they advise the use of the normal approximation (3.24) when $n^{0.31}p \ge 0.47$; for $n^{0.31}p < 0.47$ they advise using the Poisson approximation (3.32).

Simons and Johnson (1971) were able to use a result due to Vervaat (1969) to show that, if $n \to \infty$ and $p \to 0$ with $np = \theta$, then

$$\sum_{j=0}^{\infty} \left| \binom{n}{j} p^j q^{n-j} - \frac{e^{-\theta}\theta^j}{j!} \right| h(j) \to 0 \tag{3.33}$$

for any $h(j)$ for which $\sum_{j=0}^{\infty} \theta^j (j!)^{-1} h(j)$ converges.

Ivchenko (1974) studied the ratio

$$\frac{\sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j}}{\sum_{j=0}^{x} [e^{-np}(np)^j/j!]}.$$

Hald (1967) and Steck (1973) constructed Poisson approximations to cumulative binomial probabilities by seeking solutions in θ to the equation

$$\sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j} = \sum_{j=0}^{x} \frac{e^{-\theta}\theta^j}{j!}.$$

Steck gave bounds for θ, while Hald obtained the approximation

$$\theta \approx \frac{(n - x/2)p}{1 - p/2}.$$

The accuracies of a number of Poisson approximations to the binomial distribution have been studied by Morice and Thionet (1969) and by Gebhardt (1969). Gebhardt used as an index of accuracy the maximum absolute difference between the approximate and the exact cdf's. Romanowska (1978) has made similar comparisons using the sum of the absolute differences between approximate and exact values.

The *Poisson Gram–Charlier approximation* for the cumulative distribution function is

$$\Pr[X \le x] \approx \sum_{j=0}^{x} [P(j, np) + 0.5(j - np)\,\Delta P(j, np)], \tag{3.34}$$

where $P(j, np) = e^{-np}(np)^j/j!$ and the forward-difference operator Δ operates on j.

Kolmogorov's approximation is

$$\Pr[X \le x] \approx \sum_{j=0}^{x} [P(j, np) - 0.5np^2\,\nabla^2 P(j, np)]. \tag{3.35}$$

The next term is $np^3\nabla^3 P(j, np)/3$. It is a form of Gram–Charlier type B expansion. A detailed comparison of these two approximations is given in Dunin-Barkovsky and Smirnov (1955).

Galambos (1973) gave an interesting generalized Poisson approximation theorem. Let $S_x(n)$ denote the sum of the $\binom{n}{x}$ probabilities associated with different sets of x among n events $(E_1, \ldots, E_n)$. Then the conditions $S_1(n) \to a$ and $S_2(n) \to a^2/2$ as $n \to \infty$ are sufficient to ensure that the limiting distribution of the number of events that have occurred is Poissonian with parameter a.

Molenaar (1970a) provided a systematic review of the whole field of approximations among binomial, Poisson, and hypergeometric distributions. This is an important source of detailed information on relative accuracies of various kinds of approximation. For "quick work" his advice is to use

$$\Pr[X \le x] \approx \begin{cases} \Phi(\{4x+3\}^{1/2}q^{1/2} - \{4n-4x-1\}^{1/2}p^{1/2}) & \text{for } 0.05 < p < 0.93, \\ \Phi(\{2x+1\}^{1/2}q^{1/2} - 2\{n-x\}^{1/2}p^{1/2}) & \text{for } p \ge 0.93, \end{cases}$$

or

$$\Pr[X \le x] \approx \sum_{j=0}^{x} \frac{e^{-\lambda}\lambda^j}{j!} \tag{3.36}$$

with $\lambda = (2n - x)p/(2 - p)$ for p "small" (Molenaar suggests $p \le 0.4$ for $n = 3$, $p \le 0.3$ for $n = 30$, and $p \le 0.2$ for $n = 300$).

Wetherill and Köllerström (1979) derived further interesting and useful inequalities among binomial, Poisson, and hypergeometric probabilities, with special reference to their use in the construction of acceptance sampling schemes.

3.6.2 Bounds

In the previous section we gave approximations to various binomial probabilities; in this section we examine bounds. Generally approximations are closer to the true values than bounds. Nevertheless, bounds provide one-sided approximations, and they often give useful limits to the magnitude of an approximation error.

Feller (1945) showed that, if $x \geq (n+1)p$, then

$$\Pr[X = x] \leq \Pr[X = m] \exp\left[-\frac{p[x-(n+1)p+1/2]^2}{2(n+1)pq} + \left[m-(n+1)p+\tfrac{1}{2}\right]^2 \right], \tag{3.37}$$

where m is the integer defined by $(n+1)p - 1 < m \leq (n+1)p$ and

$$\binom{n}{m}\left(\frac{m+1}{n+1}\right)^m \left(1-\frac{m+1}{n+1}\right)^{n-m} \leq \binom{n}{m} p^m q^{n-m} \leq \binom{n}{m}\left(\frac{m}{n}\right)^m \left(1-\frac{m}{n}\right)^{n-m}. \tag{3.38}$$

A number of formulas give bounds on the probability $\Pr[|X/n - p| \geq c]$, where c is some constant, that is, on the probability that the difference between the relative frequency X/n and its expected value p will have an absolute value greater than c; see Uspensky (1937), Lévy (1954), and Okamoto (1958). Kambo and Kotz (1966) and Krafft (1969) discussed these, sharpened Okamoto's bounds, and obtained the following improvement on Lévy's bound:

$$\Pr[|X/n - p| \geq c] < \sqrt{2}(c\sqrt{n})^{-1} \exp\left(-2nc^2 - \tfrac{4}{3}nc^4\right) \tag{3.39}$$

if $p, q \geq \max(4/n, 2c)$ and $n > 2$.

Both upper and lower bounds for $\Pr[X \geq x]$ were obtained by Bahadur (1960). Starting from the hypergeometric series representation

$$\Pr[X \geq x] = \binom{n}{x} p^x q^{n-x} \times q\, {}_2F_1[n+1, 1; x+1; p],$$

he obtained

$$\frac{q(x+1)}{x+1-(n+1)p}\left(1+\frac{npq}{(x-np)^2}\right)^{-1} \leq \frac{\Pr[X \geq x]}{\binom{n}{x} p^x q^{n-x}} \leq \frac{q(x+1)}{x+1-(n+1)p}. \tag{3.40}$$

Slud (1977) developed further inequalities starting with the inequalities

$$\sum_{j=x+1}^{\infty} \frac{e^{-np}(np)^j}{j!} \le \Pr[X \ge x+1] \qquad \text{for } x \le \frac{n^2 p}{n+1}, \tag{3.41}$$

$$\sum_{j=x}^{\infty} \frac{e^{-np}(np)^j}{j!} \ge \max\left[\Pr[X \ge x],\, 1 - \Phi\left(\frac{x-np}{\sqrt{(npq)}}\right)\right] \qquad \text{for } x \ge (np+1), \tag{3.42}$$

and

$$\Pr[X \ge x] \ge \sum_{j=x}^{\infty} \frac{e^{-np}(np)^j}{j!} \ge 1 - \Phi\left(\frac{x-np}{\sqrt{(npq)}}\right) \qquad \text{for } x \le np. \tag{3.43}$$

The second inequality in (3.43) is valid for *all* x.

Prohorov (1953) quoted the following upper bound on the total error for the Poisson approximation to the binomial:

$$\sum_{j=0}^{\infty} \left| \binom{n}{j} p^j q^{n-j} - \frac{e^{-np}(np)^j}{j!} \right| \le \min\{2np^2, 3p\} \tag{3.44}$$

[see Sheu (1984) for a relatively simple proof]. Guzman's (1985) numerical studies suggest that in practice this bound is rather conservative.

We note the following inequalities for the ratio of a binomial to a Poisson probability when the two distributions have the same expected value:

$$e^{np}\left(1 - \frac{x}{n}\right)^x (1-p)^n \le \frac{\binom{n}{x} p^x q^{n-x}}{e^{-np}(np)^x/x!} \le e^{np}(1-p)^{n-x}. \tag{3.45}$$

Neuman's (1966) inequality is

$$\Pr[X \le np] > \frac{1}{2} + \frac{1+q}{3\sqrt{2\pi}}(npq)^{-1/2} - \frac{3q^2+12q+5}{48}(npq)^{-1} - \frac{1+q}{36\sqrt{2\pi}}(npq)^{-3/2}. \tag{3.46}$$

3.6.3 Transformations

Methods of transforming data to satisfy the requirements of the normal linear model generally seek to stabilize the variance, or to normalize the errors, or to remove interactions in order to make effects additive. Transformations are often used in the hope that they will at least partially fulfill more than one objective.

A widely used variance stabilization transformation for the binomial distribution is

$$u(X/n) = \arcsin\sqrt{\frac{X}{n}}. \tag{3.47}$$

Anscombe (1948) showed that replacing X/n by $\left(X + \frac{3}{8}\right) / \left(n + \frac{3}{4}\right)$ gives better variance stabilization; moreover, it produces a rv that is approximately normally distributed with expected value $\arcsin(\sqrt{p})$ and variance $1/(4n)$. Freeman and Tukey (1950) suggested the transformation

$$u\left(\frac{X}{n}\right) = \arcsin\sqrt{\frac{X}{n+1}} + \arcsin\sqrt{\frac{X+1}{n+1}}; \tag{3.48}$$

this leads to the same approximately normal distribution. Tables for applying this transformation were provided by Mosteller and Youtz (1961). For p close to 0.5, Bartlett (1947) suggested the transformation

$$u\left(\frac{X}{n}\right) = \ln\left(\frac{X}{n-X}\right). \tag{3.49}$$

3.7 COMPUTATION, TABLES, AND COMPUTER GENERATION

3.7.1 Computation and Tables

Recursive computation of binomial probabilities is straightforward. Since

$$\Pr[X = n] \gtrless \Pr[X = 0] \quad \text{according as} \quad p \gtrless 0.5,$$

forward recursion from $\Pr[X = 0]$ using

$$\Pr[X = x + 1] = \frac{(n - x)p}{(x + 1)q} \Pr[X = x]$$

is generally recommended for $p \leq 0.5$, and backward recursion from $\Pr[X = n]$ using

$$\Pr[X = x - 1] = \frac{xq}{(n - x + 1)p} \Pr[X = x]$$

for $p > 0.5$. Partial summation of the probabilities then gives the tail probabilities. When all the individual probabilities are required, computation with low overall rounding errors will result when an assumed value is taken for $\Pr[X = x_0]$ and both forward and backward recursion from x_0 are used; the resultant values must then be divided by their sum in order to give the true probabilities. Either the integer part of $n/2$ or an integer close to the mode of the distribution would be a sensible choice for x_0.

If only some of the probabilities are required, then recursion from the mode can be achieved using C. D. Kemp's (1986) very accurate approximation for the modal probability that was given in the previous section. This is the basis of his method for the computer generation of binomial rv's.

There are a number of tables giving values of individual probabilities and sums of these probabilities. Tables of the incomplete beta function ratio (Pearson, 1934) contain values to eight decimal places of

$$\Pr[X \geq k] = I_p(k, n-k+1)$$

for $p = 0.01(0.01)0.99$ and $\max(k, n-k+1) \leq 50$. Other tables are as follows:

Biometrika Tables for Statisticians (Pearson and Hartley, 1976)

Tables of the Binomial Probability Distribution (National Bureau of Standards, 1950)

Binomial Tables [Romig, 1953 (this supplements the tables of the National Bureau of Standards)].

Details concerning these and some other tables of binomial probabilities were given in the first edition of this book.

A method for computing $\Pr[X \geq x]$ (the binomial survival function) was devised by Bowerman and Scheuer (1990). It was designed to avoid underflow and overflow problems and is especially suitable for large n.

Stuart (1963) gave tables from which values of $[pq(n_1^{-1} + n_2^{-1})]^{1/2}$ can be obtained to four decimal places (this is the standard deviation of the difference between two independent binomial proportions with common parameter p).

Nomographs for calculating sums of binomial probabilities have been developed [see, e.g., Larson (1966)]. Such nomographs have also been constructed and labeled for the equivalent problem of calculating values of the incomplete beta function ratio (Hartley and Fitch, 1951).

3.7.2 Computer Generation

If large numbers of rv's are required from a binomial distribution with constant parameters, then the ease of computation of its probabilities coupled with the bounded support for the distribution makes nonspecific methods very attractive.

However, when successive calls to the generator are for random binomial variates with changing parameters, distribution-specific methods become important. A slow but very simple method is to simulate the flip of a biased coin n times and count the number of successes. When $p = 0.5$, it suffices to count the number of 1's in a random uniformly distributed computer word of n bits. For $p \neq 0.5$, the method requires n uniforms per generated binomial variate, making it very slow (recycling uniform random numbers in order to reduce the number required is not generally recommended). An ingenious improvement (the beta, or median, method) was devised by Relles (1972); see also Ahrens and Dieter (1974).

Devroye (1986) gives two interesting waiting-time methods based on the following features of the binomial distribution: First, let $G_1, G_2, \ldots$ be iid geometric rv's with parameter p, and let X be the smallest integer such that $\sum_{i=1}^{X+1} G_i > n$; then X is binomial with parameters n, p. Second, let $E_1, E_2, \ldots$ be iid exponential rv's, and let X be the smallest integer such that

$$\sum_{i=1}^{X+1} \frac{e_i}{n-i+1} > -\ln(1-p);$$

then X is binomial with parameters n, p. Both methods can be decidedly slow because of their requirement for very many uniform random numbers; however, as for the coin-flip method, their computer programs are very short.

The C. D. Kemp (1986) algorithm, based on inversion of the cdf by unstored search from the mode, competes favorably with the Ahrens and Dieter (1974) algorithm. Other unstored search programs are discussed in Kemp's paper.

Acceptance–rejection using a Poisson envelope was proposed by Fishman (1979). Kachitvichyanukul and Schmeiser's (1988) algorithm BTPE is a very fast, intricate composition–acceptance–rejection algorithm. Stadlober's (1991) algorithm is simpler, but not quite so fast; it uses the ratio of two uniforms.

3.8 ESTIMATION

3.8.1 Model Selection

The use of binomial probability paper in exploratory data analysis is described by Hoaglin and Tukey (1985). Binomially distributed data should produce a straight line with slope and intercept that can be interpreted in terms of estimates of the parameters.

Other graphical methods include the following:

1. a plot of the ratio of sample factorial cumulants $\tilde{\kappa}_{[r+1]}/\tilde{\kappa}_{[r]}$ against successive low values of r (Hinz and Gurland, 1967; see also Douglas, 1980);
2. a plot of xf_x/f_{x-1} against successive low values of x, where f_x is the observed frequency of x (Ord, 1967a), see Table 2.2; and
3. marking the position of $(\tilde{\kappa}_3/\tilde{\kappa}_2, \tilde{\kappa}_2/\tilde{\kappa}_1)$ on Ord's (1970) diagram of distributions, where $\tilde{\kappa}_1$, $\tilde{\kappa}_2$, and $\tilde{\kappa}_3$ are the first three sample cumulants.

Further graphical methods have been developed by Gart (1970) and Grimm (1970).

3.8.2 Point Estimation

Usually n is known. The method of moments, maximum likelihood, and minimum χ^2 estimators of p are then all equal to $\overline{x}/n$. This estimator is unbiased. Given k samples of size n, its variance is pq/nk, which is the Cramer–Rao lower bound for unbiased estimators of p; the estimator is in fact the minimum variance unbiased estimator (MVUE) of p. Its expected absolute error has been investigated by Blyth (1980).

An approximately median-unbiased estimator of p is

$$\frac{\overline{x} + \frac{1}{6}}{n + \frac{1}{3}}$$

(Crow, 1975; see also Birnbaum, 1964). A helpful expository account of estimation for p (including Bayesian estimation) is in Chew (1971). A useful summary of results is in Patel, Kapardia, and Owen (1976).

Estimation of certain functions of p (when n is known) has also been investigated. Sometimes an estimate of $\Pr[\alpha < X < \beta]$ is required. The MVUE of this polynomial function of p is

$$\sum_{\alpha<\xi<\beta} \binom{n}{\xi}\binom{n(N-1)}{T-\xi} \Big/ \binom{nN}{T}$$

where N is the number of observations, $T = \sum_{j=1}^{N} x_j$, and ξ takes integer values. From this expression it can be seen that the MVUE of a probability $\Pr[X \in \omega]$, where ω is *any* subset of the integers $\{0, 1, \ldots, n\}$, has the same form with the range of summation $\alpha < \xi < \beta$ replaced by $\xi \in \omega$. Rutemiller (1967) studied the estimator of $\Pr[X = 0]$ in some detail, giving tables of its bias and variance. Pulskamp (1990) showed that the MVUE of $\Pr[X = x]$ is admissible under quadratic loss when $x = 0$ or n, but is inadmissible otherwise. He conjectured that the maximum likelihood estimator (MLE) is always admissible.

Another function of p for which estimators have been constructed is $\min(p, 1 - p)$. A natural estimator to use, given a single observed value x, is $\min(x/n, 1 - x/n)$. The moments of this statistic have been studied by Greenwood and Glasgow (1950) and the cumulative distribution by Sandelius (1952).

Cook, Kerridge, and Pryce (1974) have shown that, given a single observation x, $\psi(x) - \psi(n)$ is a useful estimator of $\ln(p)$; they also showed that

$$\left|E[\psi(x) - \psi(n)] - \ln(p)\right| < \frac{q^{n+1}p}{n+1},$$

where $\psi(y)$ is the derivative of $\ln \Gamma(y)$ (i.e., the psi function of Section 1.1.2). They also obtained an "almost unbiased" estimator of the entropy $p \ln(p)$ and used the estimator to construct an estimator of the entropy for a multinomial distribution.

Unbiased sequential estimation of $1/p$ has been studied by Gupta (1967), Sinha and Sinha (1975), and Sinha and Bose (1985). DeRouen and Mitchell (1974) constructed minimax estimators for linear functions of $p_1, p_2, \ldots, p_r$ corresponding to r different (independent) binomial variables.

Suppose now that $X_1, X_2, \ldots, X_k$ are independent binomial rv's and that X_j has parameters n_j, p, where $j = 1, 2, \ldots, k$. Then, given a sample of k

observations $x_1, \ldots, x_k$, comprising one from each of the k distributions, the maximum–likelihood estimator of p is the overall relative frequency

$$\hat{p} = \frac{\sum_{j=1}^{k} x_j}{\sum_{j=1}^{k} n_j}. \tag{3.50}$$

Moreover $\sum_{j=1}^{k} x_j$ is a sufficient statistic for p. Indeed, since $\sum_{j=1}^{k} X_j$ has a binomial distribution with parameters $\sum_{j=1}^{k} n_j$, p, the analysis is the same as for a single binomial distribution.

The above discussion assumes that $n_1, n_2, \ldots, n_k$ (or at least $\sum_{j=1}^{k} n_j$) are known. The problem of estimating the values of the n_j's was studied by Student (1919), Fisher (1941), Hoel (1947), and Binet (1953); for further historical details see Olkin, Petkau, and Zidek (1981).

Given a single observation of a rv X having a binomial distribution with parameters n, p, then, if p is known, a natural estimator for n is x/p. This is unbiased and has variance nq/p.

The equation for the MLE $\hat{n}$ of n when p is known is

$$\sum_{j=0}^{R-1} A_j(\hat{n} - j)^{-1} = -N \ln(1 - p); \tag{3.51}$$

A_j is the number of observations that exceed j and $R = \max(x_1, \ldots, x_N)$ (Haldane, 1941). When N is large,

$$\sqrt{N}\ \mathrm{Var}(\hat{n}) \approx \left[\sum_{j=1}^{n} \left(\Pr[X = j] \sum_{i=0}^{j-1} (n - i)^{-2} \right) \right]^{-1}. \tag{3.52}$$

The consistency of this estimator was studied by Feldman and Fox (1968).

Dahiya (1981) has constructed a simple graphical method for obtaining the maximum-likelihood estimate of n, incorporating the integer restriction on n. In Dahiya (1986) he examined the estimation of m (an integer) when $p = \theta^m$ and θ is a known constant.

Suppose now that $X_1, X_2, \ldots, X_k$ are independent rv's all having the same binomial distribution with parameters n, p. Then equating the observed and expected first and second moments gives the moment estimators $\tilde{n}$ and $\tilde{p}$ of n and p as the solutions of $\overline{x} = \tilde{n}\tilde{p}$ and $s^2 = \tilde{n}\tilde{p}\tilde{q}$. Hence

$$\tilde{p} = 1 - \frac{s^2}{\overline{x}} \tag{3.53}$$

$$\tilde{n} = \frac{\overline{x}}{\tilde{p}}. \tag{3.54}$$

Note that, if $\overline{x} < s^2$, then $\tilde{n}$ is negative, suggesting that a binomial distribution is an inappropriate model.

Continuing to ignore the limitation that n must be an integer, the MLEs $\hat{n}$, $\hat{p}$ of n, and p satisfy the equations

$$\hat{n}\hat{p} = \overline{x}, \tag{3.55}$$

$$\sum_{j=0}^{R-1} A_j(\hat{n} - j)^{-1} = -N \ln\left(1 - \frac{\overline{x}}{\hat{n}}\right); \tag{3.56}$$

A_j is the number of observations that exceed j and $R = \max(x_1, \ldots, X_N)$. The similarity between (3.54) and (3.55) arises because the binomial distribution is a PSD; see Section 2.2. Unlike the method-of-moments equations, the maximum-likelihood equations require iteration for their solution. DeRiggi (1983) has proved that a maximum-likelihood solution exists iff the sample variance is less than $\overline{x}$ and that, if a solution exists, it is unique.

If N is large,

$$\operatorname{Var}(\hat{n}) \approx \frac{n}{N}\left[\sum_{j=2}^{n}\left(\frac{p}{q}\right)^{j} \frac{(j-1)!(N-j)!}{j(N-1)!}\right]^{-1}, \tag{3.57}$$

and the asymptotic efficiency of $\tilde{n}$, relative to $\hat{n}$, is

$$\left[1 + 2\sum_{j=1}^{n-1}\left(\frac{p}{q}\right)^{i} \frac{j!(N-j-1)!}{(j+1)(N-2)!}\right]^{-1} \tag{3.58}$$

(Fisher, 1941).

Olkin, Petkau, and Zidek (1981) found theoretically and through a Monte Carlo study that, when both n and p are to be estimated, the method of moments and maximum-likelihood estimation both give rise to estimators that can be highly unstable; they suggested more stable alternatives based on (1) ridge stabilization and (2) jackknife stabilization. Blumenthal and Dahiya (1981) also recognized the instability of the MLE of n, both when p is unknown and when p is known; they too gave an alternative stabilized version of maximum-likelihood estimation.

In Carroll and Lombard's (1985) study of the estimation of population sizes for impala and waterbuck, these authors stabilized maximum-likelihood estimation of n by integrating out the nuisance parameter p using a beta distribution with parameters a and b. The need for a stabilized estimator of n has been discussed by Casella (1986).

The idea of minimizing the likelihood as a function of n, with p integrated out, can be interpreted in a Bayesian context. For a helpful description of the principles underlying Bayesian estimation for certain discrete distributions, including the binomial, see Irony (1992). Geisser (1984) has discussed and contrasted ways of choosing a prior distribution for binomial trials.

An early Bayesian treatment of the problem of estimating n is that of Draper and Guttman (1971). For p known they chose as a suitable prior distribution for n the rectangular distribution with pmf $1/k$, $1 \le n \le k$, k some large preselected integer. For p unknown, they again used a rectangular prior for n and, like Carroll and Lombard, adopted a beta prior for p, thus obtaining a marginal distribution for n of the form

$$p(n|x_1, \ldots, x_N) \propto \frac{(Nn - T + b - 1)!}{(Nn + a + b - 1)!} \prod_{j=1}^{N} \frac{n!}{(n - x_j)!},$$

where $\max(x_i) \le n \le k$, $T = \sum_{j=1}^{N} x_i$. Although this does not lead to tractable analytical results, numerical results are straightforward to obtain. Kahn (1987) has considered the tailweight of the marginal distribution for n after integrating out p when n is large; he has shown that the tailweight is determined solely by the prior density on n and p. This led him to recommend caution when adopting specific prior distributions. Hamedani and Walter (1990) have reviewed both Bayesian and non-Bayesian approaches to the estimation of n.

Empirical Bayes methods have been created by Walter and Hamedani (1987) for unknown p and by Hamedani and Walter (1990) for unknown n. In their 1990 paper they used an inversion formula and Poisson–Charlier polynomials to estimate the prior distribution of n; this can then be smoothed if it is thought necessary. Their methods are analogous to the Bayes–empirical Bayes approach of Deely and Lindley (1981). Barry (1990) has developed empirical Bayes methods, with smoothing, for the simultaneous estimation of the parameters p_i for many binomials in both one-way and two-way layouts.

Serbinowska (1996) has considered estimation of the number of changes in the parameter p in a stream of binomial observations.

Research concerning the estimation of the parameter n when p is known has been reviewed and extended by Zou and Wan (2003). Casella and Strawderman (1994) have investigated the simultaneous estimation of $n_1, n_2, \ldots$ for several binomial samples.

Kyriakoussis and Papadopoulos (1993) deal with the Bayesian estimation of p for the zero-truncated binomial distribution.

3.8.3 Confidence Intervals

The binomial distribution is a discrete distribution, and so it is not generally possible to construct a confidence interval for p with an exactly specified confidence coefficient using only a set of observations.

Let $x_1, x_2, \ldots, x_N$ be values of independent random binomial variables with exponent parameters $n_1, n_2, \ldots, n_N$ and common second parameter p. Then approximate $100(1 - \alpha)\%$ limits may be obtained by solving the following equations for p_L and p_U:

$$\sum_{j=T}^{n} \binom{n}{j} (p_L)^j (1 - p_L)^{n-j} = \frac{\alpha}{2}, \tag{3.59}$$

$$\sum_{j=0}^{T} \binom{n}{j} (p_U)^j (1 - p_U)^{n-j} = \frac{\alpha}{2}, \tag{3.60}$$

where

$$n = \sum_{j=1}^{N} n_j \qquad \text{and} \qquad T = \sum_{j=1}^{N} x_j.$$

This is the standard approach of Clopper and Pearson (1934) and Pearson and Hartley (1976).

The values of p_L and p_U depend on T, and the interval (p_L, p_U) is an *approximate* $100(1-\alpha)\%$ *confidence interval* for p. Values of p_L and p_U can be found in the following publications: Mainland (1948), Clark (1953), Crow (1956, 1975), Pachares (1960), and Blyth and Hutchinson (1960). Also Pearson's (1934) tables of the incomplete beta function ratio can be used to solve (3.59), provided that $\max(x, n - x + 1) \leq 50$, and to solve (3.60), provided that $\max(x + 1, n - x) \leq 50$. The identity

$$I_p(x, n - x + 1) = \sum_{j=x}^{n} \binom{n}{j} p^j q^{n-j}$$

is used.

Equation (3.59) can be rewritten in the form

$$I_{p_L}(x, n - x + 1) = \tfrac{1}{2}\alpha.$$

Because the lower $100\beta\%$ point, $F_{v_1,v_2,\beta}$, of the F distribution with v_1, v_2 degrees of freedom satisfies the equation

$$I_c\left(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2\right) = \beta,$$

where $c = \nu_1 F/(\nu_1 + \nu_2 F)$ (Johnson et al., 1995, Chapter 26), it follows that

$$p_L = \frac{\nu_1 F_{\nu_1,\nu_2,\alpha/2}}{\nu_2 + \nu_1 F_{\nu_1,\nu_2,\alpha/2}} \tag{3.61}$$

with $\nu_1 = 2x$, $\nu_2 = 2(n - x + 1)$. There is a similar formula for p_U, with $\alpha/2$ replaced by $1 - \alpha/2$ and $\nu_1 = 2(x + 1)$, $\nu_2 = 2(n - x)$. Tables of percentage points for the F distribution can therefore be used to obtain values of p_L and p_U (Satterthwaite, 1957). Charts from which confidence limits for p can be read off are given in Pearson and Hartley's (1976) *Biometrika Tables.*

A confidence interval does not have to have probability $1 - \alpha$ of covering p and *equal* probabilities $\alpha/2$ and $\alpha/2$ of lying entirely above or entirely below p. Alternative methods for constructing confidence intervals for the binomial parameter p have been examined by Angus and Schafer (1984) and Blyth and Still (1983). The latter authors gave a one-page table of 95 and 99% confidence intervals with certain desirable properties such as monotonicity in x, monotonicity in n, and invariance with respect to the transformation $X \to n - X$ and the induced transformation $p \to 1 - p$.

Approximations to p_L and p_U can be obtained using a normal approximation to the binomial distribution (Section 3.6.1). The required values are the roots of

$$(x - np)^2 = \lambda_{\alpha/2}^2 np(1 - p), \tag{3.62}$$

where

$$(2\pi)^{-1/2} \int_{\lambda_{\alpha/2}}^{\infty} e^{-u^2/2}\,du = \frac{\alpha}{2}.$$

More accurate values can be obtained using continuity corrections. An in-depth comparison of the accuracy of various normal approximations to confidence limits for the binomial parameter p was carried out by Blyth (1986).

Nayatani and Kurahara (1964) considered that the crude, though often used, approximation

$$\frac{x}{n} \pm \lambda_{\alpha/2} \left[\frac{x}{n}\left(1 - \frac{x}{n}\right)\right]^{1/2} \tag{3.63}$$

can be used, provided that $np > 5$, $p \le 0.5$.

Stevens's (1950) method for exact confidence statements for discrete distributions involves the use of a uniform random number u where $0 \le u \le 1$. Let $Y = X + U$. Then

$$\Pr[Y \ge y] = \Pr[X > x] + \Pr[X = x]\Pr[U \ge u], \tag{3.64}$$

yielding an exact lower bound; similarly for an exact upper bound. Kendall and Stuart (1961) have applied the procedure to the binomial distribution.

Patel, Kapardia, and Owen (1976) have quoted formulas for an asymptotic confidence interval for p when n is large and known.

When n trials are conducted and no success is observed, a $100(1 - \alpha)\%$ confidence interval for p is

$$\{0 \le p \le 1 - \alpha^{1/n}\}.$$

Louis (1981) commented that this interval is exactly the $100(1 - \alpha)\%$ Bayesian prediction interval based on a uniform prior distribution for p. Also he reinterpreted $S_n = n(1 - \alpha^{1/n})$ as the "number of successes" in a future experiment of the same size.

The conservativeness of conventional Clopper–Pearson confidence intervals, especially for sample sizes less than 100, is demonstrated very clearly in Figure 1 of Brenner and Quan (1990). These authors have developed an alternative, Bayesian, technique for obtaining confidence limits based on the special beta prior $f(p) = (1 + \beta)p^{\beta}$ (a power function prior distribution) and on the restraint that the length of the interval should be as short as possible. Their exact confidence limits were obtained numerically and are presented graphically.

The sample mean is a complete and sufficient statistic for the nuisance parameter p, and hence approximations applied to the conditional likelihood

$$L(n) = \prod_{i=1}^{N} \binom{n}{x_i} \Big/ \binom{Nn}{x_1 + \cdots + x_N}$$

can be used to find approximate confidence limits for n and to derive an approximate test for $H_0 : n = n_0$ against $H_1 : n > n_0$ (Hoel, 1947). Bain, Engelhardt, and Williams (1990) have applied a partition-generating algorithm to this conditional likelihood; by generating all possible ordered outcomes, they have obtained tables of the confidence limits for n without the use of either approximations or simulation. Comparison with Hoel's approximate bounds showed that the latter are appropriate more widely than had previously been thought.

Zheng and Loh (1995) have used a bootstrap approach for obtaining confidence intervals for the parameter p. Lu (2000) constructed confidence intervals for p using a Bayesian method. Lutsenko and Maloshevskii (2003) have adopted a game-theoretic approach to finding the confidence probability for a fixed-width confidence interval.

Little attention has been given to the problem of deriving confidence limits for n when p is known. Approximate confidence limits (n_L, n_U) for n can be obtained by solving the equations

$$\begin{aligned} \sum_{j=0}^{x} \binom{n_U}{j} p^j (1-p)^{n_U - j} &= \frac{\alpha}{2}, \\ \sum_{j=0}^{x} \binom{n_L}{j} p^j (1-p)^{n_L - j} &= 1 - \frac{\alpha}{2}. \end{aligned} \tag{3.65}$$

Hald and Kousgaard (1967) have given appropriate tables.

Confidence intervals for the difference between two independent binomial parameters, $p_1 - p_2$, have been constructed by Lee, Serachitopol, and Brown (1997), Feigin and Lumelskii (2000), and Zhou, Tsao, and Qin (2004). Nam (1995) has obtained confidence limits for the ratio p_1/p_2.

3.8.4 Model Verification

Goodness-of-fit tests for discrete distributions have not been researched as extensively as those for continuous distributions. Two techniques for the binomial distribution that are very widely used are Pearson's χ^2 goodness-of-fit test and

Fisher's binomial index-of-dispersion test. These are described in detail in, for example, Lloyd (1984). Rayner and Best (1989) have discussed goodness-of-fit tests in general and have provided details of their own smooth goodness-of-fit procedure, given the assumption of a binomial distribution.

Louv and Littell (1986) studied how to combine a number of one-sided tests concerning the parameter p. The need to test $H_0 : p_i = p_{i0}$ for all i against $\mathrm{H}_1 : p_i < p_{i0}$ for at least one i arose in a study concerning equal-employment opportunity. Louv and Littell investigated median significance levels, asymptotic relative efficiencies, and accuracy of null distribution approximations for six different procedures. They made a number of specific recommendations; in particular they advocated the use of the Mantel–Haenszel statistic to detect a consistent pattern of departure from the null hypothesis.

D'Agostino, Chase, and Belanger (1988) examined tests for the equality of p for two independent binomial distributions by generating a complete enumeration of all possible sample configurations for some 660 combinations of p, n_1, and n_2. This enabled them to construct the exact distributions of the test statistics for the following four tests: Pearson's χ^2 with Yates's continuity correction; Fisher's exact test; Pearson's χ^2 without Yates's correction; and the two-independent-samples t-test. The first two tests were found to be extremely conservative. They reached the conclusion that the use of the two-independent-samples t-test should be encouraged.

Oluyede (1994) and Kulkarni and Shah (1995) studied test statistics for the one-sided alternative hypothesis that, given several binomial samples, at least one of the parameters p_i is larger (or smaller) than specified.

Chernoff and Lander (1995) were interested in a problem arising in genetics concerning whether a mixture of two binomial distributions with parameters (k, p) and $(k, 0.5)$ is a simple binomial $(k, 0.5)$; they studied this via the asymptotic distribution of the likelihood statistic.

Rahme and Joseph (1998), Katsis and Toman (1999), and Katsis (2001) have been concerned with optimal sample size determination for binomial experiments. The latter two papers adopt a Bayesian-type methodology.

3.9 CHARACTERIZATIONS

Lukacs (1965) characterized the binomial distribution by the property of constancy of regression of a particular polynomial statistic on the sample mean.

The starting point for a number of other characterization results is the Rao–Rubin theorem (Rao and Rubin, 1964). This supposes that X is a discrete rv with values $0, 1, \ldots$ and that $d(r|n)$ is the probability that the value n of X is reduced by a damage (ruin) process to the value r; the resultant rv, also taking the values $0, 1, \ldots,$ is denoted by Y. Suppose also that

$$\Pr[Y = s|\text{ruined}] = \Pr[Y = s|\text{not ruined}]. \tag{3.66}$$

Then, (i) if

$$d(r|n) = \binom{n}{r}\pi^r(1-\pi)^{n-r}$$

for all n, where π is fixed, it follows that X is a Poisson variable. Moreover, (ii) if X is a Poisson rv with parameter λ and (3.66) is satisfied for all λ, then the ruin process is binomial. This result has considerable practical importance.

Yet other characterizations have stemmed from a general theorem of Patil and Seshadri (1964). A corollary of their theorem is that, if the conditional distribution for X given $X+Y$ is hypergeometric with parameters m and n, then X and Y have binomial distributions with parameters (m, θ) and (n, θ), respectively, where θ is arbitrary.

Kagan, Linnik, and Rao (1973) have developed a number of characterization theorems by considering binomial random walks on a lattice of integer points in the positive quadrant, corresponding to sequential estimation plans with Markovian stopping rules.

Let X be a nonnegative rv with pmf p_x and hazard (failure) rate

$$h_x = p_x \Big/ \sum_{j=x}^{n} p_j.$$

Then Ahmed (1991) has shown that X has a binomial distribution with parameters n and p iff

$$E[X|X \geq x] = np + qxh_x \tag{3.67}$$

and a Poisson distribution with parameter λ iff

$$E[X|X \geq x] = \lambda + xh_x. \tag{3.68}$$

Haight (1972) has tried to unify various isolated characterizations for discrete distributions by the use of Svensson's (1969) theorem; this states that, if X is a rv with support on the nonnegative integers, then there exists a two-dimensional rv (X, Y) with the property that the pgf of $\Pr[X = x|X+Y]$ is of the form $G(z) = (1-p+pz)^n$ iff

$$\pi_{X+Y}(z) = \pi_X\left(\frac{z-1+p}{p}\right),$$

where $\pi_Y(z)$ is the generating function of Y.

3.10 APPLICATIONS

The binomial distribution arises whenever underlying events have two possible outcomes, the chances of which remain constant. The importance of the distribution has extended from its original application in gaming to many other areas.

Its use in genetics arises because the inheritance of biological characteristics depends on genes that occur in pairs; see, for example, Fisher and Mather's (1936) analysis of data on straight versus wavy hair in mice. Another, more

Table 3.1 Comparison of Hypergeometric and Binomial Probabilities

	Hypergeometric			Binomial
x	$M = 50$	$M = 100$	$M = 300$	$(M = \infty)$
0	0.0003	0.0006	0.0008	0.0010
1	0.0050	0.0072	0.0085	0.0098
2	0.0316	0.0380	0.0410	0.0439
3	0.1076	0.1131	0.1153	0.1172
4	0.2181	0.2114	0.2082	0.2051
5	0.2748	0.2593	0.2525	0.2461

recent application in genetics is the study of the number of nucleotides that are in the same state in two DNA sequences (Kaplan and Risko, 1982).

The number of defectives found in random samples of size n from a stable production process is a binomial variable; acceptance sampling is a very important application of the test for the mean of a binomial sample against a hypothetical value.

Seber (1982b) has given a number of instances of the use of the binomial distribution in animal ecology, for example, in mark-recapture estimation of the size of an animal population. Boswell, Ord, and Patil (1979) gave applications in plant ecology. Cox and Snell (1981) provided a range of data analysis examples; the binomial distribution underlies a number of them.

The importance of the binomial distribution in model building is evidenced by Khatri and Patel (1961), Katti (1966), and Douglas (1980). Johnson and Kotz (1977) have given many instances of urn models with Bernoulli trials.

The binomial distribution is the sampling distribution for the test statistic in both the sign test and McNemar's test.

It is also a limiting form for a number of other discrete distributions [see Kemp (1968a,b)], and hence, for suitable values of their parameters, it may be used as an approximation. For instance, it is often taken as an approximation to the hypergeometric distribution. The latter distribution (see Chapter 6) is appropriate for sampling without replacement from a finite population of size M consisting of two types of entities. Table 3.1 indicates the way in which the binomial limit is approached as $M \to \infty$ when the sample size n is 10 and $p = 0.5$.

Although appealing in their simplicity, the assumptions of independence and constant probability for the binomial distribution are not often precisely satisfied. Published critical appraisals of the extent of departure from these assumptions in actual situations are rather rare, though an interesting discussion concerning the applicability of the distribution to mortality data has been provided by Seal (1949a). Nevertheless, the model often gives a sufficiently accurate representation to enable useful inferences to be made. Even when the assumptions are known to be invalid, the binomial model provides a reference mark from which departures can be measured. Distributions related to the binomial by various relaxations of the assumptions are described in Section 3.12.

3.11 TRUNCATED BINOMIAL DISTRIBUTIONS

A *doubly truncated binomial distribution* is formed by omitting both the values of x such that $0 \le x < r_1$ and the values of x such that $n - r_2 < x \le n$ (with $0 < r_1 < n - r_2 < n$). For the resulting distribution

$$\Pr[X = x] = \binom{n}{x} p^x q^{n-x} \Bigg/ \sum_{j=r_1}^{n-r_2} \binom{n}{x} p^x q^{n-x}, \qquad x = r_1, \ldots, n - r_2. \quad (3.69)$$

A *singly truncated binomial distribution* is formed if *only* the values $0, 1, \ldots, r_1 - 1$, where $r_1 \ge 1$, *or* the values $n - r_2 + 1, \ldots, n$, where $r_2 \ge 1$, are omitted.

The distribution formed by omission of only the zero class, giving

$$\Pr[X = x] = \binom{n}{x} \frac{p^x q^{n-x}}{1 - q^n}, \qquad x = 1, 2, \ldots, n, \quad (3.70)$$

is called the *positive binomial* (*zero-truncated binomial*) distribution. (Sometimes, however, the untruncated binomial distribution has been referred to as the positive binomial in order to distinguish it from the negative binomial distribution, for which see Chapter 5.)

The rth moment about zero of a rv having the positive binomial distribution (3.70) is equal to $\mu'_r/(1 - q^n)$, where μ'_r is the rth moment about zero for the untruncated binomial distribution. Hence

$$\mu = \frac{np}{1 - q^n} \qquad \text{and} \qquad \mu_2 = \frac{npq}{1 - q^n} - \frac{n^2 p^2 q^n}{(1 - q^n)^2}. \quad (3.71)$$

It is not possible to obtain a simple expression for the negative moments (Stephan (1945)), though Grab and Savage (1954) noted the approximation

$$E[X^{-1}] \approx (np - q)^{-1},$$

which has two-figure accuracy for $np > 10$. Mendenhall and Lehman (1960) obtained approximate formulas by first approximating to the positive binomial with a (continuous) beta distribution (see Johnson et al., 1995, Chapter 24), making the first and second moments agree. They found that

$$E[X^{-1}] \approx \frac{1 - 2/n}{np - q}, \quad (3.72)$$

$$\operatorname{Var}(X^{-1}) \approx \frac{(1 - 1/n)(1 - 2/n)q}{(np - q)^2(np - q - 1)}. \quad (3.73)$$

Formula (3.72) gives two-figure accuracy for $np > 5$.

Situations in which n is known and it is required to estimate p from data from a doubly truncated binomial distribution are uncommon. An interesting example of a practical application has, nevertheless, been described by Newell (1965).

Finney (1949) showed how to calculate the maximum-likelihood estimator (MLE) of p; this was the method used by Newell for the data of his example.

Shah (1966) gave a method of estimating p using the sample moments calculated from a sample of N observed values $x_1, x_2, \ldots, x_N$ each having the distribution (3.69). The first three moments about zero of (3.69) are

$$\begin{aligned}\mu_1' &= A - B + np,\\ \mu_2' &= r_1 A - (n - r_2 + 1)B + \mu_1'(n-1)p + np,\\ \mu_3' &= r_1^2 A - (n - r_2 + 1)^2 B + \mu_2'(n-2)p + \mu_1'(2n-1)p + np,\end{aligned} \tag{3.74}$$

where

$$A = r_1 q \binom{n}{r_1} p^{r_1} q^{n-r_1} \Big/ \sum_{j=r_1}^{n-r_2} \binom{n}{j} p^j q^{n-j} \tag{3.75}$$

and

$$B = (n - r_2 + 1) q \binom{n}{n - r_2 + 1} p^{n-r_2+1} q^{r_2-1} \Big/ \sum_{j=r_1}^{n-r_2} \binom{n}{j} p^j q^{n-j}. \tag{3.76}$$

The moment estimator of p is obtained by eliminating A and B from these equations and replacing μ_s', $s = 1, 2, 3$, by $\sum_{j=1}^{N} x_j^s/N$. Shah calculated the asymptotic efficiency of this moment estimator relative to the MLE to be over 90% for the case $r_1 = r_2 = 1$. This value seems remarkably high, especially since his method uses the third sample moment.

The zero-truncated (positive) binomial distribution (3.70) is of frequent occurrence in demographic enquiries wherein families are chosen for investigation on the basis of an observed "affected" individual, so that there is at least one such individual in each family in the study. The MLE $\hat{p}$ of p, based on N independent observations from the distribution (3.70), satisfies the first-moment equation

$$\overline{x} = \frac{n\hat{p}}{1 - \hat{q}^n}, \tag{3.77}$$

where $\hat{q} = 1 - \hat{p}$. An alternative estimator proposed by Mantel (1951) is

$$\tilde{p} = \frac{\overline{x} - f_1/N}{n - f_1/N},$$

where f_1/N is the observed relative frequency of unity.

For large samples

$$\mathrm{Var}(\hat{p}) = \frac{pq(1-q^n)^2}{Nn(1-q^n-npq^{n-1})}, \tag{3.78}$$

$$\mathrm{Var}(\tilde{p}) = \frac{pq(1-q^n)(1-2q^{n-1}+npq^{n-1}+q^n)}{Nn(1-q^{n-1})^2}. \tag{3.79}$$

The asymptotic efficiency of $\tilde{p}$ relative to $\hat{p}$ would seem to be at least 95%; see Gart (1968).

When p is large, the effect of truncation is small and $\tilde{p}$ differs little from $\hat{p}$. Thomas and Gart (1971) carried out a thorough study of $\hat{p}$ and $\tilde{p}$, and they concluded that $\tilde{p}$ has comparable accuracy and is less biased than $\hat{p}$. They suggested certain refinements.

If the sample observations come from positive binomial distributions with a common value of p but differing parameters $n_1, \ldots, n_N$, an estimator analogous to $\tilde{p}$ is $(\overline{x} - f_1/N)/(\overline{n} - f_1/N)$, where $\overline{n} = \sum_{j=1}^{N} n_j/N$.

Suppose now that $a \in \mathbb{R}^+$, $\tilde{n} = a$ if a is an integer, and $\tilde{n} = [a+1]$ if a is not an integer (where $[\cdot]$ denotes the integer part). Then a rv X has a continuous parameter binomial distribution if

$$\Pr[X = x] = p_x = C\binom{a}{x}p^x(1-p)^{a-x}, \quad x = 0, 1, \ldots, \tilde{n}, \quad C = \frac{1}{\sum_{x=0}^{\tilde{n}} p_x}.$$

When $a \notin \mathbb{Z}^+$, the moments are not equal to those of the corresponding binomial distribution with $a \in \mathbb{Z}^+$, as has been emphasized by Winkelmann (2000). Explicitly, for $a < 3$ he states that

$$\mu = \begin{cases} \dfrac{ap}{1+(a-1)p}, & 0 < a < 1, \\[2ex] \dfrac{ap[1+(a-2)p]}{1+(a-2)p+(a-2)(a-1)p^2/2}, & 1 < a < 2, \\[2ex] \dfrac{ap[1+(a-3)p+(a-3)(a-2)p^2/2]}{1+(a-3)p+(a-3)(a-2)p^2/2+(a-3)(a-2)(a-1)p^3/6}, & \\ & 2 < a < 3, \end{cases}$$

and in general

$$\mu = ap\left(\frac{\theta^{\tilde{n}-1}(a-\tilde{n}, p)}{\theta^{\tilde{n}}(a-\tilde{n}, p)}\right) > ap, \qquad [a] < a < [a+1],$$

where $\theta^{\tilde{n}}(k, p) = \sum_{i=0}^{\tilde{n}} (k)_i p^i / i!$; see Winkelmann, Signorino, and King (1995) for further details.

3.12 OTHER RELATED DISTRIBUTIONS

3.12.1 Limiting Forms

The central importance of the binomial distribution in statistics is shown by the fact that it is related to a wide variety of standard distributions. We now summarize some of the relationships concerning limiting forms that appear elsewhere in the book. Mixtures of binomial distributions are discussed in the context of mixture distributions in Sections 8.2.4 and 8.3.3. Rao's damage model receives attention in Section 9.2.

The limiting form of the *standardized binomial distribution* as $n \to \infty$ is the *normal distribution* (this is the De Moivre–Laplace theorem; see Sections 3.4 and 3.6.1). If $n \to \infty$ and $p \to 0$ with $np = \theta$ (fixed), the *Poisson distribution* is obtained (Section 4.2.1). The binomial distribution is itself a *limiting form of the hypergeometric distribution* (Section 6.4).

Numerical relationships between the binomial and the Poisson distributions have been studied by a number of authors; see Section 3.6.1. For work on numerical relationships between the binomial and the hypergeometric distributions, see Sections 6.4 and 6.5.

3.12.2 Sums and Differences of Binomial-Type Variables

If $X_i, i = 1, 2, \ldots,$ are independent binomial rv's with parameters (n_i, p), p constant, then $\sum_i X_i$ has a binomial distribution with parameters $(\sum_i n_i, p)$; this is the reproductive property of the binomial distribution.

If X_1 and X_2 are independent binomial rv's with parameters (n_1, p_1) and (n_2, p_2), respectively, then $X = X_1 + X_2$ has the pgf

$$G(z) = (1 - p_1 + p_1 z)^{n_1}(1 - p_2 + p_2 z)^{n_1} \tag{3.80}$$

$$= \left(\frac{1-\theta_1 z}{1-\theta_1}\right)^{n_1}\left(\frac{1-\theta_2 z}{1-\theta_2}\right)^{n_2} \tag{3.81}$$

where $\theta_i = p_i/(p_i - 1)$, $i = 1, 2$.

Ong (1995a) has commented that several physical models give rise to the distribution of $X = X_1 + X_2$. These include (i) the number of busy lines at a given time in the lossless trunking model of Jensen (1948); (ii) the continuous-time Ehrenfest urn model of Karlin and McGregor (1965); (iii) Enns's (1966) machine maintenance problem [as solved by Lee (1967)]; (iv) Lam and Lampard's (1981) drug receptor interaction model for a single drug; (v) the N-capacity queuing system studied by Giorno, Negri, and Nobile (1985); and (vi) the dam model by Phatarfod and Mardia (1973), which used the distribution of X as the conditional distribution for Markovian inputs.

Ong has also provided three further stochastic modes of genesis. First, he proved that, given a Markov chain with binomial transition probability

$$\Pr(y|x) = \binom{N-x}{y}\left(\frac{p}{1-p}\right)^{y}\left(1 - \frac{p}{1-p}\right)^{N-x-y}, \qquad 0 < \frac{p}{1-p} < 1,$$

$N \in \mathbb{N}^+$, then the pgf for the k-step transition probabilities has the form (3.80). Second, he showed that when $X|t$ is a binomial rv with parameters $(m+t, \alpha)$, where t is a binomial rv with parameters (n, p), the unconditional distribution of X has the pgf $G_x(z) = (1-\alpha+\alpha z)^m(1-\alpha p+\alpha pz)^n$ and interpreted this model in a marketing context. Third, he showed that the distribution can be obtained as a thinned stochastic process of the kind studied by Cox and Isham (1980).

There are many equivalent formulas for the probabilities of $X = X_1 + X_2$. Kemp (1979) gave the following expansions of the pgf (3.81) and related them to previous formulas for the probabilities that had been given by McKendrick (1926), Bailey (1964), and Irwin (1963):

$$\begin{aligned}
G(z) &= C\sum_{r\ge 0}\binom{n_1}{r}\,{}_2F_1\left[-n_2, -r; 1+n_1-r; \frac{\theta_2}{\theta_1}\right] \times (-\theta_1 z)^r \\
&= C\sum_{r\ge 0}\binom{n_1}{r}\,{}_2F_1\left[1+n_1+n_2-r, 1+n_1; 1+n_1-r; \frac{\theta_2}{\theta_1}\right] \\
&\qquad \times \left(\frac{\theta_1-\theta_2}{\theta_1}\right)^{1+n_1+n_2}(-\theta_1 z)^r \\
&= C\sum_{r\ge 0}\binom{n_1}{r}\,{}_2F_1\left[-n_2, 1+n_1; 1+n_1-r; \frac{\theta_2}{\theta_2-\theta_1}\right] \\
&\qquad \times \left(\frac{\theta_1-\theta_2}{\theta_1}\right)^{n_2}(-\theta_1 z)^r \\
&= C\sum_{r\ge 0}\binom{n_1}{r}\,{}_2F_1\left[1+n_1+n_2-r, -r; 1+n_1-r; \frac{\theta_2}{\theta_2-\theta_1}\right] \\
&\qquad \times \left(\frac{\theta_1-\theta_2}{\theta_1}\right)^{r}(-\theta_1 z)^r \\
&= C\sum_{r\ge 0}\binom{n_1+n_2}{r}\,{}_2F_1\left[-n_2, -r; -n_1-n_2; \frac{\theta_1-\theta_2}{\theta_1}\right] \times (-\theta_1 z)^r \\
&= C\sum_{r\ge 0}\binom{n_1+n_2}{r}\,{}_2F_1\left[-n_1, -n_1-n_2+r; -n_1-n_2; \frac{\theta_1-\theta_2}{\theta_1}\right] \\
&\qquad \times \left(\frac{\theta_2}{\theta_1}\right)^{r-n_1}(-\theta_1 z)^r,
\end{aligned} \tag{3.82}$$

where $C = (1-\theta_1)^{-n_1}(1-\theta_2)^{-n_2}$, provided that the appropriate hypergeometric series either terminates or converges. Further expansions are obtainable by interchanging the suffixes.

The probabilities for $x \geq 2$ can conveniently be calculated via the recurrence relation

$$(x+1)\Pr[X = x+1] = [(x-n_1)\theta_1 + (x-n_2)\theta_2]\Pr[X = x] + (n_1+n_2-x+1)\theta_1\theta_2 \Pr[X = x-1], \quad (3.83)$$

using

$$\Pr[X=0] = (1-\theta_1)^{-n_1}(1-\theta_2)^{-n_2} = (1-p_1)^{n_1}(1-p_2)^{n_2},$$
$$\Pr[X=1] = \Pr[X=0](-n_1\theta_1 - n_2\theta_2) = \Pr[X=0]\left(\frac{n_1p_1}{1-p_1} + \frac{n_2p_2}{1-p_2}\right). \quad (3.84)$$

The moment properties can be obtained in many ways; one of the simplest is via the fcgf

$$\ln G(1+t) = n_1 \ln(1+p_1t) + n_2 \ln(1+p_2t),$$

giving $\kappa'_{[r]} = (n_1p_1^r + n_2p_2^r)(-1)^{r-1}(r-1)!$. The mean and variance are

$$\mu'_1 = \kappa_{[1]} = n_1p_1 + n_2p_2,$$
$$\mu_2 = \kappa'_{[2]} + \kappa'_{[1]} = n_1p_1(1-p_1) + n_2p_2(1-p_2), \quad (3.85)$$
$$\vdots$$

The sum of several independent nonidentical Bernoulli rv's arises in Poissonian binomial sampling; this is considered in the next section.

The pmf for the distribution of the difference of two independent binomial rv's is

$$\Pr[X_1 - X_2 = x] = \sum_y \binom{n_1}{y}\binom{n_2}{y-x} p_1^y q_1^{n_1-y} p_2^{y-x} q_2^{n_2+x-y}, \quad (3.86)$$

where the summation is between the limits $\max(0, x) \leq y \leq \min(n_1, n_2 + x)$.

The pgf for $X_1 - X_2$ is

$$\left(\frac{1-\theta_1 z}{1-\theta_1}\right)^{n_1}\left(\frac{1-\theta_2/z}{1-\theta_2}\right)^{n_2} = z^{-n_2}\left(\frac{1-\theta_1 z}{1-\theta_1}\right)^{n_1}\left(\frac{1-z/\theta_2}{1-1/\theta_2}\right)^{n_2}. \quad (3.87)$$

This is the pgf for $X_1 + X_3$ shifted to support $-n_2, -n_2+1\ldots, n_1$, where X_3 is a binomial rv with parameters $(n_2, 1-p_2)$. The properties of the distribution of the difference of two independent binomial rv's can therefore be deduced from those of the sum of two binomial rv's.

When $p_1 = p_2 = 0.5$,

$$\Pr[X_1 - X_2 = x] = \binom{n_1 + n_2}{n_2 + x} 2^{-n_1 - n_2}, \qquad -n_2 \le x \le n_1;$$

$X_1 - X_2$ therefore has a binomial distribution of the more general form mentioned in Section 3.1.

The pgf $(1 - p + pz)^m = (1 - \theta z)^m/(1 - \theta)^m$ is only valid when either $0 < p < 1$ (that is, $\theta < 0$) and m is a positive integer (the binomial distribution) or $p < 0$ (that is, $0 < \theta < 1$) and $m < 0$ (the negative binomial distribution; see Chapter 5). No other valid combination of values of p and m is possible. Clearly the sum of two binomials, or of two negative binomials, or of a binomial with a negative binomial produces a valid distribution.

Kemp (1979) also introduced the concept of a binomial pseudovariable, that is, an entity having a nonvalid pgf of the form $(1 - \theta z)^m/(1 - \theta)^m$. She examined conditions under which sums (convolutions) involving Bernoulli or geometric variables with pseudo-Bernoulli variables could give valid distributions and found that there are exactly two possibilities:

(i) $$g_1(z) = \left(\frac{1 - \theta}{1 - \theta z}\right)\left(\frac{1 - \theta^\star z}{1 - \theta^\star}\right), \quad \text{i.e., } G * ({}^P B_1), \tag{3.88}$$

where $0 \le \theta^\star \le \theta < 1$; G denotes a geometric distribution and ${}^P B_1$ denotes a pseudo-Bernoulli distribution with pgf $(1 - \theta^\star z)/(1 - \theta^\star)$ (which is not on its own a valid pgf), and

(ii) $$g_2(z) = \left(\frac{1 - \theta}{1 - \theta z}\right)\left(\frac{1 - \theta^\star}{1 - \theta^\star z}\right), \quad \text{i.e., } G * ({}^P B_3), \tag{3.89}$$

where $0 \le -\theta^\star \le \theta < 1$ and $(1 - \theta^\star)/(1 - \theta^\star z)$ is denoted by ${}^P B_3$ in Kemp's notation and is not on its own a valid pgf.

Kemp proved that in both cases the distribution is infinitely divisible.

Case (i) is the geometric-with-zeroes distribution (zero-inflated geometric distribution) with pgf

$$g_1(z) = (1 - \alpha) + \frac{\alpha(1 - \theta)}{1 - \theta z}, \quad \text{where} \quad 0 < \alpha < 1, \qquad 0 < \theta < 1. \tag{3.90}$$

(When $1 < \alpha < 1/\theta$, the outcome is a zero-deflated geometric distribution; this is a Bernoulli$*$ geometric convolution). Reparameterized, (3.90) becomes Mullahy's (1986) hurdle-geometric distribution with pgf

$$g_1 = p_0 + (1 - p_0)\frac{z(1 - \theta)}{1 - \theta z}, \quad \text{where} \quad p_0 = 1 - \alpha\theta. \tag{3.91}$$

Moreover it is the distribution of the number of individuals in the kth generation for the simple discrete-time branching process with one initial individual (the Bienaymé–Galton–Watson process; see Section 9.1). Here the parameterization is

$$g_1(z) = \frac{kc + (1-kc)z}{1 + kc - kcz}, \qquad kc > 1,$$

and $g_1(g_1(z)) = [2kc + (1-2kc)z]/(1+2kc-2kcz)$. The property that $g(g(z))$ has a similar form to $g(z)$ has led to the name *linear fractional* in some books on stochastic processes. Zheng (1997) calls it the *Möbius* distribution; he uses the parameterization

$$g(z) = \frac{a + (1-a-b)z}{1-bz}, \qquad (a, b) \in (0, 1),$$

and gives further mathematical properties.

Case (i) also arises from the nonhomogeneous birth–death process with one initial individual; the pgf here has a parameterization of the form

$$g_1(z) = \frac{\zeta + (1-\eta)(z-1)}{\zeta + \eta(z-1)}.$$

Other modes of genesis are as a Bernoulli distribution generalized by a geometric or modified geometric distribution, and as a geometric distribution generalized by a geometric or modified geometric distribution. This distribution has a long history in the literature; see inter alia Lotka (1939), Jackson and Nichols (1956), Holgate (1964, 1966), Cane (1967), Gross and Harris (1985), and Rubinovitch (1985).

Kemp found that case (ii) arises as a geometric distribution on the nonnegative even integers, also as a geometric distribution generalized with a shifted Bernoulli distribution, and also from the M/M/1 queue with arrivals in pairs, for which

$$g_2(z) = \frac{(1-\rho)(1-z)}{1 + \rho z^3/2 - (1+\rho/2)z}, \quad \text{where} \quad 0 < \rho < 1.$$

It is also a special case of a class of distributions researched by Steyn (1984); see Section 11.2.17.

There is detailed discussion of convolutions involving a negative binomial and a pseudobinomial rv in Section 11.2.7, under the title Feller-Arley and Gegenbauer distributions.

3.12.3 Poissonian Binomial, Lexian, and Coolidge Schemes

The model for *Poissonian binomial sampling* is sometimes called a *Poisson trials model*. It gives rise to a form of distribution known in the earlier literature [e.g., in Aitken's (1945) useful and concise account] as the *binomial distribution of Poisson*. The distribution has also sometimes been called the *Poisson binomial distribution*, an appellation that leads to confusion with the Poisson–binomial distribution of Section 9.5, with pgf $\exp[\lambda(1-p+pz)^n - \lambda]$.

Poisson (1837) was interested in the problem of n trials with p varying from trial to trial. Consider k independent throws of n dice, where p_{ij} is the probability of success for die number i on throw number j; each toss of each die is then a Bernoulli trial with probability p_{ij}. A binomial model holds when $p_{ij} = p$ (constant) for all i, j. For Poissonian binomial sampling we require that $p_{ij} = p_i$ for all k throws of die number i; we also require that the dice always behave independently and that not all of the p_i are equal. The pgf for the number of successes per throw is no longer $(1 - p + pz)^n$ but instead is

$$G_P(z) = \prod_{i=1}^{n}(q_i + p_i z), \qquad 0 \le p_i \le 1\,\forall i, \qquad q_i = 1 - p_i. \tag{3.92}$$

Because the pgf can be regarded as a convolution of Bernoulli pgf's, the cumulants are the sums of the individual Bernoulli cumulants. In particular

$$\begin{aligned}
\mu &= \sum_{i=1}^{n} p_i = n\overline{p},\\
\mu_2 &= \sum_{i=1}^{n} p_i(1 - p_i) = n\overline{p}(1 - \overline{p}) - \sum_{i=1}^{n}(p_i - \overline{p})^2 = n\overline{p}(1 - \overline{p}) - n\sigma_w^2,\\
\mu_3 &= \sum_{i=1}^{n} p_i(1 - p_i)(1 - 2p_i),\\
&\ \ \vdots
\end{aligned} \tag{3.93}$$

where σ_w^2 is the within-throw variance of the p_i. In an interesting and lucid article Nedelman and Wallenius (1986) have discussed the somewhat surprising result that the variance is less than that for a binomial distribution with parameters n, $\overline{p}$, that is, with the same mean. The variance is greatest when $p_1 = p_2 = \cdots = p_n$ and is least when $\sum_{i=1}^{n} p_i = n\overline{p}$ with some of the p_i equal to 0 and the remainder equal to 1.

As n becomes large in such a way that the largest p_i tends to zero but the sum $\sum_{i=1}^{n} p_i = \theta$ remains constant, the limiting form of the binomial distribution of Poisson is a Poisson distribution with parameter θ (Feller, 1968, p. 282).

Let $Y_1, Y_2, \ldots, Y_n$ be independent Bernoulli rv's with means $\mu_1, \mu_2, \ldots, \mu_n$, respectively, let $S_n = \sum_{i=1}^{n} Y_i$, and let U be a Poisson rv with mean $\lambda \sum_{i=1}^{n} \mu_i$. Then the total variation distance between S_n and U is

$$\begin{aligned}
d(S_n, U) &= \sup_j |\Pr[S_n = j] - \Pr[U = j]|\\
&= \frac{1}{2}\sum_{j=0}^{\infty} |\Pr[S_n = j] - \Pr[U = j]| \le \frac{1 - e^{-\lambda}}{\lambda}\sum_{i=1}^{n} \mu_i^2;
\end{aligned}$$

see Barbour and Hall (1984). Yannaros (1991) has given an analogous result for a random number of independent nonidentical Bernoulli rv's. Consider now partial sums of a sequence of independent nonidentical Bernoulli rv's. Turner, Young, and Seaman (1995) have obtained the following Kolmogorov-type inequality:

$$\Pr\left[\sup_{k\geq m}(\overline{Y}_k - \overline{p}_k) \geq \epsilon\right] \leq e^{-2m\epsilon^2}, \qquad \epsilon \geq 0,$$

where $\overline{Y}_k = \sum_{i=1}^{k} Y_i/k$ and $\overline{p}_k = \sum_{i=1}^{k} p_i/k$.

Let Ψ be a convex function on $[0, n]$, n constant. Then Shaked and Shantikumar (1994) have shown that

$$E\left[\Psi\left(\sum_{i=1}^{n} X_i\right)\right] \leq E\left[\Psi\left(\sum_{i=1}^{n} Y_i\right)\right]$$

where $X_1, X_2, \ldots, X_n$ are independent rv's on [0, 1] and $Y_1, Y_2, \ldots, Y_n$ are independent Bernoulli rv's such that $E[X_i] = E[Y_i] = \mu_i$. Now let Y be a binomial rv with mean $n\mu$, where $n\mu = \sum_{i=1}^{n} \mu_i$. Then León and Perron (2003) have shown that

$$E\left[\Psi\left(\sum_{i=1}^{n} Y_i\right)\right] \leq E[\Psi(Y)]. \tag{3.94}$$

León and Perron's paper also obtains optimal exponential bounds for the probabilities of large deviations of $\sum_{i=1}^{n} Y_i$ from their mean and gives references to earlier work.

Consider now particular instances of Poissonian binomial sampling. Brainerd (1972), in a study of textual type and token counts, showed that for his model I the Markov chain with transition probabilities

$$\begin{aligned} \Pr[X_1 = j+1|X_0 = j] &= 1, \\ \Pr[X_n = j+1|X_{n-1} = j] &= g(n-1) > 0 \quad \text{for } n > 1, \\ \Pr[X_n = j|X_{n-1} = j] &= 1 - g(n-1) \end{aligned} \tag{3.95}$$

leads to Poisson's binomial scheme with pgf

$$G_P(z) = \prod_{i=0}^{n-1}[1 + (z-1)g(i)]; \tag{3.96}$$

when $g(i) = e^{-\alpha i}$,

$$G_P(z) = \prod_{i=0}^{n-1}(1 - e^{-\alpha i} + e^{-\alpha i}z) \tag{3.97}$$

and the p_i are in geometric progression (Gani, 1975).

Thomas and Taub (1975) drew attention to a different problem—the problem of computing rapidly the probability of obtaining x hits from a sequence of n shots when each shot has a different probability of success due to changes in the distance from the target. Danish and Hundley's (1979) technical note gives a computer algorithm for this. Thomas and Taub's (1982) algorithm is recursive and considerably faster. They applied it to the case where the probability of success is assumed to increase linearly with the p_i in arithmetic progression (the nearer the target becomes, the greater the probability that a shot will be effective). Kemp's (1987a) weapon defense model led to the assumption that the p_i are in a two-parameter geometric progression, that is, a *loglinear relationship* of the form

$$\ln p_i = \ln C + (i - 1)\ln Q, \qquad i = 1, 2, \ldots, n. \tag{3.98}$$

The resultant pgf is

$$G_P(z) = \prod_{i=0}^{n-1}[1 + CQ^i(z - 1)], \qquad 0 < C \le 1, \qquad 0 < Q < 1. \tag{3.99}$$

This is a terminating q-series distribution; see Section 10.8.1. Gani's distribution (3.97) is the special case $C = 1$, $Q = e^{-\alpha}$.

A *loglinear-odds relationship* was adopted by A. W. Kemp and C. D. Kemp (1991) in their study of plausible nonbinomial models for Weldon's dice data. Here $p_i = cq^{i-1}/(1 + cq^{i-1})$, that is,

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \ln c + (i - 1)\ln q, \qquad i = 1, 2, \ldots, n, \tag{3.100}$$

and the pgf is

$$G_P(z) = \prod_{i=0}^{n-1}\frac{1 + cq^i z}{1 + cq^i}, \qquad 0 < c, \qquad 0 < q < 1; \tag{3.101}$$

this also is a terminating q-series distribution (Section 10.8.1).

Kemp and Kemp developed an alternative two-parameter Poisson trials model for the dice data based on the hypothesis of one dud die. The pgf now becomes $(1 - P + Pz)(1 - p + pz)^{n-1}$, where p is the probability of success for the fair dice and P is the probability of success for the dud die. An interesting feature of Kemp and Kemp's four sets of calculations for dice data is the closeness of the fits given by the loglinear-odds and one-dud-die models. They commented that "the closeness of the results when fitting these models suggests that other Poisson trial models would give very similar results and that, unless one is interested in the parameters $\{p_i\}$ per se, the extra effort involved in estimation for more complicated Poisson trials models is unlikely to be worthwhile" (Kemp and Kemp, 1991, p. 222).

The *Lexian sampling scheme* is a different variant of binomial sampling. It was introduced by Lexis (1877), who was dissatisfied with the then commonly held and often erroneous assumption of homogeneity in sampling. In the context of dice throwing, let p_{ij} again be the probability of success for die i on throw j, but now assume that $p_{ij} = p_j$ (constant) for all n dice on throw j, with independence between all of the tosses. This corresponds to the assumption that for any particular throw the dice all have the same degree of bias, with the degree of bias varying from throw to throw. The pgf for the number of successes per throw is now

$$G_L(z) = \sum_{j=1}^{k} \frac{(1 - p_j + p_j z)^n}{k} \tag{3.102}$$

with mean

$$\mu = \sum_{j=1}^{k} \frac{np_j}{k} = n\overline{p} \tag{3.103}$$

and second factorial moment $\sum_{j=1}^{k} n(n-1)p_j^2/k$ (obtained by differentiating the pgf). Hence the variance is

$$\begin{aligned} \mu_2 &= \sum_{j=1}^{k} \frac{n(n-1)p_j^2}{k} + n\overline{p} - n^2\overline{p}^2 \\ &= n\overline{p}(1-\overline{p}) + n(n-1)\sum_{j=1}^{k} \frac{n(n-1)(p_j - \overline{p})^2}{k} \\ &= n\overline{p}(1-\overline{p}) + n(n-1)\sigma_b^2, \end{aligned} \tag{3.104}$$

where σ_b^2 is the between-throws variance of the p_{ij} (see, e.g., Aitken, 1945). The variance now exceeds the variance of a binomial variable with parameters $n, \overline{p}$; the excess increases with $n(n-1)$. The outcome of the model is of course a *mixed binomial distribution*; see Section 8.2.6 for situations where the p_j for the different throws are predetermined and Section 8.3.4 for situations where the p_j can be regarded as a sample from some population. Stuart and Ord (1987, pp. 165, 171) have commented that the Lexian model is also equivalent to a special case of cluster sampling.

For the *Coolidge scheme* the p_{ij} are assumed to vary both within and between throws. The pgf for the number of successes per throw is now

$$G_C(z) = k^{-1} \sum_{j=1}^{k} \prod_{i=1}^{n} (1 - p_{ij} + p_{ij} z); \tag{3.105}$$

the mean is

$$\mu = \sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij} k^{-1}, \tag{3.106}$$

and the variance is

$$\mu_2 = n\overline{p}(1-\overline{p}) + n^2 \sum_{j=1}^{k} (p_j - \overline{p})^2 - \sum_{i=1}^{n} \sum_{j=1}^{k} (p_{ij} - \overline{p})^2 k^{-1}, \tag{3.107}$$

which will often be very similar to the variance for the Lexian scheme; see Aitken (1945, pp. 53–54). The difficulty with this model is the impossibility of estimating the excessively large number of parameters from a data set.

Ottestad (1943) seems to have been unaware of the work of Coolidge (1921) and Aitken (1939) and gave the name *Poisson–Lexis* to the Coolidge scheme. He referred to previous work by Charlier (1920) and Arne Fisher (1936). For all three situations (Poisson, Lexis, and Poisson–Lexis) he examined the cases (1) n fixed from sample to sample and (2) n varying from sample to sample.

The Lexis ratio provides a test for the homogeneity of binomial sampling; Zabell (1983) provides relevant references.

3.12.4 Weighted Binomial Distributions

Weighted distributions are ones that are modified either by the method of ascertainment or by the method of recording; when an event occurs, it may not necessarily be included in the recorded sample. Let X be a rv with pmf p_x and suppose that when the event $X = x$ occurs the probability that it is recorded is $w(x)$. Then the pmf of the *recorded* distribution is

$$\Pr[X_w = x] = \frac{w(x)p_x}{\sum_x w(x)p_x}. \tag{3.108}$$

If $\sum_x w(x)p_x z^x = E[w(x)z^x]$, then the pgf of the recorded distribution is

$$G(z) = \frac{E[w(x)z^x]}{E[w(x)]}. \tag{3.109}$$

The sampling chance $w(x)$ is called the weighting factor. The concept of a distribution weighted in this manner has been developed by Rao. Two important papers are Rao (1965) on the earlier theory of weighted distributions and Rao's (1985) comprehensive review of their applications. Patil, Rao, and Zelen (1986) have produced an extensive computerized bibliography.

When $w(x) = x$, the recorded variable, X^*, is said to be *size biased*; when $w(x) = x^\alpha$, $\alpha > 0$, it is said to be *size biased of order* α and is written $X^{*\alpha}$. Patil, Rao, and Ratnaparkhi (1986) suggested 10 types of weight functions thought to be useful in scientific work:

(i) $w(x) = x^\alpha, \quad 0 < \alpha \le 1,$
(ii) $w(x) = [x(x-1)/2]^\alpha,$
(iii) $w(x) = x(x-1)\cdots(x-\alpha+1),$
(iv) $w(x) = e^{ax},$
(v) $w(x) = \alpha x + \beta,$
(vi) $w(x) = 1-(1-\beta)^x, \quad 0 < \beta \le 1,$
(vii) $w(x) = (\alpha x + \beta)/(\delta x + \gamma),$
(viii) $w(x) = \Pr[Y \le x]$ for some rv Y,
(ix) $w(x) = \Pr[Y > x]$ for some rv Y,
(x) $w(x) = r(x)$, where $r(x)$ is the probability of "survival" of observation x.

Patil and Rao (1978) showed that, if $B(n, p)$, $P(\lambda)$, and $NB(k, p)$ denote the binomial, Poisson, and negative binomial distributions, then their size-biased forms are $1 + B(n-1, p)$, $1 + P(\lambda)$, and $1 + NB(k+1, p)$, respectively, and gave similar results for other distributions. A number of further results, with particular emphasis on models, mixtures of distributions, and form invariance under size-biased sampling, with appropriate references, are in Patil and Rao (1978) and Patil, Rao, and Ratnaparkhi (1986); see also Patil, Rao, and Zelen (1988).

Kocherlakota and Kocherlakota (1990) have concentrated on weighted binomial distributions. Let $B_n(x; p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$; then

$$G(z) = \frac{\sum w(x) B_n(x; p) z^x}{\sum w(x) B_n(x; p)} \tag{3.110}$$

denotes the pgf of a weighted binomial distribution. We have

$$\frac{\partial^r}{\partial z^r} G(z) = \frac{n! p^r}{(n-r)!} \frac{\sum w(x+r) B_{n-r}(x; p) z^x}{\sum w(x+r) B_{n-r}(x; p)}, \tag{3.111}$$

and setting $z = 1$ in this expression gives the rth factorial moment. Moreover the only pgf satisfying

$$\frac{\partial^r}{\partial z^r} G(z) = \frac{n! p^r}{(n-r)!} \frac{\sum w(x+r) \Pr[X = x] z^x}{\sum w(x+r) \Pr[X = x]} \tag{3.112}$$

is $G(z) = (1 - p + pz)^n$; that is, the relationship (3.111) characterizes the binomial distribution. Kocherlakota and Kocherlakota (1990) have studied estimation for weighted binomial distributions, both when $w(x)$ is specified and when $w(x) = x^\alpha$, α unknown, and have investigated tests of hypotheses when $w(x) = x^\alpha$ for changes in α and also for changes in p. They have also examined goodness-of-fit tests and tests for competing models, with numerical illustrations.

In Kocherlakota and Kocherlakota (1993) they turned their attention to the weighted beta–binomial distribution.

3.12.5 Chain Binomial Models

Consider the spread of a disease in a household of size n in discrete time. Suppose that at time t there are S_t individuals at risk and I_t infectives. If at time t each susceptible individual has an equal and independent chance p_t of catching the disease, then at time $t+1$

$$\Pr[I_{t+1} = i_{t+1}; s_t, p_t] = \binom{s_t}{i_{t+1}} p_t^{i_{t+1}} (1-p_t)^{s_i - i_{t+1}}, \qquad s_t \geq i_{t+1}. \tag{3.113}$$

It is often reasonable to assume that an infected individual is infective for exactly one time interval and that recovery infers immunity, leading to a chain of binomial probabilities.

The Reed–Frost model assumes that the risk of infection depends on the number of contacts between a susceptible and the infectives. Let ρ be the probability of a contact between an infective and a susceptible; let γ be the probability that the result of the contact is an infection. Then the probability that there is no infection due to any specified infective is $(1-\rho) + \rho(1-\beta) = 1 - \rho\beta = \alpha$, say,

$$p_t = 1 - (1-\alpha)^{i_t}. \tag{3.114}$$

The model was taught at John Hopkins University around 1930 and is explained clearly by many authors, including Frost (1976), Bailey (1975), and Daley and Gani (1999).

If the disease is spread by aerosol infection, then p_t may not be very dependent on the number of infectives. An alternative model due to Greenwood (1931) is

$$p_t = \begin{cases} \alpha & \text{if } i_t \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.115}$$

So for households of size 4 with one initial infective, that is, $S_0 = 3$, $I_0 = 1$, the possible chains of infection are as in Table 3.2.

Daley and Gani (1999) discuss, with references, several other (often more realistic) variants of these two basic models and give full accounts of methods of analysis for data on chains of infection. The entries by Longini (1998) and Farrington (1998) in *The Encyclopedia of Biostatistics* also give useful leads to the literature on chain binomial models.

3.12.6 Correlated Binomial Variables

Many data sets concerning the sex of siblings (sibs) within families exhibit underdispersion compared with the binomial distribution. Correlation between the sexes of adjacent sibs has been reported by a number of authors (see, e.g., Greenberg and White, 1965), and consequently Markov chain models that lead to underdispersion (also overdispersion) have been proposed. Edwards (1960) suggested a

Table 3.2 Chains of Infection When $S_0 = 3$, $I_0 = 1$, and $\beta = 1 - \alpha$

Chain $\{i_0, i_1, \ldots\}$	Reed–Frost Probability	Greenwood Probability	Total Number Infected
$\{1\}$	α^3	α^3	1
$\{1, 1\}$	$3\beta\alpha^4$	$3\beta\alpha^4$	2
$\{1, 1, 1\}$	$6\beta^2\alpha^4$	$6\beta^2\alpha^4$	3
$\{1, 2\}$	$3\beta^2\alpha^3$	$3\beta^2\alpha^2$	3
$\{1, 1, 1, 1\}$	$6\beta^3\alpha^3$	$6\beta^3\alpha^3$	4
$\{1, 1, 2\}$	$3\beta^3\alpha^2$	$3\beta^3\alpha^2$	4
$\{1, 2, 1\}$	$3\beta^3\alpha(1 + \alpha)$	$3\beta^2\alpha$	4
$\{1, 3\}$	β^3	β^3	4

Markov chain with $p_{mm} = p + rq$, $p_{mf} = q - rq$, $p_{fm} = p - rp$, $p_{ff} = q + p$, where p_{mf} is the transition probability for a boy followed by a girl. The correlation between successive births is then r, and the number of boys in n successive (single) births has the pgf

$$G(z) = \begin{bmatrix} pz, & q \end{bmatrix} \begin{bmatrix} (p + rq)z, & q - rq \\ (p - rp)z, & q + rp \end{bmatrix}^{n-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{3.116}$$

Other types of models for sibling data that lead to underdispersion are discussed, with references, in Brooks, James, and Gray (1991).

Consider now the sum of k identically distributed but not independent rv's. Suppose that they have a symmetric joint distribution with no second- or higher-order "interactions." Two possibilities have been suggested, depending on whether an "additive" or a "multiplicative" definition of "interaction" is adopted; the advantages and disadvantages of the two schemes were discussed by Darroch (1974).

The additive model was put forward almost simultaneously by Kupper and Haseman (1978) and Altham (1978). Let $\Pr[X_j = i] = p_i$ for $j = 1, \ldots, k$, $i = 0, 1$, and let $\Pr[X_j = i_1, X_{j'} = i_2] = p_{i_1 i_2}$. Let us suppose also that $Z_k = X_1 + \cdots + X_k$. Then

$$\frac{\Pr[X_1 = i_1, \ldots, X_k = i_k]}{p_{i_1} \cdots p_{i_k}} = \left(\sum_{1 \le a < b \le k} \frac{p_{i_a i_b}}{p_{i_a} p_{i_b}} \right) - \frac{k(k-1)}{2} + 1 \tag{3.117}$$

for $i_j = 0, 1$ with $j = 1, \ldots, k$; this gives a symmetric joint distribution defined by the two parameters p_1 and p_{11}, since $p_{00} + p_{01} = p_0 = 1 - p_1$ and $p_{01} = p_{10}$. Taking $1 - \alpha = p_{10}/(p_0 p_1)$, Altham deduced that

$$\Pr[Z_k = j] = \binom{k}{j} p_1^j (1-p_1)^{k-j} \times \left[\frac{\alpha}{2} \left(\frac{j(j-1)}{p_1} + \frac{(k-j)(k-j-1)}{1-p_1} \right) - \frac{\alpha k(k-1)}{2} + 1 \right]; \tag{3.118}$$

she showed that for certain values of α the distribution (3.118) is a mixture of three binomial distributions and fitted the distribution to some data using moment estimation. The case $\alpha = 0$ corresponds to independence between the X_i's and so gives a binomial distribution, whereas $\alpha > 0$ (< 0) corresponds to positive (negative) pairwise association. The constraints on α required in order that (3.118) is a valid pmf do not allow very strong pairwise association.

For the multiplicative model the assumption is that

$$\Pr[X_1 = i_1, \ldots, X_k = i_k] = K \prod_{1 \le a < b \le k} \phi_{i_a i_b}, \tag{3.119}$$

whence

$$\Pr[Z_k = j] = K \binom{k}{j} \phi_{00}^{(k-j)(k-j-1)/2} \phi_{01}^{(j-k)j} \phi_{11}^{j(j-k)/2}; \tag{3.120}$$

see Altham's paper. Taking

$$\theta = \frac{\phi_{01}}{(\phi_{00}\phi_{11})^{1/2}} \qquad \text{and} \qquad p = \frac{\phi_{11}^{(k-1)/2}}{\phi_{11}^{(k-1)/2} + \phi_{00}^{(k-1)/2}}$$

gives

$$\Pr[Z_k = j] = C' \binom{k}{j} p^j (1-p)^{k-j} \theta^{j(k-j)}, \tag{3.121}$$

where C' is a normalizing constant. This is an alternative two-parameter generalization of the binomial distribution, to which it reduces when $\theta = 1$. For $\theta > 1$ there is negative association producing a strongly unimodal distribution that is more peaked than the binomial; for $\theta < 1$ the association is positive, giving a flatter distribution.

The usefulness of the two distributions for toxicological experiments with laboratory animals, compared with that of the binomial distribution, has been discussed by Haseman and Kupper (1979); see also Makuch, Stephens, and Escobar (1989). From (1995) has described a weighted least-squares method for estimating the two parameters in Altham's model; this is simpler than maximum likelihood and is asymptotically efficient. Lovison (1998) has put forward an alternative formulation of Altham's distribution which is derived from Cox's loglinear representation for the joint distribution of n binary-dependent responses; the parameters are more intuitively meaningful than those used by Altham.

The beta–binomial is always overdispersed compared to a binomial distribution; see Section 6.2.2. An extended beta–binomial distribution that permits underdispersion (corresponding to negative association) has been discussed in relation to toxicological data by Prentice (1986). A more general modification of the binomial distribution that includes Altham's additive model as a special case has been proposed by Ng (1989). Paul's (1985) three-parameter, beta-correlated binomial distribution was modified in Paul (1987) to overcome certain theoretical difficulties associated with the earlier version of the distribution. Rudolfer (1990) gave a review of extrabinomial variation models in which he compared his {0, 1}-state Markov chain model of extrabinomial variation to Altham's additive and multiplicative models and to the beta–binomial model.

Qu, Greene, and Piedmonte (1993) have developed a generalization of the binomial distribution using a sum of symmetrically distributed Bernoulli rv's. Fu and Sproule (1995) have generalized the four-parameter binomial distribution with support $a, a+b, a+2b, \ldots, a+nb$.

Generalizations of the binomial distribution based on correlated Bernoulli rv's have received much attention since 1990. Drezner and Farnum's (1993) generalization allows dependence between trials and also nonconstant probabilities of success from trial to trial; it contains the usual binomial as a special case. It is defined recursively with

$$\begin{aligned}\Pr[X=x;n] &= \left[(1-\theta_n)p+\theta_n\frac{x-1}{n-1}\right]\Pr[X=x-1;n-1]\\ &\quad+\left[(1-\theta_n)(1-p)+\theta_n\frac{n-1-x}{n-1}\right]\Pr[X=x;n-1],\end{aligned} \tag{3.122}$$

with $\Pr[X=0;1]=1-p$, $\Pr[X=1;1]=p$, and boundary conditions

$$\Pr[X=-1;n]=\Pr[X=n+1;n]=0.$$

Drezner and Farnum (1994) developed a correlated Poisson distribution as a limiting form of their correlated trials binomial distribution.

Madsen (1993) has introduced EXBERT distributions. These arise as the joint distribution for a sequence of EXchangeable BERnoulli Trials and have probabilities of the form

$$\Pr[X=x;n]=\binom{n}{x}\sum_{j=0}^{n-x}\binom{n-x}{j}(-1)^j\pi_{x+j}, \tag{3.123}$$

where $\{\pi_k\}$, $k=0,1,\ldots,n$, belong to a sequence of probabilities with $\pi_0=1$ and all orders of differences nonnegative. Madsen has suggested three such sequences:

$$\begin{array}{llll}\pi_k=\rho p+(1-\rho)p^k, & k\ge 1, & 0\le\rho<1 & \text{(his "correlation" model)},\\ \pi_k=\exp^{(\ln p)\,k^a}, & 0\le a\le 1, & 0<p<1 & \text{(his "power" model)},\\ \pi_k=\beta+(1-\beta)\alpha^k, & 0\le\alpha\le 1, & 0<p<1 & \text{(his model 3)};\end{array}$$

the resultant distributions become the usual binomial distribution when $\rho = 0$ and $\beta = 0$.

Independently, George and Bowman (1995) and Bowman and George (1995) have also considered the joint distribution of a sequence of exchangeable Bernoulli trials where

$$\Pr[X = x; n] = \binom{n}{x} \sum_{j=0}^{n-x} \binom{n-x}{j} (-1)^j \lambda_{x+j},$$

$X = \sum_{i=1}^{n} X_i$ and $x = \sum_{i=1}^{n} x_i$. They developed the general theory for such models and fitted the model with

$$\lambda_k = \frac{2}{1 + \exp[\beta \log(k+1)]}, \qquad x \geq 0, \qquad \beta > 0,$$

(the "folded logistic function") to data from a large toxicity study. Vellaisamy (1996) has given simpler proofs and has developed new results for such distributions.

Song, Schlecht, and Groff's (2000) generalization of the binomial distribution uses correlated rv's converted to Bernoulli rv's.

Yu and Zelterman (2002) have discussed three models for the sum of a sequence of dependent but not identically distributed (and hence not exchangeable) Bernoulli rv's. For their "family history" model

$$\Pr[X = 0; n] = (1-p)^n,$$

$$\Pr[X = x; n] = p(p')^{x-1} \sum_{j=0}^{n-x} \binom{n-j-1}{x-1} (1-p)^j (1-p')^{n-x-j},$$

$x = 1, 2, \ldots, n$. The probabilities for their "incremental risk" model can be calculated recursively from

$$\begin{aligned} \Pr[X = x+1; n+1] = {} & C_n(x) \Pr[X = x; n] \\ & + [1 - C_n(x+1)] \Pr[X = x+1; n], \end{aligned}$$

where

$$C_n(k) = \frac{\exp(\alpha + \beta k)}{1 + \exp(\alpha + \beta k)}.$$

Yu and Zelterman have related these distributions to the binomial and beta–binomial distributions and have fitted all four, plus Altham's distribution, to data on interstitial pulmonary fibrosis.

CHAPTER 4

Poisson Distribution

4.1 DEFINITION

A random variable X is said to have a Poisson distribution with parameter θ if

$$\Pr[X = x] = \frac{e^{-\theta}\theta^x}{x!}, \qquad x = 0, 1, 2, \ldots, \qquad \theta > 0. \tag{4.1}$$

The characteristic function is $\exp[\theta(e^{it} - 1)]$, and the probability generating function (pgf) is

$$G(z) = e^{\theta(z-1)} = \frac{{}_0F_0[\ ;\ ;\theta z]}{{}_0F_0[\ ;\ ;\theta]} = {}_0F_0[\ ;\ ;\theta(z-1)]. \tag{4.2}$$

The distribution is a power series, distribution (PSD) with infinite nonnegative integer support. It belongs to the exponential family of distributions and is both a Kemp hypergeometric probability distribution and a Kemp hypergeometric factorial moment distribution.

The mean and variance are $\mu = \mu_2 = \theta$.

4.2 HISTORICAL REMARKS AND GENESIS

4.2.1 Genesis

Poisson (1837, Sections 73, pp. 189–190, and 81, pp. 205–207) published the following derivation of the distribution that bears his name. He approached the distribution by considering the limit of a sequence of binomial distributions with

$$p_{x,N} = \Pr[X = x] = \begin{cases} \binom{N}{x} p^x (1-p)^{N-x} & \text{for } x = 0, 1, \ldots, N, \\ 0 & \text{for } x > N, \end{cases}$$

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

in which N tends to infinity and p tends to zero while Np remains finite and equal to θ. It can be established by direct analysis that

$$\lim_{\substack{N\to\infty \\ Np=\theta}} \sum_w p_{x,N} = \sum_w \frac{e^{-\theta}\theta^x}{x!}, \tag{4.3}$$

where $\sum_w$ denotes summation over any (finite or infinite) subset w of the non-negative integers $0, 1, 2, \ldots$.

The result had been given previously by de Moivre (1711) in *De Mensura Sortis*, p. 219. Bortkiewicz (1898) considered circumstances in which Poisson's distribution might arise. From the point of view of Poisson's own approach, these are situations where, *in addition* to the requirements of independence of trials and consistency of probability from trial to trial, the number of trials must be very large while the probability of occurrence of the outcome under observation must be small. Although Bortkiewicz called this the *law of small numbers*, there is no need for $\theta = Np$ to be "small." It is the largeness of N and the smallness of p that are important. In Bortkiewicz's (1898) book, one of the "outcomes" considered was the number of deaths from kicks by horses per annum in the Prussian Army Corps. Here was a situation where the probability of death from this cause was small while the number of soldiers exposed to risk (in any one Corps) was large. Whether the conditions of independence and constant probability are satisfied is doubtful. However, the data available to Bortkiewicz were quite satisfactorily fitted by Poisson distributions and have been very widely quoted as an example of the applicability of this distribution. The data have been discussed in detail by Quine and Seneta (1987). Bortkiewicz also obtained satisfactory fits for fatal accident data and suicide data, including data on numbers of suicides per year for Prussian boys under 10 years old (also for girls).

Bortkiewicz gave tables of the probabilities, and he obtained many properties of the distribution, such as difference and differential equations for the probabilities, and the moments (derived as limiting forms of the moments of the binomial distribution).

Thiele (1889) gave a completely different derivation. He showed that for the distribution with cumulants $\kappa_r = ba^r$ the observed values are $0, a, 2a, 3a, \ldots$ and the relative frequency of ra is $b^r/r!$; he considered that the distribution is perhaps superior to the binomial as a representative of some skew laws of error.

Charlier (1905a) considered the binomial sampling scheme with pgf $\prod_{i=0}^{n}(1 - p_i + p_i z)$ and showed that the success probabilities p_i do not need to be constant for the Poisson limit to hold.

Charlier (1905b) derived the distribution from the differential-difference equations

$$\frac{dp_0}{dt} = -p_0, \qquad \frac{dp_r}{dt} = -p_r + p_{r-1}, \qquad r \geq 1, \tag{4.4}$$

where the probabilities are functions of time. Bateman (1910) and McKendrick (1914) also gave this model (the pure birth process with constant birth rate).

"Student" (W. S. Gosset) (1907) used the Poisson distribution to represent, to a first approximation, the number of particles falling in a small area A when a large number of such areas are spread at random over a surface large in comparison with A.

The Poisson distribution may also arise for events occurring "randomly and independently" in time. If it be supposed that the future lifetime of an item of equipment is independent of its present age (Johnson et al., 1994, Section 18.1), then the lifetime can be represented by a rv T with pdf of form

$$p_T(t) = \tau^{-1} \exp\left(-\frac{t}{\tau}\right), \qquad t \geq 0, \qquad \tau \geq 0,$$

that is, by an exponential rv. The expected value of T is τ (Johnson et al., 1994, Section 18.4). Now imagine a situation in which each item is replaced by another item with exactly the same lifetime distribution. Then the distribution of the number of failures (complete lifetimes) X in a period of length t is given by

$$\Pr[X = x] = \Pr[T_1 + T_2 + \cdots + T_{x-1} > t] - \Pr[T_1 + T_2 + \cdots + T_x > t], \tag{4.5}$$

where the lengths of successive lifetimes are denoted by $T_1, T_2, \ldots$ and $x \geq 1$. From the relationship between the χ^2 and Poisson distributions (see Section 4.12.2), it follows that

$$\begin{aligned} \Pr[X = x] &= \Pr\left[\chi^2_{2(x+1)} > \frac{2t}{\tau}\right] - \Pr\left[\chi^2_{2x} > \frac{2t}{\tau}\right] \\ &= \sum_{j=0}^{x} e^{-t/\tau} \frac{(t/\tau)^j}{j!} - \sum_{j=0}^{x-1} e^{-t/\tau} \frac{(t/\tau)^j}{j!} \\ &= e^{-t/\tau} \frac{(t/\tau)^x}{x!}, \end{aligned} \tag{4.6}$$

as in (4.1), with θ replaced by t/τ. Hence X has a Poisson distribution with parameter t/τ.

This mode of genesis underlies the use of the Poisson distribution to represent variations in the number of particles ("rays") emitted by a radioactive source in forced periods of time. Rutherford and Geiger (1910) gave some numerical data that were fitted well by a Poisson distribution; see also Rutherford et al. (1930).

A unification of the then existing theory was given by Bortkiewicz (1915), who utilized the gap distribution concept; see Haight (1967). The Poisson distribution can be characterized by the renewal counting process with exponential interarrival times.

Lévy (1937b, pp. 173–174) initiated the axiomatic approach. The Poisson distribution is the counting distribution for a Poisson process. This is a stochastic point process satisfying the following conditions: whatever the number of process points in the interval (0, t), the probability that in the interval $(t, t + \delta t)$ a point occurs is $\theta\delta t + o(\delta t)$ and the probability that more than one point occurs is $o(\delta t)$ (see, e.g., Feller, 1968, p. 447).

Parzen (1962, p. 118) adopted a more rigorous approach. He showed that the Poisson process $X(t)$ satisfies the following five axioms:

Axiom 0 $X(0) = 0$.

Axiom 1 $X(t)$ has independent increments; that is, for all t_i such that $t_0 < t_1 < \cdots < t_n$, the rv's $X(t_i) - X(t_{i-1})$, $i = 1, 2, \ldots, n$, are independent.

Axiom 2 For any $t > 0$, $0 < \Pr[X(t) > 0] < 1$.

Axiom 3 For any $t > 0$,

$$\lim_{h\to 0} \frac{\Pr[X(t+h) - X(t) \geq 2]}{\Pr[X(t+h) - X(t) = 1]} = 0.$$

Axiom 4 $X(t)$ has stationary increments; that is, for points $t_i > t_j \geq 0$ (and $h > 0$), the rv's $X(t_i) - X(t_j)$ and $X(t_i + h) - X(t_j + h)$ are equidistributed.

There are several proofs that Axioms 0–4 imply that there exists a constant θ such that

$$\Pr[X(t) = x] = \frac{e^{-\theta t}(\theta t)^x}{x!}.$$

Modification of these axioms leads to more general stochastic processes; for example, replacing Axiom 4 by

$$\lim_{h\to 0} \frac{1 - \Pr[X(t+h) - X(t) = 0]}{h} = \lambda(t)$$

leads to the nonhomogeneous process (Parzen, 1962, p. 125).

In his monograph on the Poisson distribution, Haight (1967) discussed other axiomatic approaches such as that of Fisz and Urbanik (1956), who gave a sufficient condition for a process satisfying Axioms 0–2 to be a Poisson process. Haight also described a number of special probability models, including, for example, the model of Hurwitz and Kac (1944).

Johnson and Kotz (1977) focused on urn models in relation to the Poisson distribution.

Since

$$\Pr[X > x + y | X > x] = \Pr[X > y]$$

iff X is exponentially distributed (Feller, 1957, Chapter 17; Parzen, 1962), it follows from the exponential distribution of interarrival times for a Poisson process that the distribution of the time between two Poisson events is surprisingly the same as the distribution of the time between an arbitrary point in time and the next Poisson event.

A Poisson distribution is the outcome of an aspect of maximum disorder that arises in information theory; see Rényi (1964). A Poisson arrangement of points, obtained by thoroughly shuffling other points, was studied by Maruyama (1955) and Watanabe (1956); Haight (1967) gave further references.

Metzler, Grossmann, and Wakolbinger (2002) have put forward a simple Poisson model that reflects important features of high-score gapped local alignments of independent DNA sequences.

In recent years, the Poisson distribution has been applied in an increasing number of situations. Like the binomial distribution, it often serves as a standard from which to measure departures, even when it is not itself an adequate representation of the real situation.

4.2.2 Poissonian Approximations

The Poisson distribution is a limiting form for the binomial distribution (see Section 3.6.1) and for many other distributions. In particular, it arises as

$$\lim_{\substack{k\to\infty\\ kp=\theta}} (1+p-pz)^{-k} = e^{\theta(z-1)}, \tag{4.7}$$

that is, as a limiting form of the negative binomial distribution; see Section 5.12.1. A realization of this relationship would have done much to defuse the controversy aroused by Whittaker (1914) regarding the intrinsic merits of the Poisson and negative binomial models.

Widdra (1972) obtained a Poisson distribution as the limiting distribution (as $n \to \infty$) of the number of successes in n *partially dependent* trials. He showed that, if Y_{gh} are indicator variables with $Y_{ih}Y_{jh} = 0$ if $i \neq j$, where $i, j = 1, \ldots, m$ and $h = 1, \ldots, n$, but otherwise Y_{gh} and Y_{ij} are independent with constant success probability $E[Y_{gh}] = p$ for all g, h, then

$$\sum_{i=1}^{m} \sum_{\substack{h=1\\ h<k}}^{n} \sum_{k=1}^{n} Y_{ih} Y_{ik}$$

has a limiting Poisson distribution as $m, n \to \infty$ with $mn(n-1)p^2/2 = \theta$.

Serfling (1977) obtained reliability bounds for k-out-of-n systems based on Poisson approximations.

Kreweras (1979) described two unusual modes of genesis for the Poisson distribution with $\theta = 1$. Consider all $n!$ permutations of the integers $1, 2, \ldots, n$ and let $u(n, x)$ be the number of these permutations in which there are x pairs of integers in their natural order. For example, with $n = 5$, the permutation 12543 has $x = 1$, 12534 has $x = 2$, and 12345 has $x = 4$. Then

$$\lim_{n\to\infty} \left(\frac{u(n, x)}{n!}\right) = \frac{e^{-1}}{x!}. \tag{4.8}$$

That is, the distribution of the number of pairs of integers in natural order (termed by Kreweras the *regularity* of the permutation) tends to a Poisson distribution with $\theta = 1$.

His other mode of genesis involves the consideration of partitions of the first $2n$ integers into pairs of integers. The total number of these partitions is

$$1 \times 3 \times 5 \times \cdots \times (2n-1) = \frac{2^{-n}(2n)!}{n!};$$

the proportion of them containing exactly x pairs of two consecutive integers also tends to $e^{-1}/x!$ as $x \to \infty$.

Important results by Chen (1975) [see also Stein (1972)] on convergence to the Poisson distribution for the number of successes in n dependent trials using only the first two moments have been reexamined by Arratia, Goldstein, and Gordon (1989, 1990). Further research on the Chen–Stein method for Poissonian approximations has been carried out by Brown and Xia (1995) and by Barbour and his colleagues; see Barbour, Holst, and Janson (1992).

A simple proof concerning Poisson approximations to PSDs, depending only on the continuity of pgf's, appears in Pérez-Abreu (1991).

The work of Logunov (1991) and Yannaros (1991) on Poissonian approximation to a random sum of nonnegative integer-valued rv's has been extended by Vellaisamy and Chaudhuri (1996).

Olofsson's (1999) Poissonian approximation has applications to extreme-value theory in discrete samples. Čekanavičius (1998) uses integer-centered Poissonian approximation. Čekanavičius (2002) obtains a Poissonian analog for the Markov binomial distribution of the Johnson–Simons theorem (Simons and Johnson, 1971) on Poissonian convergence for the binomial distribution. Poissonian approximations for the Poisson–binomial and mixed Poisson distributions are developed in Roos (2001, 2003).

Kathman and Terrell's (2002, 2003) improvements in Poissonian approximations for tail probabilities build on the tilting and expanding methods in Barbour, Holst, and Janson (1992) via new forms of generating functions.

4.3 MOMENTS

The simplicity of the Poisson pgf, $G(z) = \exp[\theta(z-1)]$, leads to simple expressions for the cumulant generating function (cgf), factorial moment generating function (fmgf), and factorial cumulant generating function (fcgf).

The cgf is $\theta(e^t - 1)$, so

$$\kappa_r = \theta \quad \text{for all } r \geq 1. \tag{4.9}$$

The fmgf is $e^{\theta t}$, whence

$$\mu'_{[r]} = \theta^r \quad \text{for all } r \geq 1, \tag{4.10}$$

and the fcgf is θt, whence

$$\kappa_{[1]} = \theta, \qquad \kappa_{[r]} = 0 \quad \text{for all } r > 1. \tag{4.11}$$

From the relationship between factorial moments and moments about the origin, we have

$$\mu'_r = \begin{cases} \displaystyle\sum_{j=0}^{r} \frac{\theta^j}{j!} \Delta^j 0^r = \sum_{j=0}^{r} S(r, j)\theta^j, & r = 1, 2, \ldots, \\ \theta\mu'_{r-1} + \theta \dfrac{\partial \mu'_{r-1}}{\partial\theta}, & r = 2, 3, \ldots, \end{cases} \tag{4.12}$$

where $\Delta^j 0^r$ is a difference of zero and $S(r, j)$ is a Stirling number of the second kind; see Section 1.1.3.

The moment generating function (mgf) is

$$E[e^{tX}] = \exp[\theta(e^t - 1)], \tag{4.13}$$

and the central moment generating function (cmgf) is

$$E[e^{t(X-\mu)}] = E[e^{t(X-\theta)}] = e^{\theta(e^t-1-t)}; \tag{4.14}$$

the latter yields the following recurrence relationship for the moments about the mean:

$$\mu_{r+1} = r\theta\mu_{r-1} + \theta\frac{\partial \mu_r}{\partial\theta}, \qquad r = 2, 3, \ldots \tag{4.15}$$

(Riordan, 1937). Also Kendall (1943) showed that

$$\mu_r = \theta \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j, \qquad r = 2, 3, \ldots. \tag{4.16}$$

Hence

$$\begin{aligned} \mu &= \kappa_1 = \theta, \\ \mu_2 &= \kappa_2 = \theta, \\ \mu_3 &= \kappa_3 = \theta, \\ \mu_4 &= 3\theta^2 + \theta, \\ \mu_5 &= 10\theta^2 + \theta, \\ &\vdots \end{aligned} \tag{4.17}$$

A Poisson rv with *parameter* θ is sometimes said to have a Poisson distribution with *expected value* θ.

The moment ratios $\sqrt{\beta_1}$ and β_2 are given by

$$\sqrt{\beta_1} = \theta^{-1/2}, \qquad \beta_2 = 3 + \theta^{-1}. \tag{4.18}$$

Note that the Poisson distribution has the properties

$$E[X] = \text{Var}(X) \qquad \text{and} \qquad \beta_2 - \beta_1 - 3 = 0.$$

The coefficient of variation is $\mu_2^{1/2}/\mu = \theta^{-1/2}$. The index of dispersion is $\mu_2/\mu = 1$; this is widely used in ecology as a standard against which to measure clustering (overdispersion with $\mu_2 > \mu$) or repulsion (underdispersion with $\mu_2 < \mu$).

The mean deviation of the Poisson distribution is

$$\nu_1 = E[|X-\theta|] = 2e^{-\theta}\theta^{[\theta]+1}/[\theta]! \tag{4.19}$$

(where $[\cdot]$ denotes the integer part); this was shown by Ramasubban (1958), who also gave an expression for the mean difference. Crow (1958) also discussed the mean deviation, and he showed that the rth inverse (ascending) factorial moment is

$$\begin{aligned}\mu_{-[r]} = E[\{(X+1)^{[r]}\}^{-1}] &= e^{-\theta}\sum_{j=0}^{\infty}\frac{\theta^j}{(j+r)!}\\ &= \theta^{-r}\left(1-e^{-\theta}\sum_{j=0}^{r-1}\frac{\theta^j}{j!}\right).\end{aligned} \tag{4.20}$$

For the Poisson distribution

$$\begin{aligned}\frac{\text{Mean deviation}}{\text{Standard deviation}} &= \frac{2\theta^{[\theta]+1/2}}{[\theta]!e^{\theta}}\\ &\approx \sqrt{\frac{2}{\pi}}\left(\frac{\theta}{[\theta]}\right)^{[\theta]+1/2}\exp\left[-(\theta-[\theta])-\frac{1}{12[\theta]}\right].\end{aligned} \tag{4.21}$$

As θ increases, the ratio oscillates; it tends to $\sqrt{2/\pi} \approx 0.798$.

Haight (1967) quoted a number of formulas relating to the incomplete moments.

Jones and Zhigljavsky (2004) have used the Stirling numbers of the first kind to obtain approximations for the negative moments of the Poisson distribution.

4.4 PROPERTIES

From (4.1)

$$\frac{\Pr[X=x+1]}{\Pr[X=x]} = \frac{\theta}{x+1}, \qquad x = 0, 1, \ldots, \tag{4.22}$$

whence it follows that $\Pr[X=x]$ increases with x to a maximum at $x = [\theta]$ (or to two equal maxima at $x = \theta - 1$ and $x = \theta$ if θ is an integer) and thereafter decreases as x increases; the distribution is therefore unimodal.

Note that Equation (4.19) can be written as

$$\nu_1 = 2\theta \max_x \Pr[X=x].$$

Moreover $\Pr[X=x]$ increases monotonically with θ for fixed x if $\theta \le x$ and decreases monotonically if $\theta \ge x$.

The probabilities are logconcave; that is,

$$(\Pr[X = x + 1])^2 > \Pr[X = x] \Pr[X = x + 2]. \tag{4.23}$$

They do not satisfy the logconvexity condition that is a *sufficient* condition for infinite divisibility; nevertheless,

$$e^{\theta(z-1)} = \left(e^{(\theta/n)(z-1)}\right)^n \tag{4.24}$$

for any positive integer n, and so the distribution is *infinitely divisible*.

Hadley and Whitin (1961) and Said (1958) listed a considerable number of formulas involving Poisson probabilities that might be of use in engineering and operations research. A few are listed here; others appear in various parts of this chapter. We use the notations

$$w(x, \theta) = \frac{e^{-\theta}\theta^x}{x!} \qquad \text{and} \qquad Q(x, \theta) = \sum_{j=x}^{\infty} w(j, \theta).$$

Then from (4.1) and (4.22)

$$\begin{aligned}
xw(x, \theta) &= \theta w(x - 1, \theta), \\
\sum_{j=0}^{x} w(j, \theta_1) w(x - j, \theta_2) &= w(x, \theta_1 + \theta_2), \\
\sum_{j=x}^{\infty} j w(j, \theta) &= \theta Q(x - 1, \theta), \\
x Q(x + 1, \theta) &= x Q(x, \theta) - \theta w(x - 1, \theta), \\
\sum_{j=x}^{\infty} Q(j, \theta) &= \theta Q(x - 1, \theta) + (1 - x) Q(x, \theta), \\
\sum_{j=0}^{x} Q(j, \theta) &= \theta[1 - Q(x, \theta)] - x Q(x + 1, \theta), \\
\sum_{j=x}^{\infty} j^2 w(j, \theta) &= \theta Q(x - 1, \theta) + \theta^2 Q(x - 2, \theta), \\
\sum_{j=x}^{\infty} j Q(j, \theta) &= \frac{1}{2}\theta^2 Q(x - 2, \theta) + \theta Q(x - 1, \theta) \\
&\quad - \frac{1}{2} x(x - 1) Q(x, \theta), \\
\sum_{j=0}^{x} j Q(j, \theta) &= \frac{1}{2}\theta^2 [1 - Q(x - 1, \theta)] + \theta[1 - Q(x, \theta)] \\
&\quad + \frac{1}{2} x(x + 1) Q(x + 1, \theta), \\
\frac{\partial w(x, \theta)}{\partial \theta} &= w(x - 1, \theta) - w(x, \theta)
\end{aligned} \tag{4.25}$$

[see also Haight (1967)].

The cdf is a nonincreasing function of θ for fixed x. It satisfies

$$\begin{aligned}\sum_{j=0}^{x} w(j,\theta) &\geq \sum_{j=0}^{x-1} w(j,\theta-1) \quad \text{if } x \leq \theta - 1,\\ \sum_{j=0}^{x} w(j,\theta) &\leq \sum_{j=0}^{x-1} w(j,\theta-1) \quad \text{if } x \geq \theta\end{aligned} \tag{4.26}$$

[see Anderson and Samuels (1967)].

The cdf can be expressed in terms of the incomplete gamma function (see Section 1.1.5) as

$$\sum_{j=0}^{x} \Pr[X = j] = \frac{\Gamma(x+1,\theta)}{\Gamma(x)}, \tag{4.27}$$

$$\sum_{j=x+1}^{\infty} \Pr[X = j] = \frac{\gamma(x+1,\theta)}{\Gamma(x)}. \tag{4.28}$$

This gives a very important relationship between the Poisson and χ^2 distributions; see Sections 4.5 and 4.12.2.

Kemp and Kemp (1990) have considered the behavior of the sums of two adjacent probabilities. Crow and Gardner (1959) have studied the sums of r adjacent probabilities. Cheng (1949) has studied the median of the Poisson distribution; it lies between $[\theta - 1]$ and $[\theta + 1]$ [see also Lidstone (1942)]. Teicher (1955) gave results for $\sum_{j=0}^{[\theta]} w(j,\theta)$; see also Kemp (1988b) and Section 4.5.

Because the distribution is logconcave, $1/r_x > 1/r_{x+1}$, where r_x is the hazard function (failure rate); the distribution therefore has an increasing failure rate. The entropy is

$$\theta \ln\theta - \theta - e^{-\theta} \sum_{j\geq 0} \frac{\theta^j[\ln(j!)]}{j!}. \tag{4.29}$$

If X_1 and X_2 are independent variables, each having a Poisson distribution, with expected values θ_1 and θ_2, respectively, then $X_1 + X_2$ has a Poisson distribution with expected value $\theta_1 + \theta_2$. This follows from (4.2) above. The distribution of the mean for samples of size n from a Poisson (θ) distribution therefore has pmf

$$p_{\overline{x}} = \frac{e^{-n\theta}(n\theta)^{\sum x}}{(\sum x)!}, \quad \text{where } \overline{x} = 0, 1/n, 2/n, \ldots. \tag{4.30}$$

The distribution of the difference $X_1 - X_2$ cannot be expressed in so simple a form. It will be discussed in Section 4.12.3.

The *conditional* distribution of X_1, given that $X_1 + X_2 = n$, is a binomial with parameters n, $\theta_1/(\theta_1 + \theta_2)$. If $X_1, \ldots, X_k$ is a sample of k independent observations from a Poisson (θ) distribution, then the conditional distribution of $X_1, \ldots, X_k$, given that $\sum_{j=1}^{k} X_j = N$, is multinomial with N trials and k equally likely categories.

Approximate formulas for the mean and variance of the median of a sample of n observations from a Poisson distribution were derived by Abdel-Aty (1954).

Steutel and Thiemann (1989b) used their results concerning exceedance times for a gamma process to obtain approximations for the means and variances of the *order statistics* for the Poisson distribution. They obtained the following exact results for samples of size 2:

$$\begin{aligned} E[X_{2;2}] &= \mu + \mu e^{-2\mu}[I_0(2\mu) + I_1(2\mu)], \\ \mathrm{Var}(X_{2;2}) &= \mu + \mu^2 e^{-4\mu}[I_0(2\mu) + I_1(2\mu)]^2 + \mu e^{-2\mu} I_0(2\mu), \end{aligned} \tag{4.31}$$

where $X_{1;n} \leq X_{2;n} \leq \cdots \leq X_{n;n}$ are the order statistics for a sample of size n and $I_j(\cdot)$ denotes a modified Bessel function of the first kind (Section 1.1.5). They also gave approximate formulas for $E[X_{j;n}]$ and $\mathrm{Var}(X_{j;n})$ for large μ.

The limiting form of the Poisson distribution as θ becomes large is normal; the standardized variable

$$Y = \frac{X - \theta}{\theta^{1/2}}$$

approaches a standard normal distribution. This is useful for giving an approximation for the probability tail when θ is moderately large and for obtaining approximate confidence intervals; see Section 4.7.3.

Results on sampling moments for random samples from a Poisson distribution have been obtained by Gart and Pettigrew (1970). For example,

$$E\left[k_j | X\right] = \overline{x}, \qquad j = 1, 2, \ldots, \tag{4.32}$$

where $X = \sum x_i$ and k_j is the jth k-statistic (see Section 1.2.12). They found moreover that

$$\mathrm{Var}(k_j|X) = E\left[\sum_{r=1}^{j} \frac{A_{rj} X(X-1)\cdots(X-r+1)}{n^r}\right], \tag{4.33}$$

where the A_{rj}'s can be found from the sampling cumulants of the k-statistics as given in Stuart and Ord (1987, Chapter 12). In particular

$$
\begin{aligned}
\operatorname{Var}(k_2|X) &= \frac{2X(X-1)}{n^2(n-1)},\\
\operatorname{Var}(k_3|X) &= \frac{6X(X-1)}{n^2(n-1)}\left(3+\frac{X-2}{n-2}\right),\\
\operatorname{Var}(k_4|X) &= \frac{2X(X-1)}{n^2(n-1)}\\
&\quad\times\left(49+\frac{108(X-2)}{n-2}+\frac{12(n+1)(X-2)(X-3)}{n(n-2)(n-3)}\right).
\end{aligned}
\tag{4.34}
$$

4.5 APPROXIMATIONS, BOUNDS, AND TRANSFORMATIONS

The relationship between the Poisson and χ^2 distributions (Section 4.12.2) implies that approximations to the cdf of the central χ^2 distribution can also be used as approximations to Poisson probabilities, and vice versa. Thus, if X has distribution (4.1), then

$$\Pr[X \le x] = \Pr[\chi^2_{2(x+1)} > 2\theta]. \tag{4.35}$$

The Wilson–Hilferty approximation to the χ^2 distribution (Johnson et al., 1994, Chapter 17) yields

$$\Pr[X \le x] \approx (2\pi)^{-1/2}\int_z^\infty e^{-u^2/2}\,du, \tag{4.36}$$

where

$$z = 3\left[\left(\frac{\theta}{x+1}\right)^{1/3} - 1 + \frac{1}{9(x+1)}\right](x+1)^{1/2}.$$

As $\theta \to \infty$, the standardized Poisson distribution tends to the unit-normal distribution. Modifying the crude approximation "$Z = (X-\theta)/\sqrt{\theta}$ is unit normally distributed," as in the Cornish–Fisher expansion (Johnson et al., 1994, Chapter 17), leads to $Z - \frac{1}{3}(Z^2-1)\theta^{-1/2} + \frac{1}{36}(7Z^3 - Z)\theta^{-1} - \cdots$ being unit normally distributed.

A formal Edgeworth expansion gives

$$
\begin{aligned}
\Pr(X \le x) = \Phi(z) - \phi(z)\Bigg(&\frac{z^2-1}{6\theta^{1/2}} + \frac{z^5 - 7z^3 + 3z}{72\theta}\\
&+ \frac{5z^8 - 95z^6 + 384z^4 - 129z^2 - 123}{6480\theta^{3/2}}\Bigg) + O(\theta^{-2}),
\end{aligned}
\tag{4.37}
$$

where $z = \left(x - \theta + \frac{1}{2}\right)\theta^{-1/2}$; see, for example, Matsunawa (1986).

Peizer and Pratt's (1968) very accurate approximation is $\Pr[X \le x] \approx \Phi(z)$, where

$$z = \left(x - \theta + \frac{2}{3} + \frac{\epsilon}{x+1}\right)\left[1 + T\left(\frac{x+1/2}{\theta}\right)\right]^{1/2} \theta^{-1/2},$$

$T(y) = (1 - y^2 + 2y \ln y)(1 - y)^{-2}$, $T(1) = 0$, and $\epsilon = 0$ for simplicity or $\epsilon = 0.02$ for more accuracy [Molenaar (1970a) suggested taking $\epsilon = 0.022$]. Peizer and Pratt discussed the error of this approximation.

A variance stabilization transformation is often wanted. A very simple solution is to take $2\sqrt{X}$. This is approximately normally distributed with expected value $2\sqrt{\theta}$ and variance 1. An improvement, suggested by Anscombe (1948), is to use $2\sqrt{X + \frac{3}{8}}$. Freeman and Tukey (1950) have suggested the transformed variable

$$Y = \sqrt{X} + \sqrt{X+1}. \tag{4.38}$$

Tables of this quantity to two decimal places for $x = 0(1)50$ have been given by Mosteller and Youtz (1961).

Molenaar (1970a,b, 1973) has made a detailed study of these and other approximations. As simple approximations with "reasonable" accuracy, he advises

$$P = \sum_{j=0}^{x} e^{-\theta} \frac{\theta^j}{j!} \approx \begin{cases} \Phi\left[2\left(x + \frac{3}{4}\right)^{1/2} - 2\theta^{1/2}\right] & \text{for } 0.06 < P < 0.94, \\ \Phi[2(x+1)^{1/2} - 2\theta^{1/2}] & \text{for tails.} \end{cases} \tag{4.39}$$

Individual Poisson probabilities may be approximated by the formula

$$\Pr[X = x] \approx (2\pi)^{-1/2} \int_{K_-}^{K_+} e^{-u^2/2}\, du, \tag{4.40}$$

where $K_- = \left(x - \theta - \frac{1}{2}\right)\theta^{-1/2}$ and $K_+ = \left(x - \theta + \frac{1}{2}\right)\theta^{-1/2}$.

When x is large, Stirling's expansion for $\Gamma(x+1) = x!$ has often been used; this gives

$$\Pr[X = x] \approx \frac{e^{x-\theta}}{\sqrt{2\pi x}}\left(\frac{\theta}{x}\right)^x \left(1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \frac{571}{2488320x^4} + \cdots\right)^{-1}. \tag{4.41}$$

Kemp (1988b, 1989) derived a J-fraction approximation (Hart, 1968) from Stirling's expansion:

$$\Pr[X = x] \approx \frac{e^{x-\theta}}{\sqrt{2\pi x}} \left(\frac{\theta}{x}\right)^x \left(1 - \frac{\frac{1}{12}}{x + \frac{1}{24} + 293/(8640x)}\right). \tag{4.42}$$

When $\theta = r$, where r is an integer, this gives the easily computed approximation for the modal probability

$$\Pr[X = r] \approx (2\pi r)^{-1/2} \left(1 - \frac{1}{u + \frac{1}{2} + 293/(60u)}\right), \tag{4.43}$$

where $u = 12r$. A similar expression for the modal cumulative probability for integer θ is

$$\sum_{j=0}^{r} \Pr[X = r] \approx 0.5 + 2(2\pi r)^{-1/2} 3^{-1} \left[1 - \frac{a}{u + b + c/(u + d)}\right], \tag{4.44}$$

where $a = 23/15$, $b = 15/14$, $c = 30557/4508$, $d = 138134432/105880005$ (Kemp, 1988b). Kemp gave parallel formulas for the modal probability and modal cumulative probability when $\theta = r + 0.5$, r an integer. She also gave adjustments to these formulas for $\theta = r + \alpha$ and $\theta = r + 0.5 + \beta$, where α and β are fractional. The accuracy of these approximations is discussed in Section 4.6.1.

Coming now to bounds, which have the advantage over approximations that the sign of the error is known, we have first the inequality (Teicher, 1955)

$$\Pr[x \le \theta] \ge e^{-1}. \tag{4.45}$$

(If θ is an integer, the right-hand side can be replaced by $\frac{1}{2}$.) Bohman (1963) has given the following inequalities:

$$\begin{aligned} \Pr[X \le x] &\le (2\pi)^{-1/2} \int_{-\infty}^{(x+1-\theta)/\sqrt{\theta}} e^{-u^2/2}\,du, \\ \Pr[X \le x] &\ge [\Gamma(\theta + 1)]^{-1} \int_0^x t^\theta e^{-t}\,dt. \end{aligned} \tag{4.46}$$

Samuels (1965) gave these very simple bounds:

$$\begin{aligned} \Pr[X \le x - 1] &> 1 - \Pr[X \le x] \quad &&\text{if } x \ge \theta, \\ \Pr[X \le x] &\ge \exp\left(-\tfrac{\lambda}{x+1}\right) &&\text{if } x + 1 \ge \theta. \end{aligned} \tag{4.47}$$

Devroye (1986, p. 508) suggested a number of upper bounds for the "normalized log" probabilities

$$q_j = \ln \Pr[X = m + j] + \ln(m!) + \theta - m \ln \theta, \tag{4.48}$$

where $m = [\theta]$.

Approximations for the distribution of the range of random samples from a Poisson distribution, based on the work of Johnson and Young (1960) on the multinomial distribution, have been investigated numerically by Bennett and Nakamura (1970).

Pettigrew and Mohler (1967) have used the relationship with the multinomial distribution to devise a quick verification test for the Poisson distribution; see Section 4.7.1.

4.6 COMPUTATION, TABLES, AND COMPUTER GENERATION

4.6.1 Computation and Tables

The calculation of an individual Poisson probability involves $x!$. Direct multiplication, with $x! = 1 \times 2 \times \cdots \times (x-1)x$, can be used for low values of x. However, $x!$ increases with x very rapidly, necessitating the use of an approximation such as Stirling's expansion or Kemp's (1989) J-fraction (see the previous section).

The customary method of calculation for a complete set of Poisson probabilities, or for the cumulative probabilities, is via the recurrence relation

$$\begin{aligned} \Pr[X = 0] &= e^{-\theta}, \\ \Pr[X = x] &= \frac{\theta}{x} \Pr[X = x - 1], \qquad x \geq 1. \end{aligned} \tag{4.49}$$

The problem of buildup errors through recursive calculations does not seem to be serious. C. D. Kemp and A. W. Kemp (1991) have studied the accuracy of individual and cumulative probabilities when computed recursively:

1. from the origin using

$$\Pr[X = x] = \frac{\theta}{x} \Pr[X = x - 1],$$

2. from the origin using

$$\ln \Pr[X = x] = \ln \Pr[X = x - 1] + \ln \theta - \ln x,$$

3. from the mode using Kemp's (1988b) modal approximations (see the previous section).

Their results were reported briefly in C. D. Kemp and A. W. Kemp (1991). Kemp's approximations were found to differ from the true values of the modal and cumulative probabilities by less than 10^{-7} when $\theta \geq 10.5$ and by less than 10^{-10} for $\theta \geq 30.5$.

Fox and Glynn (1988) have put forward an algorithm to compute a complete set of individual Poisson probabilities, with truncation error in the two tails rigorously bounded from above and the remaining probabilities rigorously bounded from below; the algorithm is designed to avoid underflow or overflow.

The first detailed published tables for the Poisson distribution were contained in Molina (1942); these gave six-place tables for the Poisson probabilities. Cumulative probabilities were also given.

Kitagawa's (1952) tables were more detailed. Janko (1958) also published tables of the Poisson probabilities and cumulative probabilities.

Hald and Kousgaard (1967) published values of θ satisfying the equation

$$e^{-\theta} \sum_{j=0}^{c} \frac{\theta^j}{j!} = P$$

to four significant figures for $P = 0.001$, 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 0.8, 0.9, 0.95, 0.975, 0.99, 0.995, 0.999 and $c = 0(1)50$.

Pearson and Hartley (1976) gave six-place values of $\Pr[X = x]$ for $\theta = 0.1$ $(0.1)15.0$ and $x = 0(1)\infty$; they also tabulated the probabilities $\Pr[X \leq x]$ for $x = 1(1)35$ and

$$\theta = 0.0005(0.0005)0.005(0.005)0.05(0.05)1(0.1)5(0.25)10(0.5)20(1)60.$$

The *Tables of the Incomplete Gamma Function Ratio* by Khamis and Rudert (1965) can also be used to obtain Poisson probabilities for distributions with

$$\theta = 0.00005(0.00005)0.0005(0.0005)0.005(0.005)0.5(0.025)3.0$$
$$(0.05)8.0(0.25)33.0(0.5)83.0(1)125.$$

Values are given to 10 decimal places.

Table 2 on pages 100–101 of the first edition of this book gave details of some other tables.

4.6.2 Computer Generation

The importance of the Poisson distribution in discrete distribution theory has led to major effort in the design of good and fast generators for Poisson variates. For a fixed value of the parameter θ a fast general method has much to commend it.

The generation of many small samples from different Poisson distributions requires the use of varying values of θ. Moreover Poisson variates with varying θ are useful for generating many other distributions, in particular those that can be interpreted as mixed Poisson distributions and those that are Poisson–stopped sum distributions. In such situations general methods involving the creation of tables for a specific value of θ are no longer suitable.

Here we mention only three of the simpler methods. Devroye (1986) gives details of many other algorithms. Characterizations of the Poisson distribution form an important feature of certain methods.

The *exponential-gap* method was widely used in the past. It exploits the relationship between the Poisson distribution and exponential interarrival times in a homogeneous Poisson point process (see Sections 4.2.1 and 4.8) by counting numbers of exponential variates. However, the use of this algorithm is no longer recommended, even for relatively low values of θ, because it is so dependent on the quality and speed of the requisite uniform generator.

Search-from-the-origin is a method based on inversion of the cdf. A single uniform random number U is compared first with $\Pr[X = 0]$, next with $\Pr[X = 0] + \Pr[X = 1]$, and so on, until U is exceeded by the cumulative probability $\sum_{j=0}^{k} \Pr[X = j]$. (The cumulative probabilities are calculated as required using the current value of θ and the recursion formula $\Pr[X = j] = \Pr[X = j - 1]\theta/j$.) The value of k is then delivered as the generated Poisson variate. There are two versions of this algorithm, one using *buildup* of the probabilities and another using *chop-down* of U (see, e.g., C. D. Kemp and A. W. Kemp, 1991). This method is useful for very low values of θ ($\theta < 5$ say), but it is very time consuming for values of θ greater than, say, 20.

A. W. Kemp and C. D. Kemp's (1990) composition-search algorithm KPLOW is designed for low values of θ ($\theta < 30$ say). The algorithm decomposes the Poisson distribution into three components,

$$e^{\theta(z-1)} = \sum_{j=0}^{2r-1} \pi_j z^j + \sum_{j=0}^{2r-1} (p_j - \pi_j) z^j + \sum_{j \geq 2r} p_j z^j,$$

where r is the mode and π_j is a tight lower bound for p_j such that

$$\frac{\pi_j}{p_j} = \frac{\pi_r}{p_r} = \alpha < 1 \quad \text{for } j = 0, 1, \ldots$$

(the determination of α is discussed in the paper). The first component is searched bilaterally from the mode r. In the rare situation where this fails to return a variate, it is necessary to search one or both of the other two components. Computer timings indicate that this algorithm is noticeably faster than search-from-the-origin.

Schmeiser and Kachitvichyanukul (1981) gave a very clear account of their composition-rejection algorithm PTPE and discussed various earlier methods. Ahrens and Dieter (1982) called their algorithm KPOISS; it is an acceptance-complement algorithm based for the most part on the normal distribution. Schmeiser and Kachitvichyanukul (1981) gave timings for both algorithms, showing that they are both much faster than earlier algorithms.

C. D. Kemp and A. W. Kemp's (1991) algorithm KEMPOIS is based on Kemp's (1988b) modal approximations. It employs a unilateral search from the mode, with squeezing. For approximately $10 < \theta < 700$ it is faster than either PTPE or KPOISS.

Brown and Pallant (1992) described a method for obtaining antithetic variables, in particular, from Poisson distributions. Antithetic pseudorandom variables have negative correlation between successive variables, making them useful for variance reduction purposes. Brown and Pallant's method can create the full range of negative correlation.

4.7 ESTIMATION

4.7.1 Model Selection

The Poisson distribution provides the simplest model for discrete data with infinite support. This has led to the development of a number of graphical tests for its suitability as a model given an empirical data set.

Dubey (1966b) suggested plotting f_r/f_{r+1} against r, where f_r is an observed frequency, $r = 0, 1, \ldots$. For a Poisson population this plot should be a straight line with both intersect and slope equal to $1/\theta$.

Ord (1967a, 1972) found that $u_r = rf_r/f_{r-1}$ is a better diagnostic. When plotted against r, this should give a straight line with $u_r = c_0 + c_1 r$ for a number of discrete distributions; see Section 2.3.3. For the Poisson distribution the expected relationship is $u_r = \theta$. Ord suggested an "intuitively reasonable smoothing procedure" using $v_r = 0.5(u_r + u_{r+1})$ instead of u_r. Gart (1970) pointed out that the relationship $u_r = \theta$ also holds for a truncated Poisson distribution.

Grimm (1970) discussed the construction of Poisson probability paper using the property that $X^{1/2}$ is approximately normally distributed with mean $\theta^{1/2}$ and variance 0.25; see Section 4.5. His Poisson probability paper uses contours of $\Pr[X \leq c]$ plotted on a cumulative normal scale against θ on a square-root scale for $c = 0, 1, \ldots$. When cumulative relative frequencies are marked on the intersections of the horizontal cumulative probability lines and the contours for the appropriate values of c, the marked points should lie on a vertical line. Grimm discussed a further use of his paper for the graphical determination of $100(1-\alpha)\%$ confidence intervals.

Hoaglin, Mosteller, and Tukey (1985) developed a "Poissonness" plot based on

$$\ln(r!\Pr[X = r]) = -\theta + r\ln\theta.$$

A plot of $\ln(r!f_r/N)$, where $\sum_{j\geq 0} f_j = N$, against r should therefore have intercept $-\theta$ and slope $\ln\theta$. They suggested a method of "leveling" the Poisson plot by taking an assumed value θ_0 of θ. A plot of the count parameter $\ln(r!f_r/N) + \theta_0 - r\ln\theta_0$ against r should then have intercept $\theta_0 - \theta$ and slope $\ln(\theta/\theta_0)$. A good choice of θ_0 (e.g., $\overline{x}$) should give a nearly horizontal plot. Hoaglin et al. suggested a smoothing procedure for f_r and the superimposition of approximate significance levels for the count metameter. They considered their procedure to be more resistant against outliers and less prone to bias than a method based on the ratios of successive observed frequencies. However, their method lacks the simplicity of Ord's.

Rapid graphical tests for a number of small samples from a Poisson distribution (with possibly different values of θ) have been devised by Pettigrew and Mohler (1967) using the following result of Johnson and Young (1960). The approximate percentiles R_p of the conditional distribution of R, the range of a Poisson sample of size k, given the mean $\overline{x}$, are given by $R_p = w_p(\overline{x})^{1/2}$; w_p are the percentiles of the distribution of W, where W is the range of k independent unit-normal variates. If R is plotted on an arithmetic scale and $\overline{x}$ on a square-root scale, then the theoretical percentile lines form rays through the origin and the points for the various samples should lie within a wedge-shaped area. Alternatively, if logarithmic paper is used, then the percentile lines are parallel, since

$$\ln R_p = \ln w_p + 0.5 \ln \overline{x},$$

and the plotted points should lie within them.

4.7.2 Point Estimation

Only one parameter θ is used in defining a Poisson distribution. Hence there is only the need to estimate a single parameter, though different functions of this parameter may be estimated in different circumstances.

Given n independent rv's $X_1, X_2, \ldots, X_n$, each with distribution (4.1), the maximum likelihood estimator (MLE) of θ is

$$\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} X_j. \tag{4.50}$$

This is of course the first-moment estimate of θ. The sum $\sum_{j=1}^{n} X_j$ has a Poisson distribution with parameter $n\theta$; it is a complete sufficient statistic for θ. The variance of $\hat{\theta}$ is θ/n. This is equal to the Cramèr–Rao bound; $\hat{\theta}$ is the minimum variance unbiased estimator (MVUE) of θ.

When all the data are available, the MLE ($\hat{\theta}$) of θ is so easily calculated that it has, in practice, almost always been used. However, when some of the data are omitted or are inaccurately observed, the situation is less clear-cut. The most important special situation of this kind is that in which the zero class is not observed. A rv so obtained has the *positive Poisson* (*zero-truncated Poisson*) distribution; this will be discussed in Section 4.10.1 together with appropriate methods of estimation for θ.

The problem of estimating the mean of a Poisson distribution in the presence of a nuisance parameter has been examined by Yip (1988). Ahmed (1991) considered the estimation of θ given samples from two possibly identical Poisson populations.

An intuitive estimator for $\exp(-\theta)$, the "probability of the zero class," is $\exp(-\hat{\theta})$. This is a biased estimator—its expected value is $\exp[-\theta(1-e^{-1/n})]$. The MVUE of $e^{-\theta}$ is

$$T = \left(1 - \frac{1}{n}\right)^{n\hat{\theta}}; \tag{4.51}$$

the variance of this estimator is $e^{-2\theta}(e^{\theta/n} - 1)$. The mean-square error of T is less than that of $\exp(-\hat{\theta})$ for $e^{-\theta} < 0.45$ (Johnson, 1951). (Of course, if there is doubt whether the distribution of each X_j is really Poisson, it is safer to use the proportion of the X's that are equal to zero as an estimator of the probability of the zero class.)

The estimation of the Poisson probability (4.1) for general x was considered by Barton (1961) and Glasser (1962). They showed that

$$\tilde{P}(k;\theta) = \binom{n\overline{X}}{k}\left(\frac{1}{n}\right)^{k}\left(1 - \frac{1}{n}\right)^{n\overline{X}-k} \tag{4.52}$$

is the MVUE of $P(k;\theta) = e^{-\theta}\theta^k/k!$. It is easy to verify that $\tilde{P}(k;\theta)$ is unbiased by using the result that $n\overline{X}$ has a Poisson distribution with parameter $n\theta$; it has the minimum variance because $n\overline{X}$ is a complete sufficient statistic.

Similarly the MVUE for the probability associated with the interval $k_1 \le k \le k_2$ [i.e., $\sum_{k_1}^{k_2}(e^{-\theta}\theta^k/k!)$] is $\sum_{k_1}^{k_2}\tilde{P}(k;\theta)$.

Estimation of the *ratio* of expected values θ_1, θ_2 of two Poisson distributions based on observed values of independent rv's that have such distributions has been described by Chapman (1952b). He showed that there is no unbiased estimator of the ratio θ_1/θ_2 with finite variance but that the estimator $x_1/(x_2+1)$ (in an obvious notation) is "almost unbiased."

The rationale of Bayesian inference for the Poisson distribution has been explained in detail by Irony (1992).

Empirical Bayes (EB) estimation for the Poisson distribution has been investigated by a number of authors. Robbins (1956), also Good (1953), developed an EB method for estimating θ that is known as the frequency-ratio method. Here x is the current observation, there are n past observations, and $f_n(x)$ denotes the number of past observations having the value x; this gives $(x+1)f_n(x+1)/[1+f_n(x)]$ as the EB estimator of θ. The frequency ratio estimator, graphed as a function of x, is far from smooth, however. Maritz (1969) reviewed and extended various smoothing procedures that had previously been proposed, and he made a careful assessment of their comparative performance. A thorough account of EB point estimation for the Poisson distribution, with applications, is in Maritz and Lwin (1989).

Sadooghi-Alvandi (1990) examined the estimation of θ from a sample of n observations using a LINEX loss function of the form

$$L(\theta_B,\theta) = b[e^{a(\theta_B-\theta)} - a(\theta_B-\theta) - 1], \qquad b \ge 0,\ a \ne 0, \tag{4.53}$$

introduced by Varian (1975); see also Zellner (1986). Given a gamma prior distribution with pdf $f(\theta|\gamma, \lambda) = \lambda^\gamma \theta^{\gamma-1} e^{-\lambda\theta}/\Gamma(\gamma)$, where $0 \le \theta < \infty$ and $\gamma > 0$, $\lambda > 0$, then, provided that $\gamma + n + a \ge 0$, the unique Bayesian estimate of θ, relative to (4.53), is

$$\theta_B = \left[\frac{n}{a} \ln\left(\frac{\lambda + n + a}{\lambda + n}\right)\right] \left(\overline{x} + \frac{\gamma}{n}\right). \tag{4.54}$$

Bayesian methods for the simultaneous estimation of the means of several independent Poisson distributions have also attracted attention. Clevenson and Zidek (1975) proposed a generalized Bayes method where the loss function is the sum of the component standardized squared-error losses. They showed that the method is minimax and also (both numerically and asymptotically) that it is better than using the sample means as estimators of the corresponding population means. Peng (1975) investigated the use of the aggregate squared-error loss function $\sum_i (\theta_i^* - \theta_i)^2$. This approach was extended by Hudson and Tsui (1981); see also Hwang (1982) and Tsui and Press (1982). Simultaneous estimation of Poisson means under entropy loss has been studied by Ghosh and Yang (1988) and Yang (1990).

Bayesian inference concerning the size of the zero class given data from the zero-truncated Poisson distribution was studied by Scollnik (1997).

Gupta and Liang (2002) are concerned with the Bayesian selection of the Poisson population with the smallest failure rate, given samples from a number of competing Poisson distributions; their proviso is that the chosen distribution should have a failure rate less than a specified control level.

Two tests for comparing two Poisson means, $\overline{X_1}$ and $\overline{X_2}$, are researched by Krishnamoorthy and Thomson (2003); their E-test (based on estimated p values) is shown to be more powerful than the C-test (based on the conditional distribution of $\overline{X_1}$ given $\overline{X_1} + \overline{X_2} = k$).

4.7.3 Confidence Intervals

Since the Poisson distribution is a discrete distribution, it is difficult to construct confidence intervals for θ with an exactly specified confidence coefficient of, say, $100(1 - \alpha)\%$. *Approximate* $100(1 - \alpha)\%$ confidence limits for θ given an observed value x of X, where X has the distribution (4.1), are obtained by solving the equations

$$\exp(-\theta_L) \sum_{j=x}^{\infty} \frac{\theta_L^j}{j!} = \frac{\alpha}{2}, \qquad \exp(-\theta_U) \sum_{j=0}^{x} \frac{\theta_U^j}{j!} = \frac{\alpha}{2} \tag{4.55}$$

for θ_L, θ_U, respectively, and using the interval (θ_L, θ_U).

From the relationship between the Poisson and the χ^2 distributions (Section 4.12.2), these equations can be written as

$$\theta_L = 0.5\chi^2_{2x,\alpha/2}, \qquad \theta_U = 0.5\chi^2_{2(x+1),1-\alpha/2}. \tag{4.56}$$

So the values of θ_L and θ_U can be found by interpolation (with respect to the number of degrees of freedom) in tables of percentage points of the central χ^2 distribution.

If θ is expected to be fairly large, say greater than 15, a normal approximation to the Poisson distribution might be used. Then

$$\Pr[|X-\theta| < u_{\alpha/2}\sqrt{\theta}|\theta] = 1-\alpha, \tag{4.57}$$

where

$$(2\pi)^{-1/2}\int_{u_{\alpha/2}}^{\infty} e^{-u^2/2}\,du = \frac{\alpha}{2}.$$

From (4.57)

$$\Pr[\theta^2 - \theta(2X + u_{\alpha/2}^2) + X^2 < 0|\theta] = 1-\alpha,$$

that is,

$$\begin{aligned}\Pr[X &+ 0.5u_{\alpha/2}^2 - u_{\alpha/2}(X+0.25u_{\alpha/2}^2)^{1/2}\\ &< \theta < X + 0.5u_{\alpha/2}^2 + u_{\alpha/2}(X+0.25u_{\alpha/2}^2)^{1/2}|\theta] \approx (1-\alpha).\end{aligned}$$

The limits

$$X + 0.5u_{\alpha/2}^2 \pm u_{\alpha/2}\sqrt{X+0.25u_{\alpha/2}^2}$$

thus enclose a confidence interval for θ with confidence coefficient approximately equal to $100(1-\alpha)\%$.

Molenaar (1970b, 1973) made recommendations concerning approximations for confidence limits for θ; he also cited other relevant references. Values of approximate 95% confidence limits for θ, given an observed value of X, are given by Mantel (1962). The "limit factors" shown must be *multiplied* by the observed value of X to obtain limits for θ. There is a similar table published by the Society of Actuaries (1954).

Tables for θ_L and θ_U are contained in Pearson and Hartley (1976). Crow and Gardner (1959) constructed a modified version of these tables, so that (1) the confidence belt is as narrow as possible, *measured in the direction of the observed variable*, and (2) among such narrowest belts, it has the smallest possible upper confidence limits. Condition 2 reduces, in particular, the width of confidence intervals corresponding to small values of the observed variable.

Let X_1 and X_2 be two independent Poisson rv's with parameters θ_1 and θ_2, respectively. Let $F_{m,n}(X)$ represent the distribution function of the ratio of two independent χ^2 variables χ_{2m}^2/χ_{2n}^2 with $2m$ and $2n$ degrees of freedom. Bol'shev (1965) showed that for any α, where $0 \le \alpha < 1$, the solution of the equation

$$F_{x_1,x_2+1}(X) = \alpha$$

satisfies the inequality

$$\inf_{\theta_1,\theta_2} \Pr\left[X < \frac{\theta_1}{\theta_2}\right] \geq 1 - \alpha.$$

Thus X can be taken to be the lower bound of a confidence interval for θ_1/θ_2 with minimal confidence coefficient $1 - \alpha$. Similarly the solution of

$$F_{x_1+1,x_2}(Y) = 1 - \alpha$$

satisfies the inequality

$$\inf_{\theta_1,\theta_2} \Pr\left[Y > \frac{\theta_1}{\theta_2}\right] \geq 1 - \alpha,$$

and thus Y can be taken to be the upper bound of a confidence interval for θ_1/θ_2 with minimal confidence coefficient $1 - \alpha$. These results are useful for testing the hypothesis $\theta_1/\theta_2 < c$, where c is some constant (and similarly for the hypothesis $\theta_1/\theta_2 > c$).

Chapman (1952b) used the fact that the conditional distribution of X_1, given $X_1 + X_2 = x$, is binomial with parameters x and $\theta_1/(\theta_1 + \theta_2)$; see Section 4.8. Forming approximate confidence intervals for $\theta_1/(\theta_1 + \theta_2) = (1 + \theta_2/\theta_1)^{-1}$ is then equivalent to doing the same for θ_1/θ_2.

Casella and Robert (1989) were critical of the method of randomized Poisson confidence intervals of Stevens (1957) and Blyth and Hutchinson (1961), where random noise is added to the data to obtain exact intervals. They introduced instead the concept of refined confidence intervals; this is a numerical procedure whereby the end points of a confidence interval are shifted to produce the smallest possible interval. Their paper is particularly concerned with confidence intervals for the Poisson distribution. They also considered certain other discrete distributions. Their method becomes very computer intensive.

Confidence intervals for the ratio of two Poisson variables were considered by Sahai and Khurshid (1993a). In Sahai and Khurshid (1993b) they gave a comprehensive review of confidence intervals for the Poisson parameter.

The coverage and length of new "short" confidence intervals are explored in Kabaila and Byrne (2001) and Byrne and Kabaila (2001).

4.7.4 Model Verification

If a data set comes from a Poisson distribution, its variance should be equal to its mean. This can be tested by the index-of-dispersion test, with $(N - 1)s^2/\overline{x}$ having approximately a χ^2 distribution with $N - 1$ degrees of freedom. Conditions under which the variance test is optimal for testing the homogeneity of a complete Poisson sample have been examined by Neyman and Scott (1966), Potthof and Whittinghill (1966), and Moran (1973); see also Gart (1974).

Tests for a fully specified Poisson distribution, based on the empirical distribution function, are discussed by Stephens (1986). The best known of these is Pearson's χ^2 goodness-of-fit test [see, e.g., Tallis (1983) for details and further references]. In particular, Stephens discusses Kolmogorov–Smirnov goodness-of-fit tests; see also Wood and Altavela (1978).

Goodness-of-fit tests for the Poisson distribution that are based on Cramér–von Mises statistics were investigated by Spinelli and Stephens (1997). Power studies indicated that they are good overall tests of fit. Spinelli and Stephens were particularly interested in a test for situations where the variance is close to the mean. Rueda and O'Reilly (1999) related their own test for goodness of fit to that of Spinelli and Stephens and provided power comparisons in the case of the Poisson distribution.

The goodness-of-fit test for the Poisson distribution put forward by Kyriakoussis, Li, and Papadopoulos (1998) is closely related to Fisher's index of dispersion. The equality of the mean and variance characterizes the Poisson distribution and hence $c = \mu'_{[2]}/\mu^2 = 1$. The moment estimate of c is $\tilde{c} = (s^2 - \overline{x} + \overline{x}^2)/\overline{x}^2$. Kyriakoussis et al. show that the test statistic

$$T_p = \sqrt{\frac{m}{2}}\,(\tilde{c} - 1)\overline{x} = \sqrt{\frac{m}{2}}\left(\frac{s^2}{\overline{x}} - 1\right),$$

where m is the sample size, has asymptotically an $N(0, 1)$ distribution. They develop analogous test statistics for the binomial and negative binomial distributions.

How to pool the data is a problem with the usual χ^2 goodness-of-fit test. Best and Rayner (1999) develop alternative smooth tests of fit for the Poisson distribution using the methodology in Rayner and Best (1989). They give power comparisons between the χ^2-test, their smooth tests, and a test based on a modified Kolmogorov–Smirnov statistic.

Gürtler and Henze (2000) discussed in detail a number of recent and classical tests of goodness of fit for the Poisson distribution, including two kinds of weighted Cramér–von Mises tests based on the empirical pgf. They carried out a simulation study for these and other tests, applied all the tests to four well-known data sets, and provided a useful bibliography of tests for the Poisson distribution.

4.8 CHARACTERIZATIONS

There has been much work concerning characterizations of the Poisson distribution. Work prior to 1974 is summarized in Kotz (1974).

Raikov (1938) showed that, if X_1 and X_2 are independent rv's and $X_1 + X_2$ has a Poisson distribution, then X_1 and X_2 must each have Poisson distributions. (A similar property also holds for the sum of any number of independent Poisson rv's.)

Kosambi (1949) and Patil (1962a) established that, if X has a power series distribution, then a necessary and sufficient condition for X to have a Poisson distribution is $\mu'_{[2]} = \mu^2$, where $\mu'_{[2]}$ is the second factorial moment. This implies that $\mu_2 = \mu$ characterizes the Poisson distribution among PSDs; see also Patil and Ratnaparkhi (1977). Gokhale (1980) gave a strict proof of this important property of the Poisson distribution and proved furthermore that, within the class of PSDs, the condition

$$\mu_2 = a\mu + b\mu^2, \quad \text{given } a + b\mu > 0,$$

holds iff the distribution is Poisson or binomial or negative binomial. Gupta (1977b) proved that the equality of the mean and variance characterizes the Poisson distribution within the wider class of modified PSDs, provided that the series function $f(\theta)$ satisfies $f(0) = 1$.

Moran (1952) discovered a fundamental property of the Poisson distribution. If X_1 and X_2 are independent nonnegative integer-valued rv's such that the conditional distribution of X_1 given the total $X_1 + X_2$ is a binomial distribution with a common parameter p for all given values of $X_1 + X_2$ and if there exists at least one integer i such that $\Pr[X_1 = i] > 0$, $\Pr[X_2 = i] > 0$, then X_1 and X_2 are both Poisson rv's.

Chatterji (1963) showed that, if X_1 and X_2 are independent nonnegative integer-valued rv's and if

$$\Pr[X_1 = x_1 | X_1 + X_2 = x] = \binom{x}{x_1} p_x^{x_1} (1 - p_x)^{x - x_1}, \qquad x_1 = 0, 1, \ldots, x, \tag{4.58}$$

then it follows that:

1. p_x does not depend on x but equals a constant p for all values of x;
2. X_1 and X_2 each have Poisson distributions with parameters in the ratio $p : (1 - p)$.

This characterization has been extended by Bol'shev (1965) to n variables X_i, $i = 1, \ldots, n$. The condition is that the conditional distribution of $X_1, X_2, \ldots, X_n$ given $\sum X_i$ is multinomial; see also Volodin (1965).

Bol'shev suggested that this property might be used to generate Poisson-distributed rv's for a number of Poisson distributions using only a single original Poisson variate (with a large expected value) that is split up according to a multinomial with fixed cell probabilities; Brown and Bromberg (1984) developed this idea further. Patil and Seshadri (1964) showed that Moran's characterization is a particular case of a more general characterization for a pair of independent rv's X_1 and X_2 with specified conditional distribution of X_1 given $X_1 + X_2$. Further extensions to the Moran characterization were given by Janardan and Rao (1982) and Alzaid, Rao, and Shanbhag (1986). For comments on a suggested extension to truncated rv's, see Panaretos (1983b).

Lukacs (1956) showed that, if $X_1, X_2, \ldots, X_n$ is a random sample from some distribution, then the distribution from which the sample is taken is Poisson with parameter θ iff the statistic $n\overline{x}$ has a Poisson distribution with parameter $n\theta$.

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a distribution and suppose that r and s are two positive integers. Assume also that the $(r+s)$th moment of the distribution exists, that the cdf is zero for $x < 0$, and that the cdf is greater than zero for $x \geq 0$. Then Lukacs (1965) showed that the distribution is Poisson iff the k-statistic $k_{r+s} - k_r$ has a constant regression on $k_1 = \overline{x}$, that is, iff $E[k_{r+s} - k_r | k_1] = 0$.

Daboni (1959) gave the following characterization in terms of mixtures of binomial distributions. Suppose that X is a rv distributed as a mixture of binomial distributions with parameters N, p. Then X and $N - X$ are independent iff N has a Poisson distribution.

Rao and Rubin (1964) obtained the following characterization of the Poisson distribution. If X is a discrete rv taking only nonnegative integer values and the conditional distribution of Y given $X = x$ is binomial with parameters x, p (p not depending on x), then the distribution of X is Poisson iff

$$\Pr[Y = k | Y = X] = \Pr[Y = k | Y \neq X].$$

Rao (1965) suggested the following physical basis for the model described above. X represents a "naturally" occurring quantity, which is observed in such a way that some of the components of X may not be counted. Y represents the value remaining (and actually observed) after this "destructive process." That is, suppose that an original observation is distributed according to a Poisson distribution with parameter θ and that the probability that the original observation n is reduced to r due to a destructive process is

$$\binom{n}{r} \pi^r (1 - \pi)^{n-r}, \qquad 0 \leq \pi \leq 1.$$

If Y denotes the resultant rv, then

$$\begin{aligned} \Pr[Y = r] = \Pr[Y = r | \text{undamaged}] &= \Pr[Y = r | \text{damaged}] \\ &= \frac{e^{-\theta\pi} (\theta\pi)^r}{r!}; \end{aligned} \tag{4.59}$$

furthermore condition (4.59) characterizes the Poisson distribution.

Rao and Rubin's (1964) characterization has generated a lot of attention. For instance, Srivastava and Srivastava (1970) showed that if the original observations have a Poisson distribution and if condition (4.59) is satisfied, then the destructive process is binomial.

Let X be a discrete rv with support $0, 1, \ldots$; suppose also that, if a random observation $X = n$ is obtained, then n Bernoulli trials with probability p of success are performed. Let Y and Z be the resultant numbers of successes and failures, respectively. Then Srivastava (1971) showed that X has a Poisson distribution iff Y and Z are independent.

An alternative, simpler proof of the Rao–Rubin theorem was given by Wang (1970). Wang also introduced the idea of binomial *splitting*, that is, compounding X with the binomial distribution

$$\binom{n}{r}\pi^r(1-\pi)^{n-r}, \qquad 1 \le \pi \le 1,$$

and the similarly defined notion of binomial *expanding*. He showed that the Poisson distribution is the only one that remains invariant under binomial splitting or expanding.

The Rao–Rubin characterization was extended to a pair of independent Poisson variables (and also to a multivariate set of Poisson variables) by Talwalker (1970). This extension was conjectured independently by Srivastava and Srivastava (1970) as follows.

Let $X = \binom{X_1}{X_2}$ be a non-degenerate, discrete random vector such that X_1, X_2 take nonnegative integer values. Let

$$\binom{n_1}{r_1}\pi_1^{r_1}\phi_1^{n_1-r_1} \qquad \text{and} \qquad \binom{n_2}{r_2}\pi_2^{r_2}\phi_2^{n_2-r_2}$$

be the independent probabilities that the observations n_1, n_2 on X_1, X_2 are reduced to r_1, r_2, respectively, during the destructive process. Let $Y = \binom{Y_1}{Y_2}$ denote the resultant random vector where Y_1 and Y_2 take the values $0, 1, 2, \ldots$; let $r = \binom{r_1}{r_2}$ denote the vector with elements r_1 and r_2.

Then

$$\begin{aligned}\Pr[Y = r] &= \Pr[Y = r|\text{ damaged}]\\ &= \Pr[Y = r|x_1 \text{ damaged}, \ x_2 \text{ undamaged}]\\ &= \Pr[Y = r|x_1 \text{ undamaged}, \ x_2 \text{ damaged}]\\ &= \Pr[Y = r|\text{ undamaged}]\end{aligned}$$

iff X has the double Poisson distribution with pgf

$$\exp\{\theta_1(z_1 - 1) + \theta_2(z_2 - 1)\}.$$

An elementary (and elegant) proof was given by Shanbhag (1974); see also Aczél (1972) and van der Vaart (1972). Aczél (1975) studied the problem under weaker conditions. Note that the simpler proof by Wang (1970) of the Rao–Rubin characterization holds also in the multivariate case.

Talwalker (1975, 1980) and Rao et al. (1980) have studied the characterization of the binomial, Poisson, and negative binomial trio of distributions via the generalized Rao–Rubin condition.

Let Z denote the initial random variable and let X denote the number of survivors. Assume that π and π' are the probabilities of survival for two different dosages. Then the initial distribution can be shown to be negative binomial, Poisson or binomial if

$$\begin{aligned}&\Pr[X = r|Z \text{ damaged, when probability of survival is } \pi]\\ &\qquad = \Pr[X = r|Z \text{ undamaged, when probability of survival is } \pi'],\end{aligned}$$

where $\pi' = \pi/(1 + dqa)$, $0 < p < 1$, $q = 1 - p$, $0 < a < 1$, and $d = -1, 0, +1$ for the negative binomial, Poisson, and binomial distributions, respectively.

Krishnaji (1974) obtained the following result: If $Y = X_1 + X_2$ and

$$\Pr[X_1 = x_1|Y = y] = \binom{y}{x_1} p^{x_1}(1 - p)^{y-x_1},$$

then $E[X_1|X_2] = \theta p$ iff Y has a Poisson distribution with expected value θ.

Another variant of the Rao–Rubin theorem has been proved by Shanbhag and Clark (1972). Let X have a power series distribution with pmf

$$\Pr[X = x] = \frac{a_x\theta^x}{A(\theta)}, \qquad x = 0, 1, \ldots, \tag{4.60}$$

and assume that the conditional distribution of Y given that $X = n$ has pmf $\Pr[Y = r|X = n] = g(r; n)$, $r = 0, 1, \ldots, n$, with mean and variance $n\pi$ and $n\pi(1 - \pi)$, where π does not depend on θ. Then

$$E[Y] = E[Y|Y = x] \qquad \text{and} \qquad \operatorname{Var}(Y) = \operatorname{Var}(Y|Y = x)$$

iff $\{P_x\}$ is Poisson and $g(n; n) = \pi^n$. [Note that, if $g(r; n)$ is binomial, then its mean and variance are $n\pi$ and $n\pi(1 - \pi)$, respectively.]

Further results relating to the Rao–Rubin characterization have been obtained by Shanbhag (1974), Shanbhag and Rajamannar (1974), Srivastava and Singh (1975), Shanbhag and Panaretos (1979), Gupta (1981), and Kourouklis (1986).

Korwar (1975) established that, if X and Y are nonnegative integer-valued rv's such that, for all x,

$$\Pr[Y = y|X = x] = \binom{x}{y} p^y(1 - p)^{x-y}, \qquad y = 0, 1, \ldots, x,$$

where p is some constant such that $0 < p < 1$, not dependent on x, then X has a Poisson distribution iff

$$\Pr[Y = y] = \Pr[Y = y|X = x] = \Pr[Y = y|X > Y]$$

[cf. (4.59)].

The investigations into empirical Bayes estimation for the Poisson distribution by Maritz (1969) have led to a different kind of characterization theorem. Shanbhag and Clark (1972) showed that if X has a PSD (4.60) and if θ is a rv with pdf

$$f(\theta) = k \sum_{j=0}^{\infty} b_i \theta^i,$$

where $b_i \geq 0$, $0 < \theta < c$, and c may be infinite, then $\beta(x) = E[\theta|X = x]$ is linear iff $f(\theta)$ is exponential and X has a Poisson distribution.

Characterizations based on properties of sample statistics were also developed by Kharshikar (1970) and by Shanbhag (1970a). Kharshikar showed that, if X belongs to a power series family, then the condition $E[s^2/\overline{x}\,|\,\overline{x} > 0] = 1$ (where $\overline{x}$ and s^2 are the sample mean and variance statistics) characterizes a Poisson distribution. Shanbhag proved a more general version of this theorem involving a quadratic form in the sample observations. The modification of Shanbhag's result by Wang (1972) was later shown by Shanbhag (1973) to need amendment.

Another sort of characterization (one based on a Bhattacharya covariance matrix) was obtained by Shanbhag (1972). This one would seem to have limited practical use, as it is valid also for the binomial and negative binomial distributions.

Boswell and Patil (1973) showed that the following four statements about a discrete rv $X(\theta)$, $\theta \geq 0$, $X(0) \equiv 0$, are equivalent:

1. There exists a real a such that $\Pr[X(\theta) \geq a] = 1$ and $\partial p_x/\partial\theta = p_{x-1} - p_x$, $\theta > 0$.
2. $\sum_{r=0}^{x} \partial p_r/\partial\theta = -p_x$, $\theta > 0$.
3. For any function $f(X(\theta))$ that has finite expectation,
$$\frac{\partial E[f(X(\theta))]}{\partial\theta} = E[f(X(\theta) + 1) - f(X(\theta))].$$
4. $X(\theta)$ has a Poisson distribution, $\theta > 0$.

Here $p_x = \Pr[X(\theta) = x]$. Boswell and Patil also gave a more general formulation that allows $X(\theta)$ to be either discrete or continuous.

Haight (1972) attempted to unify certain isolated results via Svensson's theorem (Svensson, 1969). This states that, if X has a distribution on the nonnegative integers, then there exists a two-dimensional rv (X, Y) with the property that the pgf of $(X|X + Y)$ is of the form

$$\phi(z) = (1 - p + pz)^n$$

iff

$$\pi_{X+Y}(z) = \pi_X\left(\frac{z - 1 + p}{p}\right),$$

where π_Y is the pgf of Y. Haight saw that this theorem permits the characterization of certain distributions of the sum of two rv's, omitting in general the assumption of independence. For instance, Chatterji's (1963) characterization theorem (4.58) is a straightforward consequence of Svensson's theorem.

Samaniego (1976) termed $Y = X_1 + X_2$ a "convoluted Poisson variable" if X_1 and X_2 are independent and X_1 has a Poisson distribution. He obtained characterizations of convoluted Poisson distributions.

The well-known characterization in terms of the conditional (multinomial) distribution of independent variables $X_1, \ldots, X_n$ given $\sum_{j=1}^{n} X_j = z$ was extended to zero-truncated X variables by Singh (1978). A characterization based on a multivariate splitting model was obtained by Rao and Srivastava (1979); another characterization based on this model was given by Ratnaparkhi (1981).

In a lengthy and detailed account of characterizations of the *Poisson process*, Galambos and Kotz (1978) gave a number of fresh results; they related these where possible to the work of previous authors. Their researches were especially concerned with characterizations based on age and residual life, on rarefactions of renewal processes, on geometric compounding, and on damage models. They were motivated by characterizations of the exponential distribution.

Gupta and Gupta (1986) commented that several authors had obtained characterizations of the exponential distribution, and hence of the Poisson, by certain properties of $u(t)$ and $v(t)$, where $u(t)$ is the age of a component in use and $v(t)$ is its remaining life (from t to failure). Notably, if $E[v(t)] < \infty$ and if $E[v(t)]$ is independent of t for all t, then the process is Poisson; see Gupta and Gupta (1986) for references. Gupta and Gupta extended this characterization by showing that, under mild conditions, if $E[G(v(t))]$ is constant, then the process is Poisson; here $G(x)$ is a monotone nondecreasing function of x having support on $x \geq 0$ with $G(0) = 0$ and $\int_0^\infty \exp(-\xi x)\, dG(x) < \infty$ for all $\xi > 0$.

A characterization of the Poisson process by properties of conditional moments was given by Bryc (1987) and extended by Wesolowski (1988).

The following characterization is due to Rao and Sreehari (1987). A nonnegative integer-valued rv X has a Poisson distribution iff

$$\sup\left(\frac{\mathrm{Var}(h(X))}{\mathrm{Var}(X)E[\{h(X+1)-h(X)\}^2]}\right) = 1, \tag{4.61}$$

where the supremum is taken over all real-valued functions $h(\cdot)$ such that $E[\{h(X+1)-h(X)\}^2]$ is finite.

Characterizations of the Poisson distribution based on discrete analogs of the Cramèr–Wold and Skitovich–Darmois theorems have been derived by McKenzie (1991).

Diaconis and Zarbell (1991) pointed out that the Poisson distribution is characterized by the identity

$$E[\theta f(X+1)] = E[Xf(X)]$$

for every bounded function $f(\cdot)$ on the integers. For the background to this identity, see Stein (1986, Chapter 9).

Sapatinas (1996) gives a detailed account of characterizations based on the generalized Rao–Rubin condition and uses variants of it to characterize certain compound Poisson distributions and hence the Poisson distribution.

In a study of modeling stationary Markov processes Sreehari and Kalamkar (1997) derive characterizations of the Poisson, exponential, and geometric distributions.

Kagan (2001) defines *right-side positive* (RSP) for a nonnegative integer-valued rv X to mean that $p_X(x) > 0 \Rightarrow p_X(x+1) > 0$. He takes

$$J_X(x) = 1 - \frac{p_X(x-1)}{p_X(x)} \quad \text{if } p_X(x) > 0$$

and $J_X(x) = 0$ if $p_X(x) = 0$. Then $I_X = E[J_X^2]$ is a discrete version of the Fisher information. Kagan proves that, if $X \in$ RSP and $Y \in$ RSP, then $Z = X + Y \in$ RSP, and if $I_x < \infty$, $I_Y < \infty$, then

$$\frac{1}{I_Z} \geq \frac{1}{I_X} + \frac{1}{I_Y};$$

this is a discrete version of the Stam inequality for Fisher information. Kagan then proves that the equality sign holds iff if X, Y, and Z have Poisson distributions, possibly shifted to other support on the nonnegative integers.

4.9 APPLICATIONS

The Poisson distribution has been described as playing a "similar role with respect to discrete distributions to that of the normal for absolutely continuous distributions" (Douglas, 1980, p. 5). It is used (1) as an approximation to the binomial and other distributions, (2) when events occur randomly in time or space, (3) in certain models for the analysis of contingency tables, and (4) in the empirical treatment of count data.

In many elementary textbooks it is introduced as a limit, and hence as an approximation, for the binomial distribution when the occurrence of an event is rare and there are many trials. This approach underlies the use of the Poisson distribution in quality control for the number of defective items per batch (see, e.g., Walsh, 1955; van der Waerden, 1960; and Chatfield, 1983). The Poisson is also a limiting form for the hypergeometric distribution (Section 6.5) and hence provides an approximation for sampling without replacement. Feller (1968, p. 59) pointed out the relevance of the Poisson limiting form for Boltzmann–Maxwell statistics in quantum statistics and in the theory of photographic plates.

Early writers on the analysis of quadrat data [e.g., Greig-Smith, 1964] justified the use of the Poisson distribution for such data via the relationship between the Poisson distribution and the binomial distribution [see also Seber (1982b, p. 24)].

Quadrat data have been collected extensively in ecology, geology, geography, and urban studies. A more fundamental explanation for the widespread applicability of the Poisson distribution for counts per unit of space or volume is the Poisson process [see, e.g., Cliff and Ord, 1981].

The Poisson process also has great importance concerning counts of events per unit of time, particularly in queueing theory. Here the intervals between successive events have independent identical exponential distributions, and consequently the number of events in a specified time interval has a Poisson distribution; see Section 4.2.1. Consider, for example, the problem of finding the number of telephone channels in use (or customers waiting in line, etc.) at any one time. Suppose that there are "infinitely many" (actually, a large number of) channels available, that the holding time (length of call) T for each call has the exponential distribution with pdf

$$p(t) = \phi^{-1} e^{-t/\phi}, \qquad \phi > 0, \qquad t > 0,$$

and that incoming calls arrive at times following a Poisson process, with the average number of calls per unit time equal to θ. Then the function $P_N(\tau)$, representing the probability that exactly N channels are being used at time τ after no channels are in use, satisfies the differential equations

$$\begin{aligned} P_0'(\tau) &= -\theta P_0(\tau) + \phi^{-1} P_1(\tau), \\ P_N'(\tau) &= -(\theta + N\phi^{-1}) P_N(\tau) + \theta P_{N-1}(\tau) \\ &\quad + (N+1)\phi^{-1} P_{N+1}(\tau), \qquad N \geq 1. \end{aligned} \tag{4.62}$$

The "steady-state" probabilities [$\lim_{\tau\to\infty} P_N(\tau) = P_N$] satisfy the equations

$$\begin{aligned} \theta P_0 &= \phi^{-1} P_1, \\ (\theta + N\phi^{-1}) P_N &= \theta P_{N-1} + (N+1)\phi^{-1} P_{N+1}. \end{aligned} \tag{4.63}$$

On solving these equations, it is found that

$$P_N = \frac{e^{-\theta\phi}(\theta\phi)^N}{N!}, \qquad N = 0, 1, \ldots, \tag{4.64}$$

which is (4.1) with θ replaced by $\theta\phi$ [see, e.g., Gnedenko and Kovalenko, 1989].

There are many excellent texts on stochastic processes that devote considerable attention to the Poisson process. Doob (1953) mentions its use for molecular and stellar distributions, while Parzen (1962) gives applications to particle counters, birth processes, renewal processes, and shot noise. Taylor and Karlin's (1998) introductory text is markedly application oriented, with applications in, inter alia, engineering, biosciences, medicine, risk theory, commerce, and demography. The Poisson process for entities in space has been discussed in depth by Ripley (1981) and Stoyan, Kendall, and Mecke (1987).

Poisson regression, that is, the analysis of the relationship between an observed count and a set of explanatory variables, is described by Koch, Atkinson, and Stokes (1986) in a very clear introductory article. Its use in the analysis of binary data is discussed by Cox and Snell (1989); see especially pages 146–147 for the interrelationships between Poisson and multinomial models. This relationship is examined by Sandland and Cormack (1984), Cormack (1989), and Cormack and Jupp (1991) in relation to their work on the loglinear analysis of capture–recapture data. Its role in the analysis of frequencies when summarizing data sets is described lucidly by Bishop, Fienberg, and Holland (1975). Fienberg (1982) gave helpful pointers to the literature about the treatment of cross-classified categorical data in general. Applications of Poisson regression include bioassay, counts of colonies of bacteria or viruses for varying dilutions and/or experimental conditions, equipment malfunctions for varying operational conditions, cancer incidence, and mortality and morbidity statistics. Applications to economic problems are described by Hausman et al. (1984). Lee (1986) has cited a number of other references; he was particularly concerned with testing a Poisson model against other discrete models, including the negative binomial.

In the empirical treatment of count data the Poisson distribution is often used as a yardstick to assess the degree and nature of nonrandomness. Mixed Poisson distributions and Poisson cluster distributions have been extensively developed as tools for dealing with nonhomogeneous and clumped data; see Chapters 8 and 9.

Chapter 7 in Haight (1967) gives a multitude of references concerning uses of the Poisson distribution, categorized by industry, agriculture and ecology, biology, medicine, telephony, accidents, commerce, queueing theory, sociology and demography, traffic flow theory, military, particle counting, and miscellaneous.

Other applications of the Poisson distribution include, for example, its use for sister chromatid exchanges in the study of DNA breakage and reunion (Margolin et al., 1986). Goldstein (1990) applied it to DNA sequence matching. The annual volumes of the *Current Index to Statistics* and the quarterly issues of *Statistical Theory and Method Abstracts* (now available on CD ROM) give leads to other recent uses, both applied and theoretical.

4.10 TRUNCATED AND MISRECORDED POISSON DISTRIBUTIONS

4.10.1 Left Truncation

The commonest form of truncation is the omission of the zero class; this occurs if the observational apparatus becomes active only when at least one event occurs. The pmf of the corresponding truncated Poisson distribution is

$$\Pr[X = x] = \frac{(1 - e^{-\theta})^{-1} e^{-\theta}\theta^x}{x!} = \frac{(e^{\theta} - 1)^{-1}\theta^x}{x!}, \qquad x = 1, 2, \ldots. \tag{4.65}$$

This is usually called the *zero-truncated* or *positive Poisson* distribution. Cohen (1960d) called it a *conditional Poisson* distribution.

The rth factorial moment of X is

$$E[X(X-1)\cdots(X-r+1)] = (1-e^{-\theta})^{-1}\theta^r. \tag{4.66}$$

The mean and variance are

$$E[X] = \frac{\theta}{1-e^{-\theta}}, \qquad \text{Var}[X] = \frac{\theta}{1-e^{-\theta}} - \frac{\theta^2 e^{-\theta}}{(1-e^{-\theta})^2}; \tag{4.67}$$

the distribution is therefore underdispersed. The expected value of X^{-1} is

$$E[X^{-1}] = \frac{1}{e^{\theta}-1}\sum_{j=1}^{\infty}\frac{\theta^j}{j!j} \tag{4.68}$$

(Grab and Savage, 1954). Tiku (1964) showed that

$$E[X^{-1}] \approx [(\theta-1)(1-e^{-\theta})]^{-1} \tag{4.69}$$

for sufficiently large θ. Grab and Savage (with addenda in the same volume) gave tables of $E[X^{-1}]$ to five decimal places.

The MLE $\hat{\theta}$ of θ, given observed values of n independent rv's $X_1, X_2, \ldots, X_n$ each having the same positive Poisson distribution, satisfies the equation

$$\overline{x} = n^{-1}\sum_{j=1}^{n} x_j = \frac{\hat{\theta}}{1-e^{-\hat{\theta}}}. \tag{4.70}$$

Equation (4.70) may be solved numerically; the process is quite straightforward. David and Johnson (1952) gave tables of the function $\hat{\theta}/(1-e^{-\hat{\theta}})$. Although Irwin (1959) derived an explicit expression for the solution of (4.70), namely

$$\hat{\theta} = \overline{x} - \sum_{j=1}^{\infty}\frac{j^{j-1}}{j!}(\overline{x}e^{-\overline{x}})^j, \tag{4.71}$$

the methods suggested above are usually easier to apply. Kemp and Kemp (1988) have shown that

$$\frac{6(\overline{x}-1)}{\overline{x}+2} < \hat{\theta} < \overline{x}[1-(e^{\overline{x}}-\overline{x})^{-1}]. \tag{4.72}$$

For large values of n, the variance of the MLE $\hat{\theta}$ is given by

$$\text{Var}(\hat{\theta}) \approx \theta(1-e^{-\theta})^2(1-e^{-\theta}-\theta e^{-\theta})^{-1}n^{-1}. \tag{4.73}$$

David and Johnson (1952) suggested an estimator based on the sample moments

$$T_r = \frac{1}{n}\sum_{j=1}^{n} x_j^r, \qquad r = 1, 2, \ldots.$$

Their estimator is

$$\theta^{\dagger} = T_2 T_1^{-1} - 1, \tag{4.74}$$

and its variance is approximately

$$(\theta + 2)(1 - e^{-\theta})n^{-1}.$$

Its asymptotic efficiency (relative to the MLE) reaches a minimum of about 70% between $\theta = 2.5$ and $\theta = 3$. The efficiency increases to 100% as θ increases (David and Johnson, 1952).

Tate and Goen (1958) have shown that the MVUE of θ is

$$\theta^* = \frac{n\overline{x}S(n\overline{x} - 1, n)}{S(n\overline{x}, n)} = \frac{\overline{x}[1 - S(n\overline{x} - 1, n - 1)]}{S(n\overline{x}, n)}, \tag{4.75}$$

where $S(a, b)$ is the Stirling number of the second kind defined in Section 1.1.3. There are tables of the multiplier of $\overline{x}$ in Tate and Goen (1958). For large n the multiplier is approximately $1 - (1 - n^{-1})^{n\overline{x}-1}$.

Let n_0 denote the number of missing values with $X = 0$. David and Johnson (1952, p. 277) posed the following question: "Are we to regard the truncated sample as one of a sequence of samples of equal size (i.e., n constant), or as the result of truncation of one of a sequence of complete samples of the same size (i.e., $n + n_0$ constant)?" The previous discussion has been based on the fixed-n case. McKendrick (1926) estimated θ on the alternative assumption of fixed $N = n + n_0$. For Poisson variables the first two factorial moments are θ and θ^2. McKendrick's method uses the first two observed factorial moments to obtain an initial estimate of n_0. The estimated mean and the expected zero frequency are then used to estimate θ, and hence n_0, iteratively. The method is of considerable historical interest and is not difficult to apply but is now rarely used. Assuming the same model (fixed $N = n + n_0$), Dahiya and Gross (1973) obtained a conditional MLE of N and hence estimated n_0. Blumenthal, Dahiya, and Gross (1978) derived the unconditional MLE of θ and also a modified MLE which appeared better in certain respects.

If the first r_1 values $(0, 1, \ldots, (r_1 - 1))$ are omitted, then we have a *left-truncated Poisson distribution* with pmf

$$\Pr[X = x] = \frac{e^{-\theta}\theta^x}{x!}\left(1 - e^{-\theta}\sum_{j=0}^{r_1-1}\frac{\theta^j}{j!}\right)^{-1}, \qquad x = r_1, r_1 + 1, \ldots. \tag{4.76}$$

The MLE $\hat{\theta}$ of θ, given values of n independent rv's $X_1, X_2, \ldots, X_n$, each distributed as in (4.76), satisfies the equation

$$\overline{x} = \left(\hat{\theta} - e^{-\hat{\theta}} \sum_{j=1}^{r_1-1} \frac{\hat{\theta}^j}{(j-1)!}\right)\left(1 - e^{-\hat{\theta}} \sum_{j=0}^{r_1-1} \frac{\hat{\theta}^j}{j!}\right)^{-1}. \tag{4.77}$$

As an initial value for use in the iterative solution of (4.77), the simple estimator

$$\theta^* = n^{-1} \sum_j (x_j | x_j > r_1) \tag{4.78}$$

that was proposed by Moore (1954) might be used. For the case $r_1 = 1$ (the positive Poisson distribution) this becomes

$$\theta^* = n^{-1} \sum_j (x_j | x_j > 1),$$

with variance

$$\mathrm{Var}\,(\theta^*) = \theta[1 + \theta(e^{\theta} - 1)^{-1}]n^{-1}$$

(Plackett, 1953), and its efficiency is quite high (over 90%).

Rider (1953) constructed an estimator for θ for the left-truncated Poisson distribution based on the uncorrected sample moments. Minimum-variance unbiased estimators have been given by Tate and Goen (1958). Maximum-likelihood estimators based on censored samples from truncated Poisson distributions have been discussed by Murakami (1961).

Rao and Rubin (1964) gave a characterization for the left-truncated Poisson distribution analogous to their characterization for the complete Poisson distribution in Section 4.8.

4.10.2 Right Truncation and Double Truncation

Right truncation (omission of values exceeding a specified value, r_2) can occur if the counting mechanism is unable to deal with large numbers. Often the *existence* of these high values is known, even if their exact magnitude is not; in this case, if there are n' values greater than r_2, then the MLE $\hat{\theta}$ satisfies the equation

$$\sum_j (x_j | x_j \leq r_2) = \hat{\theta}\left\{N - n'\left[1 + \frac{\hat{\theta}^{r_2}}{r_2!}\left(\sum_{j=r_2+1}^{\infty} \frac{\hat{\theta}^j}{j!}\right)^{-1}\right]\right\} \tag{4.79}$$

(Tippett, 1932). When this information is not available, then it is appropriate to use the *right-truncated Poisson distribution* with pmf

$$\Pr[X = x] = \frac{\theta^x}{x!}\left(\sum_{j=0}^{r_2}\frac{\theta^j}{j!}\right)^{-1}, \qquad x = 0, 1, \ldots, r_2. \tag{4.80}$$

If $X_1, X_2, \ldots, X_n$ are n independent rv's each having this pmf, then the MLE $\hat{\theta}$ of θ satisfies the following equation:

$$\sum_{j=0}^{r_2}\frac{(\overline{x} - j)\hat{\theta}^j}{j!} = 0. \tag{4.81}$$

The solution of this equation for $\hat{\theta}$ is a function of $\overline{x}$ alone and not of n. Cohen (1961) has provided tables from which $\hat{\theta}$ can be obtained, given r_2 and $\overline{x}$. When the tables are insufficient, a solution may be obtained by interpolation. Equation (4.81) can be rewritten in the form

$$\overline{x} = \frac{\hat{\theta}\sum_{j=0}^{r_2-1}(e^{-\hat{\theta}}\hat{\theta}^j/j!)}{\sum_{j=0}^{r_2}(e^{-\hat{\theta}}\hat{\theta}^j/j!)} \tag{4.82}$$

and the right-hand side evaluated either numerically or by the use of tables of the cumulative probabilities; see Section 4.6.1.

The asymptotic variance of $\hat{\theta}$ is $\theta\psi(\theta)n^{-1}$, where

$$\psi(\theta) = \frac{\left[\sum_{j=0}^{r_2}(\theta^j/j!)\right]^2}{\left[\sum_{j=0}^{r_2-1}(\theta^j/j!)\right]\left[\sum_{j=0}^{r_2-1}(\theta^j/j!) + \theta^{r_2+1}/r_2!\right] - (\theta^{r_2+1}/r_2!)\sum_{j=0}^{r_2}(\theta^j/j!)}. \tag{4.83}$$

Moore (1954) suggested the simple estimator [analogous to (4.78)]

$$\theta^* = \sum_j \frac{x_j}{m}, \tag{4.84}$$

where m is the number of values of x that are less than $r_2 - 1$; this is an unbiased estimator of θ.

The right-truncated Poisson arises in telephony when there are just n lines and any calls that arrive when all n lines are busy are not held. This gives the loss M/M/n queueing system with

$$\Pr[X = x] = \Pr[X = 0]\frac{\theta^x}{x!}, \qquad 0 \le x \le n, \tag{4.85}$$

where $\Pr[X = n]$ is the proportion of time for which the system is fully occupied, that is, the proportion of lost calls. The expression for $\Pr[X = n]$ is known as *Erlang's loss formula*.

The *doubly truncated Poisson distribution* (Cohen, 1954) is the distribution of a rv for which

$$\Pr[X = x] = \frac{\theta^x}{x!}\left(\sum_{j=r_1}^{r_2} \frac{\theta^j}{j!}\right)^{-1}, \qquad x = r_1, r_1 + 1, \ldots, r_2, \qquad 0 < r_1 < r_2. \tag{4.86}$$

If $X_1, X_2, \ldots, X_n$ are independent rv's, each with pmf (4.86), the MLE $\hat{\theta}$ of θ satisfies the equation

$$\overline{x} = \frac{\sum_{j=r_1}^{r_2} (j\hat{\theta}^j/j!)}{\sum_{j=r_1}^{r_2} (\hat{\theta}^j/j!)}. \tag{4.87}$$

This can be solved in the same manner as (4.81). Moore (1954) suggested a statistic analogous to (4.84) that is an unbiased estimator of θ.

Doss (1963) compared the efficiency of the MLEs of θ for the untruncated and for the doubly truncated Poisson distributions. He showed that the variance of $\hat{\theta}$ for untruncated data is always less than the variance of $\hat{\theta}$ for doubly truncated data. Comparisons of this kind are relevant when it is possible to control the method of observation to produce distributions either of type (4.1) or of type (4.86).

When all values of x that are greater than a certain value K are recorded as that value, the resultant distribution is

$$\begin{aligned} \Pr[X = x] &= \frac{e^{-\theta}\theta^x}{x!}, \qquad x < K, \\ \Pr[X = K] &= \sum_{j=K}^{\infty} \frac{e^{-\theta}\theta^j}{j!}. \end{aligned} \tag{4.88}$$

Newell (1965) applied this distribution to the number of hospital beds occupied when K is the total number of beds available.

4.10.3 Misrecorded Poisson Distributions

These distributions attempt to take into account errors in recording a variable which in reality does have a Poisson distribution. Suppose that the zero class alone is misrecorded. Then the pmf is

$$\begin{aligned} \Pr[X = 0] &= \omega + (1 - \omega)e^{-\theta}, \\ \Pr[X = x] &= (1 - \omega)\frac{e^{-\theta}\theta^x}{x!}, \qquad x \geq 1. \end{aligned} \tag{4.89}$$

When $0 < \omega < 1$ (that is, when there is overreporting), the distribution is known as the *Poisson-with-zeroes*, or alternatively as the *zero-inflated Poisson distribution*. For the *zero-deflated Poisson distribution* (that is, when there is underreporting), a necessary condition on ω is

$$(1 - e^{\theta})^{-1} < \omega < 0.$$

These are zero-modified Poisson distributions; see Section 8.2.3.

The distribution defined by the equations

$$\begin{aligned}
\Pr[X = 0] &= e^{-\theta}(1 + \theta\lambda), \\
\Pr[X = 1] &= \theta e^{-\theta}(1 - \lambda), \\
\Pr[X = x] &= \frac{\theta^x e^{-\theta}}{x!}, \qquad x \geq 2,
\end{aligned} \tag{4.90}$$

corresponds to a situation in which values from a Poisson distribution are recorded correctly, except that, when the true value is 1, there is a probability λ that it will be recorded as zero. Cohen (1960c) obtained the following formulas for MLEs of θ and λ, respectively:

$$\begin{aligned}
\hat{\theta} &= 0.5\{(\overline{x} - 1 + f_0) + [(\overline{x} - 1 + f_0)^2 + 4(\overline{x} - f_1)]^{1/2}\}, \\
\hat{\lambda} &= (f_0 - f_1\hat{\theta}^{-1})(f_0 + f_1)^{-1},
\end{aligned} \tag{4.91}$$

where $\overline{x}$ is the sample mean and f_0 and f_1 are the observed *relative* frequencies of zero and unity. The asymptotic variances of $\hat{\theta}$ and $\hat{\lambda}$ are, respectively,

$$\theta(1 + \theta)(1 + \theta - e^{-\theta})^{-1}n^{-1}$$

and

$$(1 + \lambda\theta - \lambda e^{-\theta})(1 - \lambda)(\theta e^{-\theta})^{-1}(1 + \theta - e^{-\theta})^{-1}n^{-1},$$

where n denotes the sample size. The asymptotic correlation between the estimators is

$$[(1 - \lambda)e^{-\theta}(1 - \theta)^{-1}(1 + \lambda\theta - \lambda e^{-\theta})^{-1}]^{1/2}.$$

Shah and Venkataraman (1962) gave formulas for the moments of the distribution; they also constructed estimators of θ and λ based on the sample moments.

Cohen (1959) considered the case where a true value of $c+1$ is sometimes reported as c (with probability α) where c is constant; the pmf is

$$\begin{aligned}
\Pr[X=x] &= \frac{e^{-\theta}\theta^x}{x!}, \qquad x=0,1,\ldots,c-1,c+2,c+3,\ldots,\\
\Pr[X=c] &= \frac{e^{-\theta}\theta^c}{c!}\left(1+\frac{\alpha\theta}{c+1}\right), \qquad (4.92)\\
\Pr[X=c+1] &= (1-\alpha)\frac{e^{-\theta}\theta^{c+1}}{(c+1)!}.
\end{aligned}$$

Cohen gave formulas for the MLEs $\hat{\theta}$ and $\hat{\alpha}$ and expressions for their asymptotic variances and covariance. In further papers Cohen (1960b,d) considered other similarly misrecorded Poisson distributions.

4.11 POISSON–STOPPED SUM DISTRIBUTIONS

Poisson–stopped sum distributions have pgf's of the form

$$G(z) = e^{\theta[g(z)-1]}; \qquad (4.93)$$

they arise as the distribution of the sum of a Poisson number of iid rv's with pgf $g(z)$. The term *Poisson–stopped sum* was introduced by Godambe and Patil (1975) and used by Douglas (1980) in his book on contagious distributions.

Because of their infinite divisibility (Section 9.3), these distributions have very great importance in discrete distribution theory. They are known by a number of other names. Feller (1943) used the term *generalized Poisson*; this usage is common, though subject to unfortunate ambiguity. Galliher et al. (1959) and Kemp (1967a) called them *stuttering Poisson*; Thyrion (1960) called them *Poisson par grappes*. The name *compound Poisson* was used by Feller (1950, 1957, 1968) and Lloyd (1980). The term "compound" is particularly confusing because it is also widely used for mixed Poisson distributions, as in the first edition of this book.

While certain very common distributions, such as the negative binomial (see Chapter 5), are both Poisson–stopped sum distributions *and* mixed Poisson distributions, there are others that belong to only one of these two families of distributions.

The term *clustered Poisson* distribution is also used. Suppose that the number of egg masses per plant laid by an insect has a Poisson distribution, that the number of larvae hatching per egg mass has pgf $g(z)$, and that egg masses hatch independently. Then the total number of larvae per plant has a Poisson–stopped sum distribution. These distributions have also been called *composed Poisson* distributions.

Let

$$g(z) = a_0 + a_1 z + a_2 z^2 + \cdots, \qquad a_i \ge 0, \qquad i=0,1,2,\ldots.$$

Then the Poisson–stopped sum distribution has pgf

$$G(z) = e^{\lambda a_1(z-1)} e^{\lambda a_2(z^2-1)} e^{\lambda a_3(z^3-1)} \cdots; \tag{4.94}$$

thus the distribution can also be interpreted as the convolution of a Poisson singlet distribution with a Poisson doublet distribution, a Poisson triplet distribution, and so on; in other words, it is the distribution of $X_1 + X_2 + X_3 + \cdots$, where X_1 has a Poisson distribution and $X_r, r = 2, 3, \ldots,$ has a Poisson distribution on $x = 0, r, 2r, \ldots$. A European term for "convolution" is "composition."

Clearly, if $g(z)$ is a pgf, then $a_i \geq 0$ for all i and $G(z)$ is infinitely divisible. Conversely, if $G(z)$ is infinitely divisible, then $g(z)$ is a valid pgf [see, e.g., Feller, 1968, p. 290].

If $g(z)$ is the pgf of a rv with finite support, then the Poisson–stopped sum of such rv's has been called an *extended Poisson distribution of order k* (Hirano and Aki, 1987); when $g(z)$ is a discrete uniform distribution with support $1, 2, \ldots, k$, the outcome is the *Poisson distribution of order k* (Philippou, 1983); see Section 10.7.4.

The best-known Poisson–stopped sum distribution is the negative binomial; here $g(z)$ is the logarithmic distribution (and has infinite support). Chapter 5 is devoted to this very important distribution and to its special form, the geometric.

The *Lagrangian Poisson distribution* has received much attention; it is a Poisson–stopped sum of Borel–Tanner rv's (Section 7.2.6). In his book Consul (1989) called it *the* "generalized Poisson distribution" and gave many earlier references.

The general theory of Poisson–stopped sum distributions is given in Chapter 9, where many special forms such as the Hermite, Neyman type A, Pólya–Aeppli, and Lagrangian Poisson are discussed in detail.

4.12 OTHER RELATED DISTRIBUTIONS

4.12.1 Normal Distribution

The limiting distribution of the standardized Poisson variable $(X - \theta)\theta^{-1/2}$, where X has distribution (4.1), is a unit-normal distribution. That is,

$$\lim_{\theta \to \infty} \Pr[\alpha < (X - \theta)\theta^{-1/2} < \beta] = (2\pi)^{-1/2} \int_\alpha^\beta e^{-u^2/2}\, du. \tag{4.95}$$

4.12.2 Gamma Distribution

The following important formal relation between the Poisson and gamma (and hence χ^2) distributions has already been noted in Sections 4.2.1 and 4.5. If Y has pdf

$$f(y) = \frac{1}{(\alpha - 1)!} y^{\alpha-1} e^{-y}, \qquad 0 < y,$$

and α is a positive integer, then

$$\begin{aligned}
\Pr[Y > y] &= \frac{1}{(\alpha-1)!}\int_y^\infty u^{\alpha-1}e^{-u}\,du \\
&= \frac{1}{(\alpha-1)!}y^{\alpha-1}e^{-y} + \frac{1}{(\alpha-2)!}\int_y^\infty u^{\alpha-2}e^{-u}\,du \\
&\qquad\qquad\text{(integrating by parts)} \\
&= \frac{1}{(\alpha-1)!}y^{\alpha-1}e^{-y} + \frac{1}{(\alpha-2)!}y^{\alpha-2}e^{-y} + \cdots + ye^{-y} + e^{-y} \\
&= \Pr[X < \alpha], \qquad (4.96)
\end{aligned}$$

where X is a Poisson rv with parameter y. If $\alpha = \nu/2$, where ν is an even integer, then

$$\Pr\left[X < \tfrac{1}{2}\nu\right] = \Pr[\chi_\nu^2 > 2y].$$

4.12.3 Sums and Differences of Poisson Variates

If X_i, $i = 1, 2, \ldots$, are independent Poisson rv's with parameters θ_i, then $\sum_i X_i$ has a Poisson distribution with parameter $\sum_i \theta_i$; this is the reproductive property of the Poisson distribution.

The sum of n iid positive (i.e., zero-truncated) Poisson rv's with parameter θ was studied by Tate and Goen (1958); the distribution is known as the *Stirling distribution of the second kind*. The pgf is

$$G(z) = \frac{(e^{\theta z}-1)^n}{(e^\theta - 1)^n} \qquad (4.97)$$

and the pmf is

$$\Pr[Y = y] = \frac{n!S(y,n)\theta^y}{(e^\theta-1)^n y!}, \qquad y = n, n+1, \ldots, \qquad (4.98)$$

where $S(y, n)$ is the Stirling number of the second kind. The mean and variance are

$$\mu = n\theta(1-e^{-\theta})^{-1} \qquad \text{and} \qquad \mu_2 = \mu\left(1+\theta-\frac{\mu}{n}\right),$$

respectively. The distribution is a GPSD, and therefore the MLE of θ is given by

$$\overline{x} = n\hat{\theta}(1-e^{-\hat{\theta}})^{-1}. \qquad (4.99)$$

Tate and Goen considered minimum-variance unbiased estimation of θ; see also Ahuja (1971a). Ahuja and Enneking (1972a) obtained a recurrence relationship for the pmf as well as an explicit expression for the cdf as a sum of a linear combination of incomplete gamma functions. Tate and Goen (1958) and Ahuja (1971a) also considered sums of iid Poisson rv's left truncated at $c > 1$. Cacoullos (1975) has developed *multiparameter Stirling distributions of the second kind* in his study of the sum of iid Poisson rv's left truncated at differing, unknown points.

Gart (1974) studied the exact moments of the statistic for the variance homogeneity test for samples from the zero-truncated Poisson distribution and gave a number of references to earlier work on the distribution.

The distribution of the difference between two independent rv's each having a Poisson distribution has attracted some attention. Strackee and van der Gon (1962, p. 17) state, "In a steady state the number of light quanta, emitted or absorbed in a definite time, is distributed according to a Poisson distribution. In view thereof, the physical limit of perceptible contrast in vision can be studied in terms of the difference between two independent variates each following a Poisson distribution." The distribution of differences may also be relevant when a physical effect is estimated as the difference between two counts, one when a "cause" is acting and the other a "control" to estimate the "background effect."

Irwin (1937) studied the case when the two variables X_1 and X_2 each have the same expected value θ. Evidently, for $y \geq 0$,

$$\begin{aligned}\Pr[X_1 - X_2 = y] &= e^{-2\theta} \sum_{j=y}^{\infty} \theta^{j+(j-y)} [j!(j-y)!]^{-1} \\ &= e^{-2\theta} I_y(2\theta), \qquad (4.100)\end{aligned}$$

where $I(\cdot)$ is a modified Bessel function of the first kind; see Section 1.1.5. (By symmetry $\Pr[X_1 - X_2 = y] = \Pr[X_2 - X_1 = y]$.) In this particular case all the cumulants of odd order are zero, and all those of even order are equal to 2θ. The pgf is $\exp[\theta(z + z^{-1} - 2)]$ and the approach to normality is rapid.

Skellam (1946), de Castro (1952), and Prekopa (1952) discussed the problem when $E[X_1] = \theta_1 \neq E[X_2] = \theta_2$. In this case, for $y > 0$,

$$\begin{aligned}\Pr[X_1 - X_2 = y] &= e^{-(\theta_1+\theta_2)} \sum_{j=y}^{\infty} \theta_1^j \theta_2^{j-y} [j!(j-y)!]^{-1} \\ &= e^{-(\theta_1+\theta_2)} \left(\frac{\theta_1}{\theta_2}\right)^{y/2} I_y(2\sqrt{\theta_1\theta_2}) \qquad (4.101)\end{aligned}$$

(note the relationship to the noncentral χ^2 distribution; Johnson et al. (1995, Chapter 28)). The tail probabilities can be formulated in terms of the tail probabilities of the noncentral χ^2 distribution, providing a generalization of the relationship between the Poisson and χ^2 distributions (Johnson, 1959).

Table 4.1 Combinations of Values of θ_1 and θ_2 in Strackee and van der Gon's Tables of $\Pr[X_1 - X_2 = y]$

θ_1	$\frac{1}{4}$	1	4	$\frac{1}{2}$	1	2	4	8	1	2	4	2	4	8	4	8
θ_2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1	2	2	2	4	8

Strackee and van der Gon (1962) gave tables of the cumulative probability $\Pr[X_1 - X_2 = y]$ to four decimal places for the combinations of values of θ_1 and θ_2 given in Table 4.1.

Their tables also show the differences between the normal approximations [see Fisz, 1953]

$$\Pr[X_1 - X_2 \leq y] \approx [2\pi(\theta_1 + \theta_2)]^{-1/2} \int_{-\infty}^{y+1/2} \exp\left[-\frac{(\theta_1 - \theta_2 - y)^2}{2(\theta_1 + \theta_2)}\right] dy \tag{4.102}$$

and the tabulated values. Romani (1956) showed that all the odd cumulants of $X_1 - X_2$ are equal to $\theta_1 - \theta_2$, and that all the even cumulants are equal to $\theta_1 + \theta_2$. He also discussed the properties of the MLE of $E[X_1 - X_2] = \theta_1 - \theta_2$. Katti (1960) studied $E[|X_1 - X_2|]$.

Ratcliffe (1964) carried out a Monte Carlo experiment to assess the accuracy with which the distribution of $(X_1 - X_2)(X_1 + X_2)^{-1/2}$ is represented by a unit-normal distribution for the special case $\theta_1 = \theta_2 = \theta$.

Consider now $X = Y + Z$, where Y is a Poisson rv with parameter θ and $Z \geq 0$ is a discrete rv independent of Y. Samaniego (1976) discussed the applicability of such a model to various situations, such as signal + noise processes and misclassification. He examined maximum-likelihood estimation of θ when the distribution of Z is known, and he related this to Crow and Gardner's (1959) concept of confidence belts (see Section 4.7.3). Huang and Fung's (1989) *intervened truncated Poisson distribution* arises when Z has a truncated Poisson distribution (left, right, or double truncation); see Section 4.12.7.

4.12.4 Hyper-Poisson Distributions

Staff (1964, 1967) defined the *displaced Poisson distribution* as a left-truncated Poisson distribution that has been "displaced" by subtraction of a constant so that the lowest value that is taken by the variable is zero. Thus the left-truncated distribution

$$\Pr[Y = y] = \frac{\theta^y}{y!}\left(\sum_{j=r+1}^{\infty} \frac{\theta^j}{j!}\right)^{-1}, \qquad y = r+1, r+2, \ldots,$$

corresponds to the displaced distribution

$$\Pr[X = x] = \frac{\theta^{x+r+1}}{(x+r+1)!}\left(\sum_{j=r+1}^{\infty} \frac{\theta^j}{j!}\right)^{-1}, \qquad x = 0, 1, 2, \ldots. \tag{4.103}$$

Staff also obtained the distribution via the recurrence relation

$$\Pr[X = x + 1] = \frac{\theta}{r + x} \Pr[X = x], \qquad x = 1, 2, \ldots .$$

If r is known, estimation of θ is straightforward. Staff also considered the situation in which both r and θ have to be estimated. The simplest formulas that he obtained were for estimators based on the sample mean $\overline{x}$, standard deviation s, and proportion f_0 in the zero class. They are

$$\begin{aligned} r^* &= (s^2 - \overline{x})[1 - f_0(1 + \overline{x})]^{-1}, \\ \lambda^* &= \overline{x} + r^*(1 - f_0) \end{aligned} \tag{4.104}$$

(for the displaced Poisson distribution r is necessarily a positive integer, though the restriction to integers is not necessary for the hyper-Poisson distribution; see below). If $r < 8$, the efficiency of r^* relative to the MLE exceeds 70%; if $r = 2$, then the efficiency of r^* exceeds 89%.

By stating its pgf in the form

$$G(z) = \frac{{}_1F_1[1; r; \theta z]}{{}_1F_1[1; r; \theta]}, \quad r \text{ an integer}, \tag{4.105}$$

we see immediately that the displaced Poisson is a special case of the more general *hyper-Poisson distribution* with pgf

$$G(z) = \frac{{}_1F_1[1; \lambda; \theta z]}{{}_1F_1[1; \lambda; \theta]}, \tag{4.106}$$

λ nonnegative real. Bardwell and Crow (1964) termed the distribution *sub-Poisson* for $\lambda < 1$ and *super-Poisson* for $\lambda > 1$. It is easy to obtain the recurrence relation

$$(\lambda + x) \Pr[X = x + 1] = \theta \Pr[X = x], \qquad x = 1, 2, \ldots . \tag{4.107}$$

Summation of both sides of the equation gives

$$\begin{aligned} \mu &= (\theta + 1 - \lambda) - (1 - \lambda) \Pr[X = 0] \\ &= (\theta + 1 - \lambda) - (1 - \lambda)({}_1F_1[1; \lambda; \theta])^{-1}. \end{aligned} \tag{4.108}$$

The distribution is a PSD (Section 2.2.1). It belongs to the generalized hypergeometric family (Section 2.4.1); this enabled Kemp (1968a,b) to obtain recurrence relationships of the form

$$\mu'_{[i+2]} = (\theta - \lambda - i)\mu'_{[i+1]} + \theta(i + 1)\mu'_{[i]}. \tag{4.109}$$

It follows that all moments of the distribution are finite and that the variance is

$$\operatorname{Var}(X) = (\theta - \lambda + 1 - \mu)\mu + \theta. \tag{4.110}$$

Barton (1966) commented that the distribution can be regarded as a mixed Poisson distribution, where the Poisson parameter has a truncated Pearson type III distribution (Section 8.3.3). Kemp (1968a) showed, more generally, that mixing a hyper-Poisson distribution with parameters λ and θ can yield a hyper-Poisson with parameters $\lambda + \eta$ and θ, where $\eta > 0$.

Bardwell and Crow (1964) described various methods of estimating λ and θ given values $x_1, x_2, \ldots, x_n$ of n independent rv's each having the distribution (4.106).

First, suppose that λ is known. Since (4.106) can be regarded as a PSD (Section 2.2.1), the MLE $\hat{\theta}$ is obtained by equating the expected value (4.108) to the sample arithmetic mean $\overline{x}$, giving

$$\overline{x} = \hat{\theta} + (1 - \lambda)\{1 - ({}_1F_1[1; \lambda; \hat{\theta}])^{-1}\}, \tag{4.111}$$

that is,

$${}_1F_1[1; \lambda; \hat{\theta}] = \frac{1 - \lambda}{1 - \lambda - \overline{x} + \hat{\theta}}. \tag{4.112}$$

The variance of $\hat{\theta}$ is approximately $\theta^2[n \operatorname{Var}(X)]^{-1}$, where $\operatorname{Var}(X)$ is as given in (4.110).

The UMVUE of θ is *approximately* $\lambda\overline{x}$ when n is large. Crow and Bardwell (1965) obtained the values

$$\begin{array}{ll} 0 & \text{when } \overline{x} = 0 \\ \lambda n^{-1} & \text{when } \overline{x} = n^{-1} \\ 2\lambda n^{-1}[1 + n^{-1}(\lambda - 1)(\lambda + 1)^{-1}]^{-1} & \text{when } \overline{x} = 2n^{-1} \end{array}$$

Second, when both λ and θ are unknown, the equations satisfied by the MLEs $\hat{\lambda}$ and $\hat{\theta}$ are

$${}_1F_1[1; \hat{\lambda}; \hat{\theta}] = \frac{1 - \hat{\lambda}}{1 - \hat{\lambda} - \overline{x} + \hat{\theta}} \tag{4.113}$$

[cf. (4.112)] and

$$\sum_{j=1}^{\infty} \hat{\theta}^j (\hat{\lambda}^{[j]})^{-1} S_j = {}_1F_1[1; \hat{\lambda}; \hat{\theta}] n^{-1} \sum_{i=1}^{n} S_{x_i}, \tag{4.114}$$

where $S_j = \sum_{i=1}^{j} (\lambda + i - 1)^{-1}$.

Crow and Bardwell (1965) also described some other estimators. All of them seem to have rather low efficiencies, so we will only mention a pair, θ^* and λ^*, that use the proportion of zeroes (f_0) and the first two sample moments $\overline{x}$ and $m_2' = n^{-1}\sum_{j=1}^{n} x_j^2$. They are

$$\begin{aligned}\theta^* &= [(1-f_0)m_2' - \overline{x}^2][1 - f_0(\overline{x}+1)]^{-1},\\ \lambda^* &= 1 + \frac{(m_2' - \overline{x}^2) - \overline{x}}{1 - f_0(\overline{x}+1)}.\end{aligned} \tag{4.115}$$

Some queueing theory, with a hyper-Poisson distribution of arrivals, has been worked out by Nisida (1962).

The hyper-Poisson is a special case of the Hall (1956) and Bhattacharya (1966) *confluent hypergeometric distribution* with pgf

$$G(z) = \frac{{}_1F_1[a;\lambda;\theta z]}{{}_1F_1[a;\lambda;\theta]}. \tag{4.116}$$

Also it is a member of the extended Katz family of distributions; see Section 2.3.1. Inference for this family has been studied by Gurland and Tripathi (1975) and Tripathi and Gurland (1977, 1979).

4.12.5 Grouped Poisson Distributions

In Section 4.2.1 it was shown that, if the times between successive events have a common exponential distribution, then the total number of events in a fixed time T has a Poisson distribution. Morlat (1952) constructed a "generalization of the Poisson law" by supposing that the common distribution of the times (t) is a gamma distribution with origin at zero (Johnson et al. 1994, Chapter 17). If the common pdf is

$$[\Gamma(\alpha)]^{-1} t^{\alpha-1} e^{-t}, \qquad t \geq 0,$$

then the probability of x events in time T is

$$\Pr[X = x] = [\Gamma(x\alpha)]^{-1} \int_0^T \left(1 - \frac{y^\alpha \Gamma(x\alpha)}{\Gamma((x+1)\alpha)}\right) y^{x\alpha-1} e^{-y}\,dy. \tag{4.117}$$

If α is an integer, we have the remarkable formula

$$\Pr[X = x] = \sum_{j=x\alpha}^{(x+1)\alpha-1} \frac{e^{-T}T^j}{j!}, \tag{4.118}$$

showing that the probabilities are given by grouping together an α number of successive Poisson probabilities.

The "generalized Poisson" distributions of Gold (1957) and Gerstenkorn (1962) [see also Godwin (1967)] have pmf's formed from cumulative Poisson probabilities. We have

$$\Pr[X = x] = q^x e^{-qu} \left(\sum_{j=x}^{\infty} \frac{u^j}{j!} \right) c^{-1}, \qquad x = 0, 1, 2, \ldots, \tag{4.119}$$

where $u > 0$, c is an appropriate normalizing constant, and either (1) $q > 0$, $q \neq 1$ or (2) $q = 1$.

In case 1 the pgf is

$$G(z) = \left(\frac{1 - q}{1 - qz} \right) \left(\frac{qze^{u(qz-1)} - 1}{qe^{u(q-1)} - 1} \right); \tag{4.120}$$

in case 2, the pgf is

$$G(z) = \frac{ze^{u(z-1)} - 1}{(1 + u)(z - 1)} \tag{4.121}$$

and now the probabilities are the normalized tail probabilities of the Poisson distribution.

These "generalized Poisson" distributions are related but not identical to the *burnt-fingers distribution* of Greenwood and Yule (1920), McKendrick (1926), Arbous and Kerrich (1951), and Irwin (1953). The burnt-fingers distribution arises if the first event occurs in time $(t, t + \partial t)$ with probability $a\,\partial t + o(\partial t)$, while subsequent events occur with probability $b\,\partial t + o(\partial t)$. The probabilities are

$$\begin{aligned} \Pr[X(t) = 0] &= e^{-at}, \\ \Pr[X(t) = x] &= e^{-bt} \frac{a}{b} \left(\frac{b}{b - a} \right)^x \sum_{j=x}^{\infty} \frac{(b - a)^j t^j}{j!}, \qquad x = 1, 2, \ldots, \end{aligned} \tag{4.122}$$

and the pgf is

$$g(z) = \frac{aze^{bt(z-1)} + e^{-at}(b - a)(z - 1)}{a - b + bz}. \tag{4.123}$$

The mean and variance are

$$\mu = bt + \left(1 - \frac{b}{a} \right)(1 - e^{-at}), \qquad \mu_2 = b^2 t^2 + 2bt - \frac{2b\mu}{a} + \mu - \mu^2.$$

The distribution can be underdispersed as well as overdispersed; see Faddy's (1994) use of the distribution in a generalized linear model.

Faddy and Bosch (2001) showed that the Poisson process can be further generalized by modeling the birth rates as

$$\lambda_n = a\left(n^b\, e^{-cn} + d\right), \qquad n \geq 1,$$

where $\lambda_n \to ad$ as $n \to \infty$; see also Faddy (1997). They demonstrated the use of the distribution for underdispersed data using numbers of fetal implants in mice subjected to different doses of a herbicide.

The simple immigration-catastrophe process of Swift (2000) also has probabilities involving grouped Poisson probabilities. Here the transitions and the rates are as follows:

Transition	**Rate**
$i \to i+1,\ i \geq 0$	α
$i \to 0,\ i \geq 1$	γ

Let $N(t)$ be the population size at time t and set $p_n(t) = \Pr[N(t) = n]$. Swift shows that

$$\begin{aligned} p_0(t) &= \frac{\gamma}{\alpha+\gamma} + \frac{\alpha}{\alpha+\gamma} e^{-(\alpha+\gamma)t}, \\ p_n(t) &= \frac{\alpha^n \gamma}{(\alpha+\gamma)^{n+1}} + \frac{\alpha^{n+1} t^n}{(\alpha+\gamma) n!} e^{-(\alpha+\gamma)t} - \frac{\alpha^n \gamma\ \Gamma(n, (\alpha+\gamma)t)}{(\alpha+\gamma)^{n+1}(n-1)!} \\ &= \frac{\alpha^n \gamma}{(\alpha+\gamma)^{n+1}} + e^{-(\alpha+\gamma)t}\left[\frac{\alpha^{n+1} t^n}{(\alpha+\gamma) n!} - \frac{\alpha^n \gamma}{(\alpha+\gamma)^{n+1}} \sum_{j=0}^{n-1} \frac{(\alpha+\gamma)^j t^j}{j!}\right] \end{aligned} \tag{4.124}$$

for $n = 1, 2, \ldots$. The pgf is

$$G(z) = \frac{\alpha(z-1)e^{\alpha t z - \alpha t - \gamma t} - \gamma}{\alpha(z-1) - \gamma} \tag{4.125}$$

and the mean and variance are

$$\mu = \frac{\alpha}{\gamma}(1 - e^{-\gamma t}), \qquad \mu_2 = \frac{\alpha}{\gamma^2}(\alpha + \gamma - \gamma e^{-\alpha} - 2\alpha\gamma e^{-\gamma} - \alpha e^{-2\gamma}).$$

Janardan et al. (1997) have modeled the oviposition tactics of weevils on beans via a stochastic process which involves random changes of transition probabilities after either one egg or two eggs are laid on the same bean. If $N(t)$ is the number of eggs on a bean at time t (and $p_n(t) = \Pr[N(t) = n]$), then

$$\begin{aligned} p_0(t) &= e^{-\lambda t}, \\ p_1(t) &= \frac{\lambda}{p\beta}(e^{-\lambda t} - e^{-\alpha t}), \\ p_n(t) &= \frac{\alpha\lambda\mu^{n-2}e^{-\mu t}}{pq^{n-1}\beta^n}\{q^{n-1}[e^{\beta t} - u(t)] - [e^{q\beta t} - v(t)]\}, \ n = 2, 3, \ldots, \end{aligned} \tag{4.126}$$

where

$$u(t) = \sum_{j=0}^{n-2} \frac{(\beta t)^j}{j!} \qquad \text{and} \qquad v(t) = \sum_{j=0}^{n-2} \frac{(q\beta t)^j}{j!}.$$

There are three parameters, p, λ, and μ; here $q = 1 - p$ and $\alpha = \mu p + \lambda q$. The pgf is

$$\begin{aligned} G(z) = e^{\mu t(z-1)} &+ \frac{(z-1)(\beta p + z\lambda q)}{p(\mu z - \beta)}(e^{-\lambda t} - e^{\mu t(z-1)}) \\ &- \frac{z(z-1)\lambda q}{p(\mu z - \beta q)}(e^{-\alpha t} - e^{\mu t(z-1)}). \end{aligned} \tag{4.127}$$

The mean is

$$E[N(t)] = \mu t - \beta\left[\frac{1 - e^{-\lambda t}}{\lambda} - \frac{\lambda q}{\beta p}\left(\frac{1 - e^{-\alpha t}}{\alpha} - \frac{1 - e^{-\lambda t}}{1 - \lambda}\right)\right];$$

the formula for the variance is allegedly "very messy."

4.12.6 Heine and Euler Distributions

The Heine and Euler distributions were first studied by Benkherouf and Bather (1988) as prior distributions for stopping-time strategies when sequentially drilling for oil. Kemp (1992a,b) showed that they can both be considered to be q-series analogs of the Poisson distribution; a number of other q-series distributions have been investigated during the last decade. These q-series distributions are dealt with in Section 10.8.

4.12.7 Intervened Poisson Distributions

Consider the number Y of cholera cases per household where the event $Y = 0$ is unobservable; assume that the distribution of Y is a zero-truncated Poisson with parameter (θ). Suppose now that new preventive measures change θ to $\theta\rho$, $0 \leq \rho < \infty$. Let Z be the total number of cholera cases that occur after the preventive measures are applied and assume that Z is a Poisson $(\rho\theta)$ rv. Then the total number of cholera cases is $X = Y + Z$ and

$$\begin{aligned}\Pr[X = x] &= \sum_{i=0}^{x-1} \Pr[Y = x - i] \Pr[Z = i | Y = x - i] \\ &= \frac{e^{-\rho\theta}}{e^{\theta} - 1}\left[(1+\rho)^x - \rho^x \right]\frac{\theta^x}{x!} \end{aligned} \tag{4.128}$$

This is the *intervened Poisson distribution* (IPD) of Shanmugam (1985). The pgf is

$$G_X(z) = \frac{e^{\rho\theta z}(e^{\theta}z - 1)}{e^{\rho\theta}(e^{\theta} - 1)}; \tag{4.129}$$

θ is called the incidence parameter and ρ is the intervention parameter. The mean and variance are

$$\mu = \theta\left(\rho + 1 + \frac{1}{e^{\theta} - 1}\right) \qquad \text{and} \qquad \mu_2 = \mu - \frac{\theta^2 e^{\theta}}{(e^{\theta} - 1)^2}$$

(the distribution is underdispersed).

The sum of n such iid rv's has the pgf $[G_X(z)]^n$; Shanmugam calls it an *intervened Stirling distribution of the second kind* (ISDSK) by analogy with the Stirling distribution of the second kind, which is the distribution of the sum of n zero-truncated Poisson rv's. He gives expressions for the pmf and the cdf in terms of generalized Stirling numbers of the second kind, a recurrence relationship for the probabilities, and an approximation for the cdf for large values of the rv. Moment estimation, with

$$\overline{x} - s_x^2 = \frac{\tilde{\theta}^2 e^{\tilde{\theta}}}{(e^{\tilde{\theta}} - 1)^2} \qquad \text{and} \qquad \tilde{\rho} = \frac{\overline{x}}{n\tilde{\theta}} - \frac{1}{(1 - e^{-\tilde{\theta}})^2},$$

is used to fit data on cholera cases in an Indian village. The objective in fitting the distribution is to estimate ρ; the closer ρ is to zero, the more effective are the intervention measures. Scollnik (1995) carried out a Bayesian analysis of these data using the Gibbs sampler and adaptive rejection sampling. Streit (1987) and Shanmugam (1992) have both studied tests for $H_0 : \rho = 1$ against $H_1 : \rho < 1$, that is, for no effect against some positive effect.

Huang and Fung's (1989) *intervened truncated Poisson distribution* arises when the original Poisson distribution is left, right, or doubly truncated (not necessarily only the zero frequency). The authors give moment and maximum-likelihood estimation equations.

Dhanavanthan (1998) considers a sequence $\{X_k; k \geq 1\}$ of N iid Bernoulli rv's with parameter λ, $0 < \lambda \leq 1$, where N is an independent IPD; he calls the distribution of $U = \sum_{i=1}^{N} X_i$ a *compound intervened Poisson distribution* (CIPD). The pmf is

$$\Pr[U = u] = \frac{e^{\theta(1-\lambda-\lambda\rho)}}{e^{\theta} - 1}\left[(1+\rho)^u - \rho^u e^{-\theta(1-\lambda)}\right]\frac{(\lambda\theta)^u}{u!} \tag{4.130}$$

and the pgf is

$$G_U(z) = \frac{e^{\lambda\rho\theta(z-1)}(e^{(z-1)\lambda\theta+\theta} - 1)}{e^{\theta} - 1}. \tag{4.131}$$

The mean and variance are

$$\mu = \lambda\theta\left(\frac{e^{\theta}}{e^{\theta} - 1} + \rho\right) \qquad \text{and} \qquad \mu_2 = \mu - \frac{e^{\theta}\theta^2\lambda^2}{(e^{\theta} - 1)^2},$$

and again there is underdispersion.

Dhanavanthan also investigated the distribution of $\sum_{i=1}^{n} U_i$, $n = 1, 2, \ldots,$ fixed, which he terms a *modified* ISDKS. He obtained an expression for the pmf and a recurrence for the probabilities. Here, briefly, and in Dhanavanthan (2000) he has examined moment estimates for the parameters and has explored their properties. He notes that maximum-likelihood estimation does not have a solution in closed form.

CHAPTER 5

Negative Binomial Distribution

5.1 DEFINITION

Many different models give rise to the negative binomial distribution, and consequently there is a variety of definitions in the literature. The two main dichotomies are (a) between parameterizations and (b) between points of support.

Formally, the negative binomial distribution can be defined in terms of the expansion of the negative binomial expression $(Q - P)^{-k}$, where $Q = 1 + P$, $P > 0$, and k is positive real; the $(x + 1)$th term in the expansion yields $\Pr[X = x]$. This is analogous to the definition of the binomial distribution in terms of the binomial expression $(\pi + \omega)^n$, where $\omega = 1 - \pi$, $0 < \pi < 1$, and n is a positive integer.

Thus the *negative binomial distribution* with parameters k, P is the distribution of the rv X for which

$$\Pr[X = x] = \binom{k + x - 1}{k - 1} \left(\frac{P}{Q}\right)^x \left(1 - \frac{P}{Q}\right)^k, \qquad x = 0, 1, 2, \ldots, \tag{5.1}$$

where $Q = 1 + P$, $P > 0$, and $k > 0$. Unlike the binomial distribution, here there is a nonzero probability for X to take any specified nonnegative integer value, as in the case of the Poisson distribution.

The probability generating function (pgf) is

$$G(z) = (1 + P - Pz)^{-k} \tag{5.2}$$

$$= {}_1F_0[k;\ ;P(z - 1)] \tag{5.3}$$

$$= \frac{{}_1F_0[k;\ ;Pz/(1 + P)]}{{}_1F_0[k;\ ;P/(1 + P)]} \tag{5.4}$$

and the characteristic function is $(1 + P - Pe^{it})^{-k}$. The mean and variance are

$$\mu = kP \qquad \text{and} \qquad \mu_2 = kP(1 + P). \tag{5.5}$$

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

This parameterization (but with the symbol p instead of P) is the one introduced by Fisher (1941).

Other early writers adopted different parameterizations. Jeffreys (1941) had $b = P/(1+P)$, $\rho = kP$, giving the pgf $[(1-bz)/(1-b)]^{\rho-\rho/b}$, and $\mu = \rho$, $\mu_2 = \rho/(1-b)$. Anscombe (1950) used the form $\alpha = k$, $\lambda = kP$, giving the pgf $(1+\lambda/\alpha - \lambda z/\alpha)^{-\alpha}$, and $\mu = \lambda$, $\mu_2 = \lambda(1+\lambda/\alpha)$.

Evans (1953) took $a = P$, $m = kP$, giving the pgf $(1+a-az)^{-m/a}$, and $\mu = m$, $\mu_2 = m(1+a)$. This parameterization has been popular in the ecological literature. Some writers, for instance Patil et al. (1984), called this the Pólya–Eggenberger distribution, as it arises as a limiting form of Eggenberger and Pólya's (1923) urn model distribution. Other authors, notably Johnson and Kotz (1977) and Berg (1988b), called the (nonlimiting) urn model distribution the Pólya–Eggenberger distribution; see Section 6.2.4. For both distributions "Pólya–Eggenberger" is quite often abbreviated to "Pólya."

A further parameterization that has gained wide favor is $p = 1/(1+P)$, that is, $q = P/(1+P)$, and $k = k$, giving

$$G(z) = \left(\frac{1-q}{1-qz}\right)^k, \tag{5.6}$$

$$\Pr[X = x] = \binom{k+x-1}{k-1} q^x (1-q)^k, \qquad x = 0, 1, 2, \ldots, \tag{5.7}$$

and $\mu = kq/(1-q)$, $\mu_2 = kq/(1-q)^2$. Sometimes $\lambda = P/(1+P)$ is used to avoid confusion with the binomial parameter q.

Clearly k need not be an integer. When k is an integer, the distribution is sometimes called the *Pascal distribution* (Pascal, 1679). The name "Pascal distribution" is, however, more often applied to the distribution shifted k units from the origin, that is, with support $k, k+1, \ldots$; this is also called the binomial waiting-time distribution.

The geometric is the special case $k = 1$ of the negative binomial distribution.

Kemp (1967a) summarized four commonly encountered formulations of pgf's for the negative binomial and geometric distributions as follows:

Formulation	**Negative Binomial**	**Geometric**	**Conditions**
1	$p^k(1-qz)^{-k}$	$p(1-qz)^{-1}$	$p+q=1$, $0<p<1$
2	$p^kz^k(1-qz)^{-k}$	$pz(1-qz)^{-1}$	
3	$(Q-Pz)^{-k}$	$(Q-Pz)^{-1}$	$Q = 1+P, P>0$ ($Q=1/p$, $P=q/p$, i.e., $p = 1/Q, q = P/Q$)
4	$z^k(Q-Pz)^{-k}$	$z(Q-Pz)^{-1}$	

In cases 1 and 3, k is positive real, the support is $0, 1, 2, \ldots$, and the distribution is a power series distribution (PSD). For cases 2 and 4, k is necessarily a positive integer; the distribution has support $k, k+1, k+2, \ldots$, and the distribution is a generalized power series distribution (GPSD) (see Section 2.1).

Case 1 shows that the distribution is a generalized hypergeometric probability distribution [with argument parameter q; cf. (5.4)], while case 3 shows that it is a generalized hypergeometric factorial moment distribution [with argument parameter P; cf. (5.3)]. Other families to which the negative binomial belongs are the exponential (provided k is fixed), Katz, Willmot, and Ord families.

5.2 GEOMETRIC DISTRIBUTION

In the special case $k = 1$ the pmf is

$$\Pr[X = j] = Q^{-1}\left(\frac{P}{Q}\right)^{j} = pq^{j}, \qquad j = 0, 1, 2, \ldots. \tag{5.8}$$

These values are in geometric progression, so this distribution is called a *geometric distribution*; sometimes it is called a *Furry distribution* (Furry, 1937).

Its properties can be obtained from those of the negative binomial as the special case $k = 1$.

The geometric distribution possesses a property similar to the "nonaging" (or "Markovian") property of the exponential distribution (Johnson et al. 1995, Chapter 18). This is

$$\Pr[X = x + j | X \geq j] = \frac{Q^{-1}(P/Q)^{x+j}}{(P/Q)^{j}} = Q^{-1}\left(\frac{P}{Q}\right)^{x} = \Pr[X = x]. \tag{5.9}$$

This property characterizes the geometric distribution (among all distributions restricted to the nonnegative integers), just as the corresponding property characterizes the exponential distribution. The distribution is commonly said to be a discrete analog of the exponential distribution. It is a special case of the grouped exponential distribution; see Spinelli (2001).

The geometric distribution may be extended to cover the case of a variable taking values $\theta_0, \theta_0 + \delta, \theta_0 + 2\delta, \ldots \quad (\delta > 0)$. Then, in place of (5.8), we have

$$\Pr[X = \theta_0 + j\delta] = Q^{-1}\left(\frac{P}{Q}\right)^{j}. \tag{5.10}$$

The characterization summarized in (5.9) also applies to this distribution with X replaced by $\theta_0 + X\delta$ and j replaced by $\theta_0 + j\delta$.

Other characterizations are described in Section 5.9.1.

Another special property of the geometric distribution is that, if a mixture of negative binomial distributions [as in (5.1)] is formed by supposing k to have the geometric distribution

$$\Pr[k = j] = (Q')^{-1}\left(\frac{P'}{Q'}\right)^{j-1}, \qquad j = 1, 2, \ldots, \tag{5.11}$$

then the resultant mixture distribution is also a geometric distribution of the form (5.8) with Q replaced by $QQ' - P'$.

The geometric, like the negative binomial distribution, is infinitely divisible; see Section 1.2.10 for a definition of infinite divisibility.

The Shannon entropy (first-order entropy) of the geometric distribution is $P \log_2 P - Q \log_2 Q$. The second-order entropy is $\log_2(1 + 2P)$.

Margolin and Winokur (1967) obtained formulas for the moments of the order statistics for the geometric distribution and tabulated values of the mean and variance to two decimal places; see also Kabe (1969). Steutel and Thiemann's (1989a) expressions for the order statistics were derived using the independence of the integer and fractional parts of exponentially distributed rv's. The computation of the order statistics from the geometric distribution has also been studied by Adatia (1991). Adatia also obtained an explicit formula for the expected value of the product of two such order statistics.

Order statistics from a continuous distribution form a Markov chain, but this is not in general true for discrete distributions. At first it was thought that exceptionally the Markov property holds for the geometric distribution (Gupta and Gupta, 1981); later this was disproved by Nagaraja (1982) and Arnold et al. (1984). Nagaraja's (1990) lucid survey article on order statistics from discrete distributions documents a number of characterizations of the geometric distribution based on its order statistics; see Section 5.9.1.

Estimation of the parameter of the geometric distribution is particularly straightforward. Because it is a PSD, the first-moment equation is also the maximum-likelihood equation. Hence

$$\hat{P} = \overline{x}. \tag{5.12}$$

A moment-type estimator for the geometric distribution with either or both tails truncated was obtained by Kapadia and Thomasson (1975), who compared its efficiency with that of the maximum-likelihood estimator (MLE). Estimation for the geometric distribution with unknown P and unknown location parameter was studied by Klotz (1970) (maximum-likelihood estimation), Iwase (1986) (minimum-variance unbiased estimation), and Yanagimoto (1988) (conditional maximum-likelihood estimation). Vit (1974) examined tests for homogeneity.

If $X_1, X_2, \ldots, X_k$ are rv's each with the geometric distribution with pmf

$$\Pr[X = x] = Q^{-1}\left(\frac{P}{Q}\right)^x, \qquad x = 0, 1, 2, \ldots,$$

then $\sum_{i=1}^k X_i$ is a negative binomial variable with parameters k and P; see Section 5.5. Using this fact, Clemans (1959) constructed charts from which confidence intervals for P, given $k^{-1}\sum_{i=1}^k x_i$, can be read off.

Applications of the geometric distribution include runs of one plant species with respect to another in transects through plant populations (Pielou, 1962, 1963), a ticket control problem (Jagers, 1973), a surveillance system for congenital malformations (Chen, 1978), and estimation of animal abundance (Seber, 1982b). Mann et al. (1974) looked at applications in reliability theory.

The distribution is used in Markov chain models, for example, in meteorological models of weather cycles and precipitation amounts (Gabriel and Neumann, 1962). Many other applications in queueing theory and applied stochastic models were discussed by Taylor and Karlin (1998) and Bhat (2002). Daniels (1961) investigated the representation of a discrete distribution as a mixture of geometric distributions and applied this to busy-period distributions in equilibrium queueing systems. Sandland (1974) put forward a building-society-membership scheme and a length-of-tenure scheme as models for the truncated geometric distribution with support $0, 1, \ldots, n-1$.

5.3 HISTORICAL REMARKS AND GENESIS OF NEGATIVE BINOMIAL DISTRIBUTION

Special forms of the negative binomial distribution were discussed by Pascal (1679). A derivation as the distribution of the number of tosses of a coin necessary to achieve a fixed number of heads was published by Montmort (1713) in his solution of the problem of points; see Todhunter (1865, p. 97). A very clear interpretation of the pmf as a density function was given by Galloway (1839, pp. 37–38) in his discussion of the problem of points. Let X be the rv representing the number of independent trials necessary to obtain k occurrences of an event that has a constant probability of occurring at each trial. Then

$$\Pr[X = k + j] = \binom{k+j-1}{k-1} p^k (1-p)^j, \qquad j = 1, 2, \ldots; \tag{5.13}$$

that is, X has a negative binomial distribution (case 2 in Kemp's list in the previous section).

Meyer (1879, p. 204) obtained the pmf as the probability of exactly j male births in a birth sequence containing a fixed number of female births; he assumed a known constant probability of a male birth. He also gave the cdf in a form that we now recognize as the upper tail of an F distribution (equivalent to an incomplete beta function; see Section 5.6).

Student (1907) found empirically that certain hemocytometer data could be fitted well by a negative binomial distribution. Whittaker (1914) continued this approach. Unfortunately she did not realize that the Poisson distribution is a limiting form for both the binomial and the negative binomial distributions (see Section 5.12.1), and she aroused considerable controversy concerning the relative merits of the Poisson and the negative binomial distributions.

Greenwood and Yule (1920) derived the following relationship between the Poisson and negative binomial distributions. Suppose that we have a mixture of Poisson distributions such that the expected values θ of the Poisson distributions vary according to a gamma distribution with pdf

$$f(\theta) = [\beta^\alpha \Gamma(\alpha)]^{-1} \theta^{\alpha-1} \exp\left(-\frac{\theta}{\beta}\right), \qquad \theta > 0, \qquad \alpha > 0, \qquad \beta > 0.$$

Then

$$\begin{aligned}\Pr[X = x] &= [\beta^\alpha \Gamma(\alpha)]^{-1} \int_0^\infty \theta^{\alpha-1} e^{-\theta/\beta} (\theta^x e^{-\theta}/x!)\, d\theta \\ &= \binom{\alpha + x - 1}{\alpha - 1} \left(\frac{\beta}{\beta+1}\right)^x \left(\frac{1}{\beta+1}\right)^\alpha. \end{aligned} \tag{5.14}$$

So X has a negative binomial distribution with parameters α and β. This type of model was used to represent "accident proneness" by Greenwood and Yule. The parameter θ represents the expected number of accidents for an individual. This is assumed to vary from individual to individual.

Another important derivation is that of Lüders (1934); see also Quenouille (1949). Here the negative binomial arises as the distribution of the sum of N independent random variables each having the same logarithmic distribution (Chapter 7), where N has a Poisson distribution. Thyrion (1960) called this an *Arfwedson process*. Boswell and Patil (1970) termed it a Poisson sum (Poisson–stopped sum) of logarithmic rv's. Let $Y = X_1 + X_2 + \cdots + X_N$, where the X_i are iid logarithmic rv's with pgf $\ln(1 - \theta z)/\ln(1 - \theta)$. Assume also that N is a Poisson rv (with parameter λ) which is independent of the X_i. Then the pgf of Y is

$$\exp\left[\lambda \left(\frac{\ln(1-\theta z)}{\ln(1-\theta)} - 1\right)\right] = \left(\frac{1-\theta}{1-\theta z}\right)^{-\lambda/\ln(1-\theta)}; \tag{5.15}$$

see Section 7.1.2.

The negative binomial as a limiting form for Pólya and Eggenberger's urn model was mentioned in Section 5.1. Consider a random sample of n balls from an urn containing Np white balls and $N(1 - p)$ black balls. Suppose that after each draw the drawn ball is replaced together with $c = N\beta$ others *of the same color*. Let the number of white balls in the sample be X. Then

$$\Pr[X = x] = \binom{n}{x} \left(\frac{p}{\beta}\right)^{[x]} \left(\frac{q}{\beta}\right)^{[n-x]} \Bigg/ \left(\frac{1}{\beta}\right)^{[n]}, \tag{5.16}$$

where $a^{[x]} = a(a + 1) \cdots (a + x - 1)$; see Section 6.2.4. The limiting form as $n \to \infty$, $p \to 0$, $\beta \to 0$ such that $np \to \eta k$, $n\beta \to \eta$ is negative binomial with pmf (5.1), where $P = \eta$; see Eggenberger and Pólya (1923, 1928).

This limiting form of (5.16) has been called a "Pólya" distribution by, for instance, Gnedenko (1961), Arley and Buch (1950), and Hald (1952). On the other hand, Bosch (1963) called (5.16) a "Pólya" distribution.

Patil and Joshi (1968) called the negative binomial a "Pólya–Eggenberger" and the distribution (5.16) simply a "Pólya" distribution. Proofs of the limiting form appear in Bosch (1963), Lundberg (1940), Feller (1968), and Boswell and Patil (1970). Thompson (1954) showed that a negative binomial distribution can also be obtained (approximately) from a modified form of Neyman's contagious distribution model (Section 9.6).

Distribution (5.16) arises also as a beta mixture of binomial distributions (Skellam, 1948); see Sections 6.2.2 and 8.3.4. Boswell and Patil (1970) derived the negative binomial as a limiting form of this mixture of binomials.

Feller (1957, p. 253) pointed out that the negative binomial can be regarded as a convolution of a fixed number of geometric distributions; here, as for the inverse sampling model (5.13), the exponent k is necessarily an integer. Maritz (1952) considered how the negative binomial could arise from the addition of a set of correlated Poisson rv's. Kemp (1968a) showed that weighting a negative binomial with parameters k and $P = q/p$ using the weight function (sampling chance) $a_x = \alpha^x$ gives another negative binomial but with parameters k and $\alpha q/(1-\alpha q)$.

Bhattacharya (1966) obtained the negative binomial by mixing his confluent hypergeometric distributions with a "generalized exponential" distribution; the pgf of the outcome is

$$\int_0^\infty \frac{{}_1F_1(a;b;\theta z)}{{}_1F_1(a;b;\theta)} \times \frac{c^a(c+1)^{b-a}\theta^{b-1}e^{-(c+1)\theta}\ {}_1F_1(a;b;\theta)\,d\theta}{\Gamma(b)} = \frac{[1-1/(c+1)]^a}{[1-z/(c+1)]^a}. \tag{5.17}$$

Bhattacharya showed that the generalized exponential mixing distribution is unique by virtue of the uniqueness of the Mellin transform (Section 1.1.10), and he applied his results to the theory of accident proneness in the case where $a = 1$ and the number of accidents sustained by an individual has a sub-Poisson distribution (Section 4.12.4).

The result of mixing negative binomials with constant exponent parameter k using a beta distribution with parameters c and $k-c$, where $k > c > 0$, is another negative binomial distribution with exponent parameter c; we have

$$\int_0^1 [(1+P\theta-P\theta z)^{-k}]\frac{\theta^{c-1}(1-\theta)^{k-c-1}\,d\theta}{B(c,k-c)} = (1+P-Pz)^{-c}. \tag{5.18}$$

A mixture of Katti (1966) type H_2 distributions (Section 6.11) using a particular beta distribution can also yield a negative binomial distribution:

$$\begin{aligned}\int_0^1 {}_2F_1[k,a;b;\,P\theta(z-1)]\frac{\theta^{b-1}(1-\theta)^{a-b-1}\,d\theta}{B(b,a-b)} &= {}_3F_2[k,a,b;b,a;\,P(z-1)]\\ &= (1+P-Pz)^{-k}.\end{aligned} \tag{5.19}$$

Also a gamma mixture of Poisson $\bigwedge$ beta distributions (Section 8.3.3) gives rise to a negative binomial:

$$\begin{aligned}\int_0^\infty {}_1F_1[a; a+b; P\theta(z-1)]\frac{e^{-\theta}\theta^{a+b-1}\,d\theta}{\Gamma(a+b)}\\ = {}_2F_1[a, a+b; a+b; P(z-1)]\\ = (1+P-Pz)^{-a}. \qquad (5.20)\end{aligned}$$

These results are special cases of those in Section 8.3.6; see also Kemp (1968a).

The negative binomial arises also from several well-known stochastic processes. The time-homogeneous birth-and-immigration process with zero initial population was first obtained by McKendrick (1914); the equivalence of the distributions arising from this process, from Greenwood and Yule's model as a gamma mixture of Poisson distributions, and from Lüders and Quenouille's Poisson–stopped sum of logarithmic distributions model was discussed by Irwin (1941). The nonhomogeneous process with zero initial population known as the *Pólya process* was developed by Lundberg (1940) in the context of risk theory. Other stochastic processes that lead to the negative binomial include the simple birth process with nonzero initial population size (Yule, 1925; Furry, 1937), Kendall's (1948) nonhomogeneous birth-and-death process with zero death rate, and the simple birth-death-and-immigration process with zero initial population of Kendall (1949).

The geometric distribution is the equilibrium distribution of queue length for the M/M/1 queue, while the negative binomial is the equilibrium queue length distribution for the M/M/1 queue with a particular form of balking; see Haight (1957) and also Bhat (2002). The negative binomial can also be obtained as the equilibrium solution for a particular type of Markov chain known as a Foster process (Foster, 1952).

5.4 MOMENTS

From the pgf $(1+P-Pz)^{-k}$, it follows that the factorial moment generation function (fmgf) is $(1-Pt)^{-k}$, and so

$$\mu'_{[r]} = \frac{(k+r-1)!}{(k-1)!}P^r, \qquad r = 1, 2, \ldots. \qquad (5.21)$$

Also the factorial cumulant generating function (fcgf) is $-k\ln(1-Pt)$, whence

$$\kappa_{[r]} = k(r-1)!P^r, \qquad r = 1, 2, \ldots. \qquad (5.22)$$

The relationship with the binomial pgf is readily apparent—replacing N by $-k$ and π by $-P$ in the well-known formulas for the moment properties of the binomial distribution gives the corresponding formulas for the negative binomial distribution. In particular

$$\begin{aligned}
\mu_r &= kPQ\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_j + P\sum_{j=0}^{r-2}\binom{r-1}{j}\mu_{j+1},\\
\mu &= \kappa_1 = kP = \frac{kq}{p},\\
\mu_2 &= \kappa_2 = kP(1+P) = \frac{kq}{p^2},\\
\mu_3 &= \kappa_3 = kP(1+P)(1+2P) = \frac{kq(1+q)}{p^3},\\
\mu_4 &= 3k^2P^2(1+P)^2 + kP(1+P)(1+6P+6P^2)\\
&= \frac{3k^2q^2}{p^4} + \frac{kq(p^2+6q)}{p^4},
\end{aligned} \tag{5.23}$$

and

$$\begin{aligned}
\sqrt{\beta_1} &= \frac{1+2P}{[kP(1+P)]^{1/2}} = \frac{1+q}{\sqrt{kq}},\\
\beta_2 &= 3 + \frac{1+6P+6P^2}{kP(1+P)} = 3 + \frac{p^2+6q}{kq},
\end{aligned} \tag{5.24}$$

where $p = 1/(1+P) = Q^{-1}$ and $q = P/(1+P) = PQ^{-1}$.

The alternative notation, with the pgf in the form $p^k(1-qz)^{-k}$, shows that the mgf is $p^k(1-qe^t)^{-k}$, the central mgf is $e^{-kqt/p}p^k(1-qe^t)^{-k}$, whence

$$\mu_{r+1} = q\frac{\partial\mu_r}{\partial q} + \frac{rkq}{p^2}\mu_{r-1},$$

and the cumulant generating function (cgf) is $k\ln p - k\ln(1-qe^t)$.

Because this is a PSD with series parameter q, the cumulants satisfy

$$\kappa_{r+1} = q\frac{\partial\kappa_r}{\partial q}, \qquad r = 1, 2, \ldots. \tag{5.25}$$

The distribution is overdispersed (variance greater than the mean), with an index of dispersion equal to $p^{-1} = 1 + P$. The coefficient of variation is $(kq)^{-1/2} = [(1+P)/(kP)]^{1/2}$.

The factorial moments of negative order are

$$\mu_{-k} = E\left[\frac{X!}{(X+k)!}\right] = \int_0^1\int_0^{t_{k-1}}\cdots\int_0^{t_1}\left(\frac{1-q}{1-qz}\right)^k ds\,dt_1\cdots dt_{k-1};$$

in particular

$$\mu_{-1} = E\left[\frac{1}{X+1}\right] = \frac{(1-q)^k}{q(1-k)}[1-(1-q)^{1-k}]$$

(Balakrishnan and Nevzorov, 2003).

The mean deviation is

$$\nu_1 = \frac{2m(k+m-1)!P^m}{m!(k-1)!Q^{m+k-1}} = \frac{2m(k+m-1)!p^{k-1}q^m}{m!(k-1)!}, \tag{5.26}$$

where $m = [kP] + 1$ (that is, m is the smallest integer greater than the mean μ); see Kamat (1965).

5.5 PROPERTIES

From the relationship

$$\frac{\Pr[X = x+1]}{\Pr[X = x]} = \frac{(k+x)P}{(x+1)Q}, \tag{5.27}$$

it can be seen that

$$\Pr[X = x+1] < \Pr[X = x] \quad \text{if } x > kP - Q$$

and that

$$\Pr[X = x] \geq \Pr[X = x-1] \quad \text{if } x \leq kP - P. \tag{5.28}$$

So when $(k-1)P$ is not an integer, there is a single mode at $[(k-1)P]$, where $[\cdot]$ denotes the integer part. When $(k-1)P$ is an integer, then there are two equal modes at $X = (k-1)P$ and $X = kP - Q$. If $kP < Q$, the mode is at $X = 0$.

Van de Ven and Weber (1993) have obtained bounds for the median of the negative binomial distribution which are valid for all parameter values. Their definition of the median is $\inf\left\{x : \Pr[X \leq x] \geq \frac{1}{2}\right\}$. Göb (1994) commented that long-standing inequalities for the percentage points of the binomial cdf provide bounds for the binomial median. He then used the relationship between the binomial and the negative binomial cdf's to obtain bounds for the negative binomial median.

For fixed values of x and k the probabilities increase monotonically with P; for fixed x and P they increase monotonically with k.

When $k < 1$, we have $p_x p_{x+2}/p_{x+1}^2 > 1$ (where $p_x = \Pr[X = x]$), and therefore the probabilities are logconvex; when $k > 1$, we have $p_x p_{x+2}/p_{x+1}^2 < 1$ and so now the probabilities are logconcave. Although the probabilities satisfy the logconvexity condition that is a sufficient condition for infinite divisibility only when $k < 1$, nevertheless the distribution is a Poisson–stopped sum of logarithmic rv's and so is infinitely divisible for all values of k.

The logconvexity/logconcavity properties imply that the distribution has a decreasing hazard (failure) rate for $k < 1$ and an increasing hazard rate for $k > 1$. For $k = 1$ the failure rate is constant. This is the no-memory (Markovian) property of the geometric distribution; see Sections 5.2 and 5.9.1.

If X_1 and X_2 are independent variables each having a negative binomial distribution with the same series parameter q but with possibly different power parameters k_1 and k_2, then $X_1 + X_2$ also has a negative binomial distribution; its pgf is

$$(1 + P - Pz)^{-k_1-k_2} = p^{k_1+k_2}(1 - qz)^{-k_1-k_2}. \tag{5.29}$$

As k tends to infinity and P to zero, with kP remaining fixed ($kP = \theta$), the right-hand side of (5.1) tends to the value $e^{-\theta}\theta^k/k!$, corresponding to a Poisson distribution with expected value θ.

Young (1970) gave formulas for the moments of the order statistics for the negative binomial distribution and tabulated $E[X_{(r)}]$ for samples of size $n = 2(1)8$ to two decimal places. He showed that when $p = Q^{-1}$ is close to unity there is a good gamma approximation, enabling Gupta's (1960a) tables for gamma order statistics to be used.

Pessin (1961, 1965) noted that, as $Q \to \infty$ with k constant, the standardized negative binomial tends to a gamma distribution.

5.6 APPROXIMATIONS AND TRANSFORMATIONS

The sum of a number of negative binomial terms can be expressed in terms of an incomplete beta function ratio and hence as a sum of binomial terms. We have

$$\begin{aligned}\sum_{j=r}^{\infty} \Pr[X = j] &= \frac{(k+r-1)!p^k q^r}{(k-1)!r!}\left(1 + \frac{(k+r)q}{(r+1)} + \cdots\right)\\ &= \frac{(k+r-1)!p^k q^r}{(k-1)!r!}\, {}_2F_1[1, k+r; r+1; q]\\ &= \frac{(k+r-1)!q^r}{(k-1)!r!}\, {}_2F_1[r, 1-k; r+1; q]\\ &= \frac{B_q(r,k)}{B(r,k)} = I_q(r,k).\end{aligned} \tag{5.30}$$

Therefore

$$\begin{aligned}\Pr[X \le r] = 1 - I_q(r+1, k) &= I_p(k, r+1)\\ &= \Pr[Y \ge k],\end{aligned} \tag{5.31}$$

where Y is a binomial rv with pgf $(q + pz)^{k+r}$. This formula has been rediscovered on many occasions. Patil (1963a) gives a list of references; see also Morris (1963). Approximations for binomial distributions (already discussed in Section 3.6.1) can thereby be applied to negative binomial distributions.

Negative binomial approximations to the negative hypergeometric probabilities have been obtained by López-Blázquez and Salamanca-Miño (2001).

Bartko (1966) studied five different approximations for cumulative negative binomial probabilities. Their accuracy is similar to that of approximations for the binomial distribution. The two most useful approximations in Bartko's opinion are as follows:

1. A corrected (Gram–Charlier) Poisson approximation

$$\Pr[X \le x] = e^{-kP} \sum_{j=0}^{x} \frac{(kP)^j}{j!} - \frac{(x - kP)}{2(1 + P)} e^{-kP} \frac{(kP)^x}{x!}. \tag{5.32}$$

2. The Camp–Paulson approximation (see Johnson et al., 1995, Chapter 26)

$$\Pr[X \le x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K} e^{-u^2/2}\, du,$$

where

$$K = \frac{\left[\dfrac{9x + 8}{x + 1} - \dfrac{(9k - 1)\{kP/(x + 1)\}^{1/3}}{k}\right]}{3\left[\dfrac{\{kP/(x + 1)\}^{2/3}}{k} - \dfrac{1}{x + 1}\right]^{1/2}}. \tag{5.33}$$

Of these (5.33) is remarkably accurate, but it is much more complicated than (5.32).

Peizer and Pratt (1968) and Pratt (1968) have obtained extremely accurate normal approximations of the form

$$\Pr[X \le r] = \int_{-\infty}^{z_i} (2\pi)^{-1/2} e^{-x^2/2}\, dx,$$

where

$$\begin{aligned} z_i &= d_i \left(\frac{1 + pg(a) + qg(b)}{(r + k + 1/6)pq}\right)^{1/2}, \qquad i = 1, 2, \\ a &= \frac{r + 0.5}{(r + k)q}, \qquad b = \frac{k - 0.5}{(r + k)p}, \qquad g(u) = \frac{1 + u}{1 - u} + \frac{2u \ln u}{(1 - u)^2}, \end{aligned} \tag{5.34}$$

and

$$d_1 = \left(r + \frac{2}{3}\right) p - \left(s - \frac{1}{3}\right) q, \qquad d_2 = d_1 + 0.02 \left(\frac{p}{r + 1} - \frac{q}{k} + \frac{p - 0.5}{r + k + 1}\right).$$

Peizer and Pratt provided a table giving evidence concerning the remarkable accuracy of the two approximations.

Guenther's (1972) approximation is based on the incomplete gamma function; it is

$$\Pr[X \leq r] \approx \Pr\left[\chi^2_{2kq} \leq (2r+1)p\right], \tag{5.35}$$

where χ^2_{2kq} is a chi-squared variable; see Johnson et al. (1994, Chapter 17). It enables tables of the incomplete gamma function to be used. Best and Gipps (1974) presented evidence that

$$\Pr[X \leq r] \approx \Pr\left[\chi^2_{8kq/(q+1)^2} \leq \frac{[4r+2+4kq/(1+q)]p}{1+q}\right] \tag{5.36}$$

provides a considerable improvement over (5.35).

A transformation that approximately normalizes and approximately equalizes the variance is useful. The formulas $E[X] = kP$, $\operatorname{Var}(X) = kP(1+P)$ suggest the transformation

$$Y_1 = \sqrt{k}\sinh^{-1}\sqrt{\frac{X}{k}}, \tag{5.37}$$

with Y_1 approximately distributed as a standard normal variable.

More detailed investigations by Anscombe (1948) indicated that the transformation

$$Y_2 = \sqrt{k-0.5}\sinh^{-1}\sqrt{\frac{X+\frac{3}{8}}{k-\frac{3}{4}}} \tag{5.38}$$

is preferable; see also Laubscher (1961).

5.7 COMPUTATION AND TABLES

Computation of individual negative binomial probabilities can be reduced to the calculation of the corresponding binomial probabilities by the use of the relationship between the tails of the binomial and negative binomial distributions; see the previous section. For low values of r the probabilities can also be computed by recursion from $\Pr[X = 0]$ using

$$\Pr[X = r+1] = \frac{(k+r)q}{r+1}\Pr[X = r]. \tag{5.39}$$

For higher values of r Stirling's expansion (Section 1.1.2) can be used for the gamma functions in the expression

$$\Pr[X = r+1] = \frac{\Gamma(k+r)p^k q^r}{\Gamma(r+1)\Gamma(k)};$$

this gives

$$\ln \Pr[X = r] \approx (k-1)\ln\left(\frac{(k+r)p}{k}\right) + (r+0.5)\ln\left(\frac{(k+r)q}{r}\right) - 0.5\ln\left(\frac{2\pi kq}{p^2}\right) - \frac{1}{12k} - \frac{k}{12r(k+r)}. \tag{5.40}$$

Cumulative negative binomial probabilities can be computed from cumulative binomial probabilities or by summation of individual probabilities; alternatively they can be approximated using appropriate formulas from the previous section.

However, for fractional values of k, and for convenience in looking up sequences of values, direct tables can be useful. Williamson and Bretherton (1963) provided comprehensive six-decimal tables of $\Pr[X = r]$.

Grimm (1962) gave values of individual probabilities and of the cumulative distribution function to five decimal places; see also Brown (1965). Taguti (1952) gave minimum values of r for which

$$\sum_{j=0}^{r} (j!)^{-1} h(h+d)\cdots[h+(j-1)d](1+d)^{-(h/d)-j} \geq \alpha$$

for $\alpha = 0.95, 0.99$; these are (approximate) percentage points of negative binomial distributions with $k = h/d$, $P = d$.

Computer generation of rv's from a geometric distribution is very straightforward. One method is to exploit the waiting-time property. Consider a stream of uniform rv's. Then geometric rv's can be generated by counting the number of uniforms needed to obtain a uniform less than p (the number of failures needed to obtain the first success). Devroye (1986, p. 498) considers that "for $p \geq \frac{1}{3}$ the method is probably difficult to beat in any programming environment."

A second way to generate a geometric rv G is by analytic inversion of the cdf. Let U be a uniform rv. Then $G = [\ln(U)/\ln(1-p)]$ (where $[\cdot]$ denotes the integer part). If a stream of exponential rv's is available, then discretizing the exponential (E) gives $G = [-E/\ln(1-p)]$. Devroye notes (in an exercise) that there may be an accuracy problem for low values of p and that one way that this may be overcome is via the expansion

$$\ln(1-p) = \frac{2}{c}\left(1 + \frac{1}{3c^2} + \frac{1}{5c^4} + \cdots\right),$$

where $c = 1 - 2/p$ (c is negative).

The negative binomial with an integer parameter $k = N$ can be generated as the sum of N geometric rv's. Except for low values of N (say $N = 2, 3, 4$), this method cannot be advocated as it requires many uniforms for a single output negative binomial rv. This argument applies a fortiori to the use of the sum of a Poisson number of logarithmic rv's.

The method generally recommended for generating negative binomial rv's with changing parameters is to generate Poisson rv's with random parameters drawn from a gamma distribution [see, e.g., algorithm NB3 in Fishman (1978)]. For fixed parameters the use of a fast general method, such as indexed table look-up, alias, or frequency table, is recommended.

Three simple stochastic models that can be used to generate *correlated* negative binomial rv's have been described by Sim and Lee (1989). Two of their methods are based on the autoregressive scheme of the first-order Markovian process. The third uses the Poisson process from a first-order autoregressive gamma sequence.

5.8 ESTIMATION

5.8.1 Model Selection

Early graphical methods for identifying whether or not a negative binomial model is appropriate for a particular type of data were based on ratios of factorial moments (Ottestad, 1939), or probability-ratio cumulants (Gurland, 1965), or ratios of factorial cumulants (Hinz and Gurland, 1967). Ord's method of plotting $u_r = rf_r/f_{r-1}$ against r (where f_r is an observed frequency) gives an upward-sloping straight line, $u_r \approx (k + r - 1)p$; see Ord (1967a, 1972) and Tripathi and Gurland (1979). Grimm's (1970) method and the methods of Hoaglin, Mosteller, and Tukey (1985) can also be used; see Section 4.7.1.

5.8.2 *P* Unknown

Consider the total number of trials $k + X$ needed to obtain k successes when the probability of a success is p (the inverse sampling model). Then the minimum variance unbiased estimator (MVUE) of $p = (1 + P)^{-1}$, based on a single observation x of X, is

$$p^{\circ} = \frac{k-1}{k+x-1}, \tag{5.41}$$

and the Cramér–Rao lower bound on its variance is

$$\operatorname{Var}(p^{\circ}) \geq \frac{p^2 q}{k}. \tag{5.42}$$

Best (1974) stated that

$$\operatorname{Var}(p^{\circ}) = p^2 \sum_{r=1}^{\infty} \binom{k+r-1}{r}^{-1} q^r; \tag{5.43}$$

Mikulski and Smith (1976) showed that

$$\operatorname{Var}(p^{\circ}) \leq \frac{p^2 q}{k - p + 2}. \tag{5.44}$$

These bounds on the variance of p° were sharpened by Ray and Sahai (1978) and Sahai and Buhrman (1979).

Given a sample of observations from a negative binomial distribution with power parameter k, consideration of the pgf in the form $G(z) = p^k(1-qz)^{-k}$ shows that the distribution is a PSD, and hence the maximum-likelihood equation for q is the first-moment equation $\overline{x} = k\hat{q}/(1-\hat{q})$; thus

$$\hat{q} = \frac{\overline{x}}{k+\overline{x}}. \tag{5.45}$$

Roy and Mitra (1957) showed that the uniformly minimum variance unbiased estimator (UMVUE) of $P = q/p$ is $\tilde{\theta}/(1-\tilde{\theta})$, where

$$\tilde{\theta} = \frac{\sum_x x f_x}{\sum_x (k+x) f_x - 1}, \tag{5.46}$$

and the f_x are the observed frequencies. The UMVUE of μ is $\sum x f_x/(k \sum f_x)$ and the UMVUE of μ_2 is $\sum x f_x \sum (x+1) f_x/[n(n+1)]$ (Guttman, 1958).

Irony (1992) commented that the steps needed to make Bayesian inferences about q parallel those needed for Bayesian inferences about the binomial parameter p.

Maynard and Chow (1972) constructed an approximate Pitman-type "close" estimator of P for small sample sizes. Scheaffer (1976) has studied methods for obtaining confidence intervals for $p = 1-q$. Gerrard and Cook (1972) and Binns (1975) considered sequential estimation of the mean $kq/(1-q)$ when k is known.

5.8.3 Both Parameters Unknown

Consider now the situation where both parameters are unknown. Because of the variability of the sample variance of the negative binomial distribution, samples with the sample variance less than the sample mean ($s^2 < \overline{x}$) will occasionally be encountered, even when a negative binomial model is appropriate. However, when this occurs, the appropriateness of the model should be examined (see, e.g., Clark and Perry, 1989).

Method of Maximum Likelihood The maximum-likelihood estimators satisfy the equations

$$\hat{k}\hat{P} = \overline{x}, \tag{5.47}$$

$$\ln(1+\hat{P}) = \sum_{j=1}^{\infty} \left((\hat{k}+j-1)^{-1} \sum_{i=j}^{\infty} f_j \right), \tag{5.48}$$

where f_j is an observed frequency; see Fisher (1941), Bliss and Fisher (1953), and Wise (1946). Iteration is required for the solution of these equations. It is important to realize that iteration may be very slow if the initial estimates are poor. Rapid explicit methods that can provide good initial estimates have therefore been studied extensively. Ross and Preece (1985) have advocated the use of the maximum-likelihood program (MLP) of Ross (1980).

Method of Moments The simplest way to estimate the parameters is by the method of moments, that is, by equating the sample mean $\overline{x}$ and sample variance s^2 to the corresponding population values.

Thus, if $x_1, x_2, \ldots, x_n$ are n observed values (supposed independent), we calculate the solutions $\tilde{k}$, $\tilde{P}$ of the equations

$$\tilde{k}\tilde{P} = \overline{x} \qquad \text{and} \qquad \tilde{k}\tilde{P}(1 + \tilde{P}) = s^2;$$

this gives

$$\tilde{P} = \frac{s^2}{\overline{x}} - 1, \qquad \tilde{k} = \frac{\overline{x}^2}{s^2 - \overline{x}}. \tag{5.49}$$

Bowman and Shenton (1965, 1966) obtained asymptotic formulas for the variances, covariances, and biases of the moment and maximum-likelihood estimators.

Method of Mean and Zero Frequency In place of (5.48) an equation obtained by equating the observed and expected numbers of zero values may be used. This equation is

$$f_0 = (1 + P^{\dagger})^{-k^{\dagger}}, \tag{5.50}$$

where f_0 is the number of zero values. Combining this equation with the equation $k^{\dagger}P^{\dagger} = \overline{x}$ gives

$$\frac{P^{\dagger}}{\ln(1 + P^{\dagger})} = -\frac{\overline{x}}{\ln f_0}. \tag{5.51}$$

Other Methods Gurland (1965) and Gurland and Tripathi (1975) put forward a method based on the solution of linear equations involving functions of the moments and/or frequencies; see also Katti and Gurland (1962a). Gurland (1965) and Hinz and Gurland (1967) concluded that estimators based on the factorial cumulants and a certain function of the zero frequency have good efficiency relative to maximum likelihood.

Pieters et al. (1977) made small-sample comparisons of various methods using simulation. Willson, Folks, and Young (1984) extended this work by considering not only the bias but also the standard deviation and the mean square error of the method of moments and maximum-likelihood estimators and by comparing these to a proposed multistage estimation procedure. In her comment on their work, Bowman (1984) pointed out the riskiness in depending on small samples when estimating k and questioned the choice of $n = 5$ as the initial sample size for the multistage procedure.

A. W. Kemp (1986) sought to explain Anscombe's findings regarding the efficiencies of certain methods by showing that the solution to an approximation to the maximum-likelihood equations that is valid for a certain region of the parameter space can give useful approximations to the MLEs over that region. A. W. Kemp and C. D. Kemp (1987) went further with this approach by using (5.47) and an approximation to the digamma function implicit in (5.48) to obtain the equations

$$k^* = \frac{\overline{x}\overline{x}_w(\eta - 1)}{\overline{x}_w - \overline{x}\eta}, \qquad P^* = \frac{\overline{x}}{k^*}, \tag{5.52}$$

where $\overline{x}_w = \sum_x x f_x \eta^x$ is a weighted mean of the sample distribution and

$$\eta = \begin{cases} \dfrac{\tilde{k}+1}{\tilde{k}+2} & \text{for } \overline{x} \le 2, \\[2ex] \dfrac{\tilde{k}+\overline{x}/2}{\tilde{k}+1+\overline{x}/2} & \text{for } \overline{x} > 2, \end{cases} \tag{5.53}$$

where $\tilde{k}$ is the moment estimate of k. A large simulation study confirmed that these explicit estimators have high efficiency relative to maximum likelihood for nearly all of the parameter space.

Anraku and Yanagimoto (1990) have adopted the parameterization

$$G(z) = (1 + \theta\mu - \theta\mu z)^{-1/\theta};$$

this is obtained by setting $P = \theta\mu$, $k = 1/\theta$ in (5.1), where μ is the population mean. In this parameterization the index of dispersion is $\mu_2/\mu = 1 + \theta\mu$. The authors' focus of attention is the estimation of θ (the "dispersion parameter" in their paper) conditional on a knowledge of μ. They commented that "$\overline{x}$ is a reasonable estimator of μ irrespective of the estimator of θ" and obtained the conditional likelihood

$$L_c(x;\theta) = \sum_{i=1}^{n}\sum_{j=1}^{x_i} \frac{j-1}{1+\theta(j-1)} - \sum_{i=1}^{t} \frac{i-1}{n+\theta(i-1)}, \tag{5.54}$$

where $t = \sum X_i$ when $s^2 > \overline{x}$. This is maximized by their conditional MLE θ_c. For $s^2 \le \overline{x}$ their estimator is defined to be zero. Anraku and Yanagimoto have investigated the biases associated with θ_c, compared with the unconditional MLE and the moment estimator of θ, both theoretically and via simulation.

Clark and Perry (1989, p. 310) advocated the use of the parameter $\beta = 1/k$ instead of k because this "avoids problems caused by infinite values of $\hat{k}$ when $s^2 = \overline{x}$; also confidence intervals for β are continuous and usually more symmetric than those for k, which may be discontinuous." Clark and Perry's maximum quasi-likelihood estimator $\hat{\beta}$ of β is obtained by solving iteratively

$$\sum_{i=1}^{n}\left[\hat{\beta}^{-2}\ln\left(\frac{1+2m}{1+\hat{\beta}x_i}\right) - \frac{x_i}{1+\hat{\beta}x_i} + \frac{1+6x_i}{2(\hat{\beta}+6+6\hat{\beta}x_i)}\right] = \frac{n}{2(\hat{\beta}+6)}, \tag{5.55}$$

where m is taken to be $\hat{\mu}$ and x_i, $i = 1, 2, \ldots, n$, are the observations. The authors indicated how a 95% profile likelihood confidence interval for β can be constructed. Their simulation results showed that maximum quasi-likelihood estimation performed slightly better than estimation by the method of moments, except when the sample size is no more than 20. Piegorsch (1990) carried out further simulation studies, showing that maximum quasi-likelihood estimation appears to be slightly less biased than maximum-likelihood estimation or the method of moments; he recommended the use of maximum quasi-likelihood estimation, provided that the sample size is greater than 20 and that β is "not very small."

Maximum-likelihood fitting of a negative binomial distribution to coarsely grouped data was described by O'Carroll (1962).

5.8.4 Data Sets with a Common Parameter

The estimation of an (assumed) common value of k from data from a number of negative binomial distributions was discussed briefly by Anscombe (1950) and in more detail by Bliss and Owen (1958). Suppose that in a sequence of n samples of sizes n_i, $i = 1, 2, \ldots, n$, the observed means and standard deviations are denoted by $\overline{x}_i$ and s_i^2, respectively, where $i = 1, 2, \ldots, n$. Then the moment estimators of k are

$$\tilde{k}_i = \frac{\overline{x}_i^2}{s_i^2 - \overline{x}_i}, \qquad i = 1, 2, \ldots, n,$$

and their approximate variances are

$$\frac{2k(k+1)(Q_i/P_i)^2}{n_i}, \qquad i = 1, 2, \ldots, n.$$

As a first approximation, the weights $w_i = n_i(\tilde{P}_i/\tilde{Q}_i)^2$ may be used, giving

$$k^{(1)} = \frac{\sum_{i=1}^{n} w_i \tilde{k}_i}{\sum_{i=1}^{n} w_i}.$$

Using this value, $k^{(1)}$, new estimates of P_i and new weights can be calculated; these can then be used to obtain an improved estimate of k.

A standard method for testing the hypothesis of a common value of k was provided by Bliss and Fisher (1953).

Testing for a common value of P has also been investigated by Meelis (1974), who distinguished three situations:

1. P known to the experimenter.
2. P unknown to the experimenter and the sample sizes for the n samples all equal.
3. P unknown to the experimenter and the sample sizes not necessarily equal.

Freeman (1980) investigated methods based on minimum χ^2 for fitting two-parameter discrete distributions to many data sets with one common parameter. He discussed fitting negative binomials with a common value of k and also with a common value of P; he examined the meanings of these two models with particular reference to Taylor's power law (Taylor, 1961).

5.8.5 Recent Developments

The past 15 years have seen a great deal of research on inference for the negative binomial distribution.

The papers by Hubbard and Allen (1991) and Heffernan (1996) are on sequential probability ratio tests for the mean. The usual sequential probability ratio test assumes that k is known. Hubbard and Allen present analytic and simulation results concerning robustness of the test to misspecification of k. Heffernan has developed a method for assessing the performance of sequential probability ratio tests without making the assumption of a common k by using a model of the dependence of k on the mean.

Wang, Young, and Johnson (2001) give a uniformly most powerful unbiased test for the equality of means from independent samples from two negative binomial populations given a common dispersion parameter. They compare the size and power of their test with several other known tests.

A score test for testing zero-inflated Poisson regression models against zero-inflated negative binomial alternatives has been put forward by Ridout, Hinde, and Demétrio (2001).

Goodness-of-fit tests for the geometric distribution were studied in depth by Best and Rayner (2003) using simulation. The tests include a Chernoff–Lehmann χ^2-test, smooth tests of the type introduced by Best and Rayner (1989) and Rayner and Best (1989), a Kolmogorov–Smirnov test, and an Anderson–Darling test. They recommended a data-dependent Chernoff–Lehmann χ^2-test and the Anderson–Darling test.

Mulekar and Young (1991, 1993) and Mulekar (1995) deal with fixed sample size selection procedures for selecting the "best" of k negative binomial populations. The exponent r is assumed to be known and the same for all populations. A large sample approximation for the least favorable configuration is put forward in the 1991 paper and the sample sizes obtained using it are compared with those obtained using the exact least favorable configuration. In the 1993 paper selection is made in a way such that the probability of correct selection is at least P^* whenever the distance between the probabilities of success is at least δ^*. The smallest sample sizes needed to meet the specifications (P^*, δ^*) are tabulated. The 1995 paper adopts an indifference zone approach.

The uniqueness of the solution of the maximum-likelihood equations has been the subject of a series of papers. Anscombe (1950) conjectured that the maximum-likelihood equations for the two-parameter negative binomial distribution have a unique solution iff $s^2 > \overline{x}$. Aragon, Eberly, and Eberly (1992) gave a proof of this conjecture, but Wang (1996) pointed out an error in their proof. The intention of Ferreri (1997) was to settle the matter in favor of uniqueness when $s^2 > \overline{x}$ (the case $s^2 < \overline{x}$ and $\overline{x} > c - 1$, where c is the largest observation, seems to be a counterexample). Kokonendji (1999) refers to a proof in Levin and Reeds (1977) and has extended it to show that the maximum-likelihood estimate and the moment estimate of the index parameter exist iff $s^2 > \overline{x}$ for Jain and Consul's (1971) generalized negative binomial distribution.

Wang (1996) also showed that an unbiased estimator of k does not exist.

The MLE and the UMVUE of $P[X \le Y]$, where X and Y are both negative binomial rv's, have been derived by Sathe and Dixit (2001).

Many researchers had realized the advantage gained by reparameterizing to μ, k. Klugman, Panjer, and Willmot (1998) point this out very clearly. In particular it means that for maximum-likelihood estimation only one equation has to be solved iteratively. In their book these authors give a detailed account of the method of moments and maximum-likelihood estimation. They mention the use of Newton–Raphson iteration, also numerical root-finding methods (bisection, secant), and discuss the scoring method.

Clark and Perry (1989) had used a large simulation study to compare the method of moments with maximum quasi-likelihood estimation. The simulations by Piegorsch (1990) compared maximum-likelihood estimation with maximum quasi-likelihood estimation and the method of moments. The focus in Anraku and Yanagimoto (1990) was on conditional maximum-likelihood estimators. Van de Ven's (1993) large-scale simulation study enabled all four methods to be compared. He used two assessment procedures. The traditional one favored conditional maximum-likelihood estimation; the newer one favored maximum-likelihood estimation.

Nakashima (1997) was worried by the lack of robustness of maximum-likelihood estimation. He calculated efficiencies for the method of moments and for pseudolikelihood maximum-likelihood estimation. He considered that both methods are highly efficient when there is not much overdispersion.

The use of the EM (expectation–maximization) algorithm for solving the maximum-likelihood equations is described briefly by McLachlan and Krishnan (1997), who refer to papers by Schader and Schmid (1985) and Achuthan and Krishnan (1992). The EM algorithm written by Adamidis (1999), unlike that of Schader and Schmid, does not involve iteration in the M-step.

5.9 CHARACTERIZATIONS

5.9.1 Geometric Distribution

The geometric distribution is characterized by the Markovian property

$$\Pr[X = x + y | X \ge y] = \Pr[X = x]; \tag{5.56}$$

see Section 5.2. This is a discrete analog of the Markovian property of the exponential distribution (Hawkins and Kotz, 1976). Many other characterizations of the geometric are similarly analogs of exponential characterizations.

Shanbhag (1970b) showed that, if X is a discrete positive integer-valued random variable, then

$$\Pr[X > a + b | X > a] = \Pr[X > [a + b] - [a]] \tag{5.57}$$

for all a and b iff X has a geometric distribution. Similarly X has a geometric distribution iff

$$E[X | X > y] = E[X] + [y] + 1 \tag{5.58}$$

for all y, where $[y]$ denotes the integer part of y.

It is also possible to characterize any geometric distribution by the distribution of the difference between two independent rv's having the same geometric distribution (Puri, 1966). Puri also showed that, if the common distribution has parameter P, then the absolute difference distribution can be constructed as the distribution of the sum of two independent rv's, one distributed binomially with parameters $n = 1$, $p = P(1 + 2P)^{-1}$ and the other distributed geometrically with parameter P.

The independence of the difference $X_1 - X_2$ and $\min(X_1, X_2)$ for two independent rv's underlies another type of characterization. Ferguson (1964, 1965) and Crawford (1966) found that $\min(X_1, X_2)$ and $X_1 - X_2$ are independent rv's iff X_1 and X_2 are either both exponential or both geometric rv's with the same location and scale parameters; see also Srivastava (1974).

Srivastava (1965) obtained an extended form of this characterization applicable to n independent discrete rv's $X_1, X_2, \ldots, X_n$. The condition that $\min(X_1, X_2, \ldots, X_n)$ and $\sum_{i=1}^{n}[X_i - \min(X_1, X_2, \ldots, X_n)]$ are mutually independent is essential.

Srivastava (1974) showed furthermore that, if X_1 and X_2 have independent nonnegative discrete distributions, then

$$\Pr[U = j, V = l] = \Pr[U = j] \Pr[V = l] \tag{5.59}$$

[where $U = \min(X_1, X_2)$ and $V = X_1 - X_2$] for all j and $l = 1, 2$ iff X_1 and X_2 have geometric distributions with possibly different parameters.

Gupta's (1970) result based on a triplet of order statistics is as follows (in the discrete case): Let $X_1, X_2, \ldots, X_n$ be iid discrete rv's with order statistics

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

Then the independence of $X_{(i)}$ and $X_{(k)} - X_{(j)}$ for one triplet i, j, k, where $1 \leq i \leq j < k \leq n$, provides a characterization of the geometric distribution. The independence of $X_{(1)}$ and $\sum_{j=2}^{k}(X_{(j)} - X_{(1)})$ also provides a characterization (Srivastava, 1974).

Suppose now that X has an arbitrary nonnegative discrete distribution satisfying the condition that for some interval $(\delta_1, \delta_2]$ we have $\Pr[\delta_1 < X \leq \delta_2] = 0$, where $0 \leq \delta_1 < \delta_2 < \infty$. Puri and Rubin (1970) proved that, if X_1 and X_2

independently have the same distribution as X, then X and the absolute difference $|X_1 - X_2|$ have the same distribution iff X has the pmf

$$\begin{aligned}\Pr[X = 0] &= \alpha, \\ \Pr[X = x\tau] &= 2\alpha(1-\alpha)(1-2\alpha)^{x-1}, \qquad x = 1, 2, \ldots,\end{aligned} \tag{5.60}$$

where $\tau > 0$ and either $\alpha = 1$ or $0 < \alpha \leq 0.5$.

Puri's (1973) characterization theorem concerning the sum of two independent nonnegative discrete rv's X_1 and X_2 states that

$$\Pr[X_1 \leq x] - \Pr[X_1 + X_2 \leq x] = a \Pr[X_1 + X_2 = x], \qquad x = 0, 1, 2, \ldots, \tag{5.61}$$

where $a > 0$, iff

$$\Pr[X_2 = x] = \left(\frac{1}{1+a}\right)\left(\frac{a}{1+a}\right)^x, \qquad x = 0, 1, 2, \ldots, \tag{5.62}$$

that is, iff X_2 has a geometric distribution.

The analog of Puri and Rubin's (1972) characterization of mixtures of exponential rv's supposes that $(X_1, X_2, \ldots, X_n)$ is a vector of n nonnegative discrete rv's such that

$$\Pr[X_i > x_i | i = 1, 2, \ldots, n] = \beta \Pr[X_i = x_i | i = 1, 2, \ldots, n], \tag{5.63}$$

where $\beta > 0$ and $x_i = 0, 1, 2, \ldots$. The only distributions of $X_1, X_2, \ldots, X_n$ satisfying this condition are mixtures of geometric distributions.

Uppuluri, Feder, and Shenton (1967) obtained a characterization based on the observation that the geometric distribution is obtained as the limiting form of the sequence $\{Y_n\}$ defined by the stochastic model

$$Y_n = 1 + V_n Y_{n-1}, \tag{5.64}$$

where V_i are iid with $\Pr[V_i = 0] = a$ and $\Pr[V_i = 1] = 1 - a$, with $0 < a < 1$; see also Paulson and Uppuluri (1972).

Lukacs (1965) gave characterizations based on moment properties. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with finite variance and put $S_j = \sum_{i=1}^n (X_i)^j$, $j = 1, 2$. Then the distribution is geometric iff

$$T = \left(\frac{n+1}{n-1}\right) S_2 - \left(\frac{2}{n-1}\right) S_1^2 - S_1 \tag{5.65}$$

has zero regression on S_1. Lukacs also showed that $0 < \kappa_1/\kappa_2 < 1$, $\kappa_1/\kappa_2 = 1/Q$, $\kappa_2 = \kappa_1^2 + \kappa_1$ characterizes the geometric distribution with characteristic function $\phi(t) = (Q - Pe^{it})^{-1}$ and mean and variance $\mu = \kappa_1 = kP$ and $\mu_2 = \kappa_2 = kP(1+P)$.

Kagan, Linnik, and Rao (1973) showed that Patil and Seshadri's (1964) general result for discrete distributions implies the following characterization: If the conditional distribution of X_1, given $X_1 + X_2$, has a discrete uniform distribution for all values of the total $X_1 + X_2$, then X_1 and X_2 both have the same geometric distribution.

Dallas (1974) found that if X has a nondegenerate distribution on the nonnegative integers, with $\Pr[X = k] \neq 0$ for all $k = 0, 1, 2, \ldots$, then X has a geometric distribution iff $\text{Var}(X|X > c) = d$, where d is constant $[= kP(1 + P)]$, for $c = -1, 0, 1, \ldots$. This is a "nonaging" property [cf. (5.56)].

Srivastava (1979) and Ahsanullah and Holland (1984) have also given two characterizations of the geometric distribution based on properties of record value distributions. Arnold (1980) has given further characterizations in terms of order statistics. Yet more characterizations have been proved by Shaked (1974), Shanbhag (1974), Chong (1977), Nagaraja and Srivastava (1987), Nagaraja (1988a), Wesolowski (1989), and Khalil, Dimitrov, and Dion (1991). Nagaraja (1990) has included in his paper on order statistics for discrete distributions a comprehensive review of characterizations for the geometric distribution based on (1) the independence of certain functions of its order statistics and (2) its distributional properties.

5.9.2 Negative Binomial Distribution

Only a few characterizations have been obtained for the negative binomial distribution.

Patil and Seshadri's (1964) general result implies that, if the conditional distribution of X_1 given $X_1 + X_2$ is negative hypergeometric with parameters m and n for all values of the total $X_1 + X_2$, then X_1 and X_2 both have negative binomial distributions, with parameters (m, θ) and (n, θ), respectively (Kagan, Linnik, and Rao, 1973).

Meredith (1971) has made the following comment on Ottestad's (1944) derivation of the negative binomial as a mixed Poisson distribution by assuming that the regression of Θ (the mixing variable) on X is linear, without making any assumption about the form of $f(\theta)$: Because the mixing distribution of a mixed Poisson is identifiable (Feller, 1943), it follows that the regression of Θ on X is linear iff Θ has a gamma distribution and the mixed Poisson distribution is negative binomial.

Some of the other modes of genesis in Section 5.3 can similarly be made the basis of characterizations. For example, a Poisson–stopped sum of a distribution on the *positive* integers is a negative binomial iff the distribution on the positive integers is logarithmic.

See Ong (1995c) and references therein for a characterization of the negative binomial distribution via a conditional distribution and a linear regression. Sathe and Ravi (1997) have shown that, if X is a nonnegative integer-valued rv with finite mean μ and if k is a positive integer, then X has a negative binomial distribution iff

$$\frac{d\Pr[X > n]}{d\mu} = \frac{n + k}{p + k} \Pr[X = n], \qquad n = 0, 1, \ldots.$$

A characterization using the property of α-monotonicity has been obtained by Sapatinas (1999).

5.10 APPLICATIONS

The negative binomial distribution has become increasingly popular as a more flexible alternative to the Poisson distribution, especially when it is doubtful whether the strict requirements, particularly independence, for a Poisson distribution will be satisfied.

Negative binomial distributions have been found to provide useful representations in many fields. Many researchers, including Arbous and Kerrich (1951), Greenwood and Yule (1920), and Kemp (1970), have applied it to accident statistics. Furry (1937) and Kendall (1949) have shown its applicability in birth-and-death processes. It has been found useful for psychological data by Sichel (1951); as a lag distribution for time series in economics by Solow (1960); and in market research and consumer expenditure by Chatfield, Ehrenberg, and Goodhardt (1966), Chatfield (1975), and Goodhardt, Ehrenberg, and Chatfield (1984). Medical and military applications have been described by Chew (1964) and by Bennett and Birch (1964). Burrell and Cane (1982) have used negative binomial models for lending-library data.

Bliss and Fisher (1953) successfully fitted the distribution to a large number of biometrical data sets. When Martin and Katti (1965) fitted the negative binomial and certain other distributions to 35 ecological data sets, they found that the negative binomial and the Neyman type A have very wide applicability. Elliott's (1979) manual highlights its usefulness for analyzing samples of freshwater fauna. Wilson and Room (1983), Binns (1986), and Perry (1984) have used the negative binomial for modeling entomological data.

The distribution has been used to model family size by Rao et al. (1973). Janardan and Schaeffer (1981) considered it as a possible alternative to the logarithmic distribution for the number of different compounds identified in water samples. The following three very diverse applications were examined by Clark and Perry (1989): cell-centers in grid squares [see also Crick and Lawrence (1975) and Diggle (1983)], red mites on apple leaves [see also Bliss and Fisher (1953)], and counts of cycles to failure of worsted yarn [see also Box and Cox (1964)].

Boswell and Patil (1970) gave a useful account of some dozen processes leading to the negative binomial distribution; these and others are given in Section 5.3. Physical applications involving queueing theory and other stochastic processes were described by Bhat (2002) and by Taylor and Karlin (1998). Autoregressive moving-average processes with geometric and negative binomial marginal distributions were discussed by McKenzie (1986).

More recent applications of the negative binomial distribution include the estimation of unobserved lost sales in retail inventory management (Agrawal and Smith, 1996). Al-Saleh and Al-Batainah (2003) estimated the proportion of sterile couples by fitting a negative binomial and also a zero-truncated negative binomial distribution to data on completed families. Cooper (1998) has explained the Haviv–Puterman differential equation relating the expected infinite-horizon λ-discounted reward to the expected total reward up to a random time determined by a negative binomial rv with parameters 2 and λ; he has also proved the more general case where the parameters are k and λ. Thurston, Wand, and Wiencke (2000) have extended the generalized additive model to handle negative binomial responses and have applied their methodology to data involving DNA adduct counts and smoking variables among exsmokers with lung cancer.

A problem can arise, however. If the negative binomial is found empirically to give a good fit for a particular kind of data, then the experimenter may still have to decide how to interpret the fit in terms of the many possible modes of genesis of the distribution. In particular, a good fit does not, on its own, distinguish between the heterogeneous (mixed) Poisson model and the Poisson–stopped sum model.

There are of course situations where a good fit is not obtainable with the negative binomial distribution, and in such cases it is usual to consider the possibility of a mixture of distributions or a contagious distribution (possibly with more than two parameters); see Chapters 8 and 9.

5.11 TRUNCATED NEGATIVE BINOMIAL DISTRIBUTIONS

In the most common form of truncation, the zeroes are not recorded, giving the pmf

$$\Pr[X = x] = (1 - Q^{-k})^{-1}\binom{k+x-1}{k-1}\left(\frac{P}{Q}\right)^x\left(1 - \frac{P}{Q}\right)^k, \qquad x = 1, 2, \ldots, \tag{5.66}$$

where $Q = 1 + P$. This situation occurs in applications of the negative binomial such as the number of offspring per family, the number of claims per claimant, and the number of occupants per car. Other examples have been given by Sampford (1955) and Brass (1958).

Boswell and Patil (1970) gave a genesis of the *zero-truncated negative binomial distribution* as a distribution for the sizes of groups (a group size distribution). They also showed that it can be derived as a mixture of zero-truncated Poisson distributions, since (amending a typo in the original)

$$\begin{aligned}
&\binom{k+x-1}{x}\left(\frac{1}{1+\theta}\right)^x\left(\frac{\theta}{1+\theta}\right)^k \bigg/ \left[1 - \left(\frac{\theta}{1+\theta}\right)^k\right] \\
&\quad = \binom{k+x-1}{x}\left(\frac{\theta^k(1+\theta)^{-x-k}}{1 - [\theta/(1+\theta)]^k}\right) \\
&\quad = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})} \times ce^{-\lambda\theta}\lambda^{k-1}(1-e^{-\lambda})\, d\lambda, \qquad x = 1, 2, \ldots,
\end{aligned} \tag{5.67}$$

where $c^{-1} = \Gamma(k)[\theta^{-k} - (1+\theta)^{-k}]$ and $\theta = 1/P$.

The moments about zero are easily calculated as

$$\mu_r' = (1 - Q^{-k})^{-1} m_r', \tag{5.68}$$

where m_r' is the corresponding uncorrected moment of the untruncated negative binomial distribution. Thus

$$\mu = E[X] = \frac{kP}{1 - Q^{-k}} \qquad \text{and} \qquad E[X^2] = \frac{kPQ + k^2P^2}{1 - Q^{-k}}, \tag{5.69}$$

whence

$$\mu_2 = \frac{kPQ + k^2P^2}{1 - Q^{-k}} - \frac{k^2P^2}{(1 - Q^{-k})^2}. \tag{5.70}$$

Rider (1962b) gave tables of $E[X^{-1}]$ and of $\mathrm{Var}(X^{-1})$ to five decimal places; Govindarajulu (1962) proposed the approximations

$$E[X^{-1}] \approx \frac{1}{kP - Q} \qquad \text{and} \qquad \mathrm{Var}(X^{-1}) \approx \frac{Q}{(kP - Q)(kP - 2Q)}. \tag{5.71}$$

The MLEs of k and Q are given by

$$\begin{aligned} 0 &= \sum_{x \geq 1} f_x \left(\frac{1}{\hat{k}} + \frac{1}{\hat{k} + 1} + \cdots + \frac{1}{\hat{k} + x - 1} - \frac{\ln \hat{Q}}{1 - \hat{Q}^{-\hat{k}}} \right), \\ 0 &= \sum_{x \geq 1} f_x \left(\frac{x}{1 - \hat{Q}} - \frac{\hat{k}}{1 - \hat{Q}^{-\hat{k}}} \right), \end{aligned} \tag{5.72}$$

where f_x is the observed frequency of the observation x; see David and Johnson (1952) who also gave expressions for the variances and covariance of the maximum-likelihood estimates.

Solution of the maximum-likelihood equations requires iteration. An early computer algorithm to solve the maximum-likelihood equations was published by Wyshak (1974).

David and Johnson (1952) found that the moment equations (obtained by equating the sample and expected values of the mean and variance) also do not have an explicit solution. Sampford (1955) proposed a trial-and-error method for solving them. However, David and Johnson found that these estimators are very inefficient and recommended using maximum-likelihood methods instead.

Brass (1958) proposed using f_1 together with $\overline{x}$ and s^2. He showed that the efficiency of his method was greater than that of Sampford's moment method when $k \leq 5$ and not much less when $k > 5$. Pichon et al. (1976) proposed the use of only $\overline{x}$ and f_1, but without investigating the properties of the method.

Schenzle (1979) made an in-depth study of the asymptotic efficiencies of these methods. He found that for small values of k (e.g., $k < 1$) and values of μ in the range $2 \le \mu \le 8$ none of them is applicable and recommended maximum-likelihood estimation instead.

Minimum-variance unbiased estimation was investigated by Cacoullos and Charalambides (1975).

Ahuja (1971b) studied the *n-fold convolution of the zero-truncated negative binomial distribution.* He observed that it is a special case of the generalized PSD with series function

$$f(\theta) = \left[(1-\theta)^{-k} - 1\right]^n,$$

where $\theta = P/Q$, and hence showed that the pmf is

$$\Pr[X = x] = \sum_{i=1}^{n} (-1)^{n-i} \binom{n}{i} \binom{x+ki-1}{x} \frac{\theta^x}{[(1-\theta)^{-k} - 1]^n} \tag{5.73}$$

for $x = n, n+1, \ldots$. He also obtained an expression for the cdf in terms of a linear combination of incomplete beta functions. Ahuja and Enneking (1974) studied the n-fold convolution of the negative binomial truncated by omission of $x = 0, 1, \ldots, c$, where c is a positive integer. Charalambides (1977b) explored the use of associated Lah numbers and generalized Lah numbers for these two convolutions. Saleh and Rahim (1972) gave a formula for the convolution of n truncated negative binomial variables as an example of their general method for investigating convolutions of truncated discrete distributions.

Hamdan (1975) obtained an expression for the correlation between the numbers of two types of children when family size is assumed to be negative binomial but the number of fertile childless families is unknown.

The *displaced negative binomial distribution* (*left-truncated negative binomial distribution*) is obtained by truncation of the first r probabilities of a negative binomial distribution. Shah (1971) showed that it has pmf

$$\Pr[X = x] = \frac{(k+x-1)!\theta^x/x!}{\sum_{x \ge r} (k+x-1)!\theta^x/x!}, \qquad x = r, r+1, \ldots, \tag{5.74}$$

and investigated maximum-likelihood and other methods of estimation; see also Cameron and Trivedi (1998, p. 118). A left-truncated geometric distribution is of course a shifted geometric distribution.

The *right-truncated negative binomial distribution* has support $0, 1, \ldots, n$ and pgf

$$G(z) = \frac{{}_2F_1[k, -n; -n; Pz/(1+P)]}{{}_2F_1[k, -n; -n; P/(1+P)]}. \tag{5.75}$$

This distribution has finite support; the condition that $0 < p = P/(1 + P) < 1$ can be relaxed to $0 < p$ (if $p > 1$, then $P < 0$). Cameron and Trivedi (1998) commented that it has a smaller mean and a smaller variance than the untruncated distribution. Gurmu and Trivedi (1992) have made a detailed analysis of the moments for both the left- and right-truncated distributions.

The distribution of the sum of t independent right-truncated negative binomial variables with differing parameters has been researched by Lingappaiah (1992). In the general case there are three sets of parameters: $q_1, q_2, \ldots, q_t$, $k_1, k_2, \ldots, k_t$, and $m_1, m_2, \ldots, m_t$, where m_i is the truncation point for the ith distribution. Closed-form expressions are given for the mean and variance. Attention is given to the special case where $q_1 = q_2 = \cdots = q_t$, $k_1 = k_2 = \cdots = k_t$, and $m_1 = m_2 = \cdots = m_t$.

5.12 RELATED DISTRIBUTIONS

5.12.1 Limiting Forms

The negative binomial distribution is a limiting form of the hypergeometric-type distributions of Section 6.2.5, with pgf

$$G(z) = \frac{{}_2F_1[-n, -a; b - n + 1; z]}{{}_2F_1[-n, -a; b - n + 1; 1]},$$

given the following conditions:

1. when $a < 0$, $b < 0$, and n is a positive integer (type IIA hypergeometric), if $n \to \infty$, $b \to -\infty$, and $n/(n - b) \to q$, $0 < q < 1$, that is, if $n \to \infty$, $b \to -\infty$, and $-n/b \to P > 0$;
2. when $a + b < n < 0 < a$ and a is a positive integer (type IIIA hypergeometric), if $a \to \infty$, $b \to -\infty$, and $-a/b \to q$, $0 < q < 1$, that is, if $a \to \infty$, $b \to -\infty$, and $-a/(a + b) \to P > 0$;
3. when $a < 0, n < 0, 0 < a + b + 1$ (type IV hypergeometric), if $a \to -\infty$, $b \to \infty$, and $-a/b \to q$, $0 < q < 1$, that is, if $a \to -\infty$, $b \to \infty$, and $-a/(a + b) \to P > 0$.

In turn, the logarithmic distribution of Chapter 7 with parameter $\alpha = q$ is the limiting form of the zero-truncated negative binomial distribution as $k \to 0$.

Relationships among the negative binomial, Poisson, and binomial distributions have been mentioned in Chapters 2, 3, and 4. It is convenient to make some comments here concerning these relationships. Each of the three distributions has a pgf of the form

$$[(1 + \gamma) - \gamma z]^{-m}.$$

For the negative binomial, $\gamma > 0$, $m > 0$, while for the binomial, $-1 < \gamma < 0$, $m < 0$ an integer. The Poisson distribution corresponds to the limiting case, where $\gamma \to 0$, $m \to \infty$, with $m\gamma = \theta$, θ fixed. That is, the Poisson distribution with parameter θ is the limiting form of the negative binomial distribution with pgf $(1-q)^k/(1-qz)^k$ as $k \to \infty$, $q \to 0$, $kq \to \theta$.

The most obvious distinction between the three distributions is in the value of the ratio of the variance to the mean. This is less than unity for the binomial, equal to unity for the Poisson, and greater than unity for the negative binomial distribution. Sampling variation may yield values of this ratio that do not accord with the underlying distribution; nevertheless, the value of the ratio of the sample variance to the sample mean is a good guide as to which of these three distributions is appropriate when fitting a given data set, *given that one of them is to be used* (see also the remarks in Section 5.8.1).

It is interesting to note that for all Poisson distributions the point (β_1, β_2) lies on the straight line $\beta_2 - \beta_1 - 3 = 0$. On the other hand, for the binomial distribution with parameters n and p we have $\beta_2 - \beta_1 - 3 = -2/n$; for the negative binomial distribution with parameters k and P we have $\beta_2 - \beta_1 - 3 = 2/k$.

5.12.2 Extended Negative Binomial Model

Closely related extensions of the negative binomial distribution that are formed via truncation of its zero frequency have been studied, often independently, under a variety of names:

- Engen's extended negative binomial distribution (Engen, 1974, 1978).
- Feller–Shreve distribution of recurrence times (Feller, 1968).
- Salvia–Bolinger distribution for reliability studies (Salvia and Bollinger, 1982).
- Beta–geometric distribution (Weinberg and Gladen, 1986).
- Mittag–Leffler distribution (Pillai, 1990; Pillai and Jayakumar, 1995).
- Sibuya and scaled Sibuya distributions (Devroye, 1993; Christoph and Schreiber, 2000).
- Cluster size distribution for the Poisson–Hougaard process (Lee and Whitmore, 1993).
- Extended truncated negative binomial distribution as a loss model (Klugman, Panjer, and Willmot, 1998).

Engen (1974, 1978) put forward his extended negative binomial distribution as a model for species frequency data. Suppose that each member of a population of individuals (elements) belongs to one and only one of a number of species (classes) $C_1, C_2, \ldots, C_s$, where s is the number of species (possibly infinite). Let R_j, $j = 1, \ldots, N$, be the number of species represented by exactly j individuals in a sample of individuals and suppose that sampling is such that

$$E[R_j] \propto \frac{w^k \Gamma(k+j)(1-w)^j}{\Gamma(k+1)j!}, \qquad j = 1, 2, \ldots, \qquad 0 < w < 1. \tag{5.76}$$

The number of species represented by no individuals is unascertainable. For a given total number of observed species the conditional distribution of species frequency therefore has the pgf

$$\begin{aligned} G(z) &= \frac{\sum_{j \geq 1} \alpha w^k (k+1) \cdots (k+j-1)(1-w)^j z^j / j!}{\sum_{j \geq 1} \alpha w^k (k+1) \cdots (k+j-1)(1-w)^j / j!} \\ &= \frac{[1-(1-w)z]^{-k} - 1}{[1-(1-w)]^{-k} - 1} \\ &= z \frac{{}_2F_1[k+1, 1; 2; (1-w)z]}{{}_2F_1[k+1, 1; 2; (1-w)]}. \end{aligned} \tag{5.77}$$

Expression (5.77) shows that it is a special case of a shifted extended generalized Waring distribution (Section 6.11.1).

Feller (1968) derived the distribution as the outcome of a random walk. It was obtained by Lee and Whitmore (1993) as the *cluster size distribution* for the Poisson–Hougaard process with the parameterization

$$k \Rightarrow -\alpha, \qquad w \Rightarrow \frac{\kappa}{\kappa + 1},$$

where $0 < \alpha < 1$, $0 < \kappa$. Willmot (1988a) commented that the distribution belongs to Sundt and Jewell's (1981) family of distributions (Section 2.3.2).

As $k \to 0$, Engen's model tends to Fisher's logarithmic series model (see Section 7.1.2); the corresponding conditional (one-parameter) pgf is $g(z) = \ln[1-(1-w)z]/\ln w$. However, the form of (5.76) led Engen (1978, p. 45) to remark that "it appears that the natural lower bound for k seems to be -1 (Engen, 1974) and not zero as claimed by Fisher et al. (1943)." Values of k in the range $-1 < k < 0$ do indeed give a valid distribution.

The constraints on the parameters are therefore $-1 < k$ and $0 < w < 1$; the support is $1, 2, \ldots$ and the pmf is

$$\Pr[X = x] = \frac{w^k \Gamma(k+j)(1-w)^j / [\Gamma(k+1)j!]}{[1-(1-w)]^{-k} - 1}. \tag{5.78}$$

The mean and variance are

$$\mu = \frac{k(1-w)}{w(1-w^k)} \qquad \text{and} \qquad \mu_2 = \frac{k(1-w)}{w^2(1-w^k)} \left[1 - \frac{k(1-w)w^k}{(1-w^k)}\right]. \tag{5.79}$$

Klugman, Panjer, and Willmot (1998), in their study of discrete models for actuarial loss data, called it the *extended truncated negative binomial distribution*; they also considered a zero-modified form of the distribution.

The limiting case as $w \to 0$ has the pgf

$$\alpha z\, {}_2F_1[1, 1-\alpha; 2; z], \qquad 0 < \alpha = -k < 1. \tag{5.80}$$

It is a particular shifted Waring distribution and hence is a beta mixed geometric distribution (see, e.g., Weinberg and Gladen, 1986). Shifted to support $0, 1, 2, \ldots,$ it was put forward as a discrete reliability distribution by Salvia and Bollinger (1982). Under the name *1-displaced Salvia–Bollinger distribution* it was studied by Wimmer and Altmann (2001) as the self-same partial-sum distribution that satisfies

$$\Pr[X = x] = \begin{cases} \dfrac{a}{x}(\Pr[X = x] + \Pr[X = x+1] + \cdots), & x = 1, 2, \ldots, \qquad (5.81) \\[2ex] \dfrac{a(1-a)\cdots(x-a-1)}{x!}, & x = 2, 3, \ldots, \qquad (5.82) \end{cases}$$

with $\Pr[X = 0] = 0$, $\Pr[X = 1] = a = \alpha$. It has the hazard function a/x and so has a decreasing failure rate; see Roy and Gupta (1999) and Gupta, Gupta, and Tripathi (1997). The probabilities are monotonically decreasing; there are no finite moments. When shifted to the origin, it becomes the special case $a = 1 - \alpha$, $c = 1$ of the Waring distribution.

The *Sibuya*(γ) *distribution* of Devroye (1993) and Christoph and Schreiber (1998) has the pgf

$$G(z) = 1 - (1-z)^{\gamma}, \qquad 0 < \gamma < 1.$$

Christoph and Schreiber (2000) have studied in depth the mathematical properties of the *scaled Sibuya*(γ) *distribution*; this has the pgf

$$G(z) = 1 - \lambda(1-z)^{\gamma}, \qquad \gamma \in (0, 1], \qquad \lambda \in (0, 1].$$

It has strictly decreasing probabilities. Also it is self-decomposable at least for $\lambda \le (1-\gamma)/(1+\gamma)$, and is infinitely divisible if $0 < \lambda \le (1-\lambda)$.

The *discrete Mittag–Leffler* and *discrete Linnik* distributions are also closely related; see Section 11.2.5.

5.12.3 Lagrangian Generalized Negative Binomial Distribution

The binomial–binomial Lagrangian distribution of Consul and Shenton (1972) is discussed as a member of the Lagrangian family in Section 7.2. The special case with $p = p'$ (and hence with $q = q'$) was introduced into the literature by Jain and Consul (1971); it is mentioned here because they called it the "*generalized* *negative binomial distribution*," and this name has continued to be used. When $p = p'$ (and $q = q'$) the pmf is

$$\Pr[X = x] = \frac{n}{n + mx}\binom{n + mx}{x} p^x q^{n+mx-x} \tag{5.83}$$

where $x = 0, 1, \ldots$ and the restrictions on the parameters are $0 < p < 1$, $n > 0$, $p < mp < 1$. The first three moments are

$$\begin{aligned}
\mu &= np(1 - mp)^{-1},\\
\mu_2 &= npq(1 - mp)^{-3}, \qquad (5.84)\\
\mu_3 &= \mu_2\left[\frac{3mpq}{(1 - mp)^2} + \frac{4q - 1}{1 - mp}\right] - \frac{2npq^2}{(1 - mp)^4}.
\end{aligned}$$

The distribution has received much attention; see, for instance, Gupta (1974) and Kumar and Consul (1979). Normal and inverse-Gaussian limiting forms were examined by Consul and Shenton (1973). Characterization theorems appear in Jain and Consul (1971), Consul (1974), and Consul and Gupta (1980). The distribution belongs to Charalambides's (1986a) family of Gould distributions. The negative moments have been investigated by Kumar and Consul (1979). Jain and Consul (1971) fitted the distribution to data using estimation by the method of moments, while minimum-variance unbiased estimation has been investigated by Kumar and Consul (1980) and Consul and Famoye (1989). A stochastic urn model for the distribution was devised by Famoye and Consul (1989). The difference of two such distributions is the subject of Consul (1989).

Famoye (1997) has studied and compared four methods for estimating the parameters: (a) maximum likelihood, (b) first two moments and proportion of zeros, (c) first two moments and ratio of the first two frequencies, and (d) minimum chi-square estimation. He derived their asymptotic relative efficiencies. He has found that method (d) is more efficient than (b) or (c). His simulation results indicate that method (b) is preferable concerning the bias and variance of the estimators.

5.12.4 Weighted Negative Binomial Distributions

The theory of *weighted negative binomial distributions* is analogous to that for weighted binomial distributions (see Section 3.12.4).

An interesting variant is that of Bissell (1972a,b); see also Scheaffer and Leavenworth (1976). These authors studied estimation for the negative binomial distribution when the sampling units vary in size; they were interested in counts of flaws in strips of cloth where the strips are of varying length. This required a form of weighting according to the size of the sampling unit. Bissell showed that the way that the weighting factor enters into the pmf depends on the appropriate negative binomial model; it is not the same for the gamma mixture of Poisson models as for the model corresponding to a Poisson–stopped sum of logarithmic variables.

5.12.5 Convolutions Involving Negative Binomial Variates

Convolutions of distributions were defined in Section 1.2.11. The convolution of a Poisson with a negative binomial distribution (the *Poisson * negative binomial distribution*) appeared in a little-known paper by Lüders (1934) as his *Formel II distribution*.

We consider first Lüders' Formel I distribution. He obtained this as a convolution of variables that are Poisson singlets, doublets, triplets, and so on, with parameters

$$\lambda_1 = \frac{\beta}{1+\gamma}, \qquad \lambda_2 = \frac{\beta\gamma}{2(1+\gamma)^2},$$
$$\lambda_3 = \frac{\beta\gamma^2}{3(1+\gamma)^3}, \qquad \ldots \qquad \lambda_r = \frac{\beta\gamma^{r-1}}{r(1+\gamma)^r}, \qquad \ldots,$$

respectively. The pgf for this convolution is then

$$G(z) = \exp\left[\frac{\beta}{\gamma}\sum_{r\geq 1}\left(\frac{\gamma}{1+\gamma}\right)^r\left(\frac{z^r-1}{r}\right)\right] = (1+\gamma-\gamma z)^{-\beta/\gamma}. \qquad (5.85)$$

Lüders' Formel I distribution is therefore a negative binomial distribution.

Lüders' Formel II distribution is the outcome when the Poisson singlet, doublet, triplet, ... parameters are

$$\lambda_1 = \lambda, \qquad \lambda_2 = \frac{b}{2}, \qquad \lambda_3 = \frac{bq}{3}, \qquad \ldots \qquad \lambda_r = \frac{bq^{r-2}}{r}, \qquad \ldots,$$

respectively. This is a three-parameter distribution. Lüders derived a general formula for the probabilities and expressions for the first three moments; he used the latter to fit the distribution by the method of moments to the well-known hemocytometer data of "Student" (1907).

The pgf can be shown to be

$$\begin{aligned} G(z) &= \exp\left[\left(\lambda - \frac{b}{q}\right)(z-1) + \frac{b}{q^2}\sum_{r\geq 1}\frac{q^r(z^r-1)}{r}\right] \\ &= e^{(\lambda-b/q)(z-1)}(1-q)^{b/q^2}(1-qz)^{-b/q^2}, \end{aligned} \qquad (5.86)$$

that is, the convolution of a Poisson pgf with a negative binomial pgf. Using the convolution property, the pmf is found to be

$$\Pr[X = x] = \frac{e^{-\theta}\theta^x(1-q)^k}{(k-1)!x!}\sum_{j=0}^{x}\binom{x}{j}\left(\frac{q}{\theta}\right)^j, \qquad x = 0, 1, \ldots, \qquad (5.87)$$

where $\theta = \lambda - b/q$ and $k = b/q^2$. The cumulants are the sums of Poisson and negative binomial cumulants.

We note that the *Poisson*binomial distribution* (a convolution of a Poisson and a binomial distribution) arises in stochastic process theory as the equilibrium solution for a simple immigration–death process (Cox and Miller, 1965, p. 168); this mode of genesis can be reinterpreted as the equilibrium solution for Palm's trunking problem concerning a telephone exchange with an infinite number of channels (an M/M/∞ queue) (Feller, 1957, pp. 414, 435). The properties of the distribution are analogous to those for the Poisson*negative binomial distribution. Ong (1988) has expressed the pmf in terms of Charlier polynomials and has studied models for the distribution.

The Poisson*negative binomial distribution is known in the actuarial literature as the *Delaporte distribution*, where it has arisen as a three-parameter mixture of Poisson distributions; see Section 8.3.3. It has been put forward as an alternative to the more usual assumption of a two-parameter gamma mixture (i.e., the negative binomial distribution) in the theory of insurance claims; see Delaporte (1959) as well as Willmot (1989a) and Willmot and Sundt (1989), who have provided useful bibliographies.

The mixing distribution is assumed to have the density function

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\lambda - \gamma)^{\alpha-1}e^{\beta(\gamma-\lambda)}, \qquad \alpha, \beta > 0, \qquad \lambda > \gamma \geq 0, \tag{5.88}$$

yielding a mixed Poisson distribution with pgf

$$\begin{aligned} G(z) &= \int_\gamma^\infty e^{\lambda(z-1)} f(\lambda)\, d\lambda \\ &= e^{\gamma(z-1)} \left(\frac{\beta+1}{\beta} - \frac{z}{\beta}\right)^{-\alpha}. \end{aligned} \tag{5.89}$$

Willmot (1989a) has examined its tail behavior; certain asymptotic results are given in Willmot (1989b).

Ong and Lee (1979) derived their *noncentral negative binomial distribution* as a mixture of negative binomial variables with pmf $\binom{t+x-1}{x}(1-a)^t a^x$, $x = 0, 1, \ldots,$ where T is a rv, $T = Y + v$, v is constant, and Y is a Poisson rv with parameter λ. The resultant mixture has the pmf

$$\begin{aligned} \Pr[X = x] &= \sum_{y=0}^\infty \frac{e^{-\lambda}\lambda^y}{y!} \binom{y+v+x-1}{x} (1-a)^{y+v} a^x \\ &= e^{-\lambda}(1-a)^v a^x \binom{v+x-1}{x} {}_1F_1[v+x; v; \lambda(1-a)] \end{aligned} \tag{5.90}$$

and pgf

$$G(z) = \left(\frac{1-a}{1-az}\right)^v \exp\left[\lambda\left(\frac{1-a}{1-az} - 1\right)\right]; \tag{5.91}$$

this is the pgf for the convolution of a negative binomial rv with a Pólya–Aeppli rv (see Section 9.7 for the Pólya–Aeppli distribution). The cumulants of the distribution are therefore the sums of the component Poisson and Pólya–Aeppli cumulants.

Ong and Lee (1979) showed that the distribution arises also as a mixture of Poisson distributions using a continuous mixing distribution that involves a Bessel function and is related to the noncentral chi-square distribution of Johnson et al. (1995, Chapter 28). This mode of genesis has applications to neural counting mechanisms [see McGill (1967)]. It also has applications to photon counting [see Teich and McGill (1976)]. Ong and Lee gave a useful recurrence relation for the probabilities. They also obtained a characterization and demonstrated various estimation procedures. Several minimum chi-squared procedures based on moment and probability relationships, also maximum-likelihood estimation, were investigated by Gurland, Chen, and Hernandez (1983), who independently derived the distribution using the parameters $\{\alpha, \theta, \phi\}$, where $v = 1 + \alpha$, $a = \theta$, and $\lambda = \phi/(1 - \theta)$.

Ong and Toh (2001) reexpressed the pgf (5.91) in the form

$$G(z) = \exp\{\lambda[zg(z) - 1]\}$$

where $zg(z)$ is the pgf of the cluster size distribution. They established logconvexity and infinite divisibility for the (shifted) pgf $g(z)$ and investigated its tail behavior and parameter estimation.

The Ong and Lee (1986) generalization of the noncentral negative binomial distribution is the convolution of two binomial-type distributions; see Kemp (1979) and Section 3.12.2. Ong and Lee obtained formulations via mixing processes and gave examples of fits of this distribution to data using moment estimation.

The distribution obtained by Kemp and Kemp (1986) as a model for the spread of drug abuse is a convolution of Ong and Lee's distribution (5.91) with either a binomial or a pseudobinomial distribution.

5.12.6 Pascal–Poisson Distribution

The *Pascal–Poisson distribution* was first derived as a negative binomial–stopped sum of Poisson distributions (the *negative binomial* $\bigvee$ *Poisson distribution*) by Subrahmaniam (1966) as a limiting form of the more general contagious distribution arising from a larvae survival model that is described in Section 9.9. (Stopped-sum distributions are defined in Section 9.1.)

The pgf of the Pascal–Poisson distribution is

$$\begin{aligned} G(z) &= {}_1F_0\left[a; ; \frac{\mu}{a\phi}\{\exp[\phi(z-1)] - 1\}\right] \\ &= \left(1 + \frac{\mu}{a\phi} - \frac{\mu}{a\phi}e^{\phi(z-1)}\right)^{-a}, \end{aligned} \tag{5.92}$$

where a, ϕ, and μ are all positive. The form of the pgf (5.92) shows, by Gurland's theorem (Section 8.3.2), that the Pascal–Poisson distribution is both a negative binomial–stopped sum of Poisson variables and a Poisson mixture of negative binomial distributions. It is a three-parameter distribution; it was reexamined in a paper by Kathleen Subrahmaniam (1978).

Kocherlakota Subrahmaniam (1966) gave a recursion formula for the probabilities and showed that the mean and variance are

$$\mu \quad \text{and} \quad \mu_2 = \mu\left(1 + \phi + \frac{\mu}{a}\right). \tag{5.93}$$

He fitted the distribution to Beall and Rescia's (1953) data sets by an ad hoc method with $a = 1, 2$.

Subrahmaniam (1978) set $\phi = c$ and obtained the following formulas for the probabilities:

$$\Pr[X = 0] = \left(1 - \frac{\mu}{ac}[\exp(-c) - 1]\right)^{-a},$$

$$\Pr[X = x] = \left(\frac{ac}{ac + \mu}\right)^{a} \frac{c^x}{\Gamma(a)\Gamma(x+1)} \sum_{j=1}^{\infty} \frac{\Gamma(a+j)}{\Gamma(j+1)} \left(\frac{\mu}{ac+\mu}\right)^{j} j^x e^{-jc} \tag{5.94}$$

for $x = 1, 2, \ldots$. She concentrated on estimation for the distribution assuming that a is known and gave moment and maximum-likelihood estimators for μ and c; she derived expressions for their variances. The relative efficiency of moment estimation was found to decrease as μ increases (a and c fixed) and to decrease as c increases (μ and a fixed). Maximum-likelihood estimation proved much superior. The Pascal–Poisson distribution gave a somewhat better fit than a negative binomial to data on molecular evolutionary events in a study on amino acids.

5.12.7 Minimum (Riff–Shuffle) and Maximum Negative Binomial Distributions

Consider two packs of cards, A and B, each containing m cards. At each trial a card is taken independently with probability p from pack A and probability $q = 1 - p$ from pack B. Let X be the rv that represents the number of cards chosen from one pack when all the m cards of the other pack have been chosen. That is, trials continue until one of the two packs is exhausted; the number of cards remaining in the other pack is then $m - X$.

Uppuluri and Blot (1970) [see also Lingappaiah (1987)] showed that the appropriate pmf is

$$\Pr[X = x] = \binom{m + x - 1}{x}(p^m q^x + q^m p^x), \qquad x = 0, 1, \ldots, m. \tag{5.95}$$

(We have $\sum_{x=0}^{m} \Pr[X = x] = 1$, since

$$\sum_{x=0}^{m} \binom{m+x-1}{x} p^m q^x = I_p(m, m),$$

where $I_p(m, m)$ is an incomplete beta function and $I_p(m, m) + I_q(m, m) = 1$.) The distribution can be regarded as a mixture of two tail-truncated negative binomial distributions with parameters (m, p) and (m, q), where $p + q = 1$.

From Uppuluri and Blot's graphical analysis of the distribution it would appear to be unimodal. It is very skewed with a long tail to the left when $p = 0.5$, is nearly symmetrical when $0.3 < p < 0.4$ and when $0.6 < p < 0.7$, and becomes very skewed with a long tail to the right as p approaches 0 or 1 (the properties of the distribution are symmetric in p and $q = 1 - p$).

Uppuluri and Blot gave expressions for the mgf and for the mean and variance in terms of incomplete beta functions; they found moreover that, as m becomes large,

$$\mu \approx \begin{cases} m - 2\sqrt{\dfrac{m}{\pi}} & \text{when } p = 0.5 = q, \\ \dfrac{mq}{p} & \text{when } p \neq 0.5, \end{cases} \tag{5.96}$$

$$\mu_2 \approx \begin{cases} 2m - \dfrac{4m}{\pi} - 2\sqrt{\dfrac{m}{\pi}} & \text{when } p = 0.5 = q, \\ \dfrac{mq}{p^2} & \text{when } p \neq 0.5. \end{cases} \tag{5.97}$$

They showed also that as m becomes large $(X - mq/p)/(mq/p^2)^{1/2}$ tends to normality, and they considered applications to baseball series, the Banach matchbox problem, and the genetic code problem.

Lingappaiah (1987) investigated Bayesian estimation of the parameter p.

Zhang, Burtness, and Zelterman (2000) gave the name *minimum negative binomial distribution* to the riff–shuffle distribution because the variable can be regarded as the smallest number of Bernoulli trials needed to obtain a fixed number either of successes or of failures.

For their *maximum negative binomial distribution* the variable is the smallest number of Bernoulli trials needed to obtain at least c successes and c failures. They showed that the pmf is

$$\Pr[X = x] = \binom{2c+x-1}{c-1}(p^x + q^x)(pq)^c, \qquad x = 0, 1, \ldots, \tag{5.98}$$

where $p + q = 1$. Their graphs of the pmf show that the pmf can have five distinct shapes. They obtained an expression for the mgf and hence for the mean and variance; their approximate values for large values of c are

$$\mu \approx \frac{c(q-p)}{p} \qquad \text{and} \qquad \mu_2 \approx \frac{cq}{p^2}. \tag{5.99}$$

For large values of c and $p = 0.5$ the rv $Y = (2c)^{-1/2}X$ has approximately a half-normal distribution. Zhang, Burtness, and Zelterman also described how to achieve maximum-likelihood estimation via the EM algorithm.

Their interest in the maximum negative binomial distribution arose from the need to infer the probability of an abnormal Ki-*ras* gene in the screen of colon cancer patients. They wanted to design a study to test whether or not p is close to 0 or 1 (in which case they thought that screening would not be worthwhile).

5.12.8 Condensed Negative Binomial Distributions

The Erlang process is a generalization of the Poisson process in which the time between events has an Erlang distribution (a gamma distribution with an integer exponent parameter k). It gives rise to the *asynchronous counting distribution* with pmf

$$\begin{aligned}
\Pr[X=0] &= \sum_{j=0}^{k-1}\left(1-\frac{j}{k}\right)\frac{e^{-\lambda t}(\lambda t)^j}{j!},\\
\Pr[X=x] &= \sum_{j=1-k}^{k-1}\left(1-\frac{|j|}{k}\right)\frac{e^{-\lambda t}(\lambda t)^{xk+j}}{(xk+j)!}, \qquad x=1,2,\ldots,
\end{aligned} \tag{5.100}$$

(Haight, 1967). From data that they had seen, Chatfield and Goodhardt (1973) and Murthi, Srinivasan, and Tadikamalla (1993) considered that the Erlang distribution with $k = 2$ and pdf

$$f(t) = \lambda^2 te^{-\lambda t}, \qquad 0 < \lambda, \qquad 0 < t < \infty,$$

[with mode $= (\lambda t)^{-1}$, mean $= 2(\lambda t)^{-1}$, and variance $= 2(\lambda t)^{-2}$] is a more realistic interpurchase time distribution than an exponential one [with mode zero and mean $(\lambda t)^{-1}$ and variance $= (\lambda t)^{-2}$]. The resultant distribution for number of purchases, X, has the pmf

$$\begin{aligned}
\Pr[X=0] &= e^{-\lambda t}\left(1+\frac{\lambda t}{2}\right),\\
\Pr[X=x] &= e^{-\lambda t}\left(\frac{(\lambda t)^{2x-1}}{2(2x-1)!}+\frac{(\lambda t)^{2x}}{(2x)!}+\frac{(\lambda t)^{2x+1}}{2(2x+1)!}\right), \qquad x=1,2,\ldots.
\end{aligned} \tag{5.101}$$

The mean and variance are

$$\mu = \frac{\lambda t}{2} \qquad \text{and} \qquad \mu_2 = \frac{\lambda t + e^{-\lambda t}\sinh \lambda t}{4}.$$

It is the asynchronous counting distribution with $k = 2$; in the marketing context it was called the *condensed Poisson distribution* because $\mu_2 < \mu$ for all λt.

Chatfield and Goodhardt and Murthi et al. recognized that, while different purchasers could be assumed to have the same interpurchase behavior, their purchasing intensities could differ. Their final model for X is a gamma mixture of condensed Poisson variables; they called this a *condensed negative binomial distribution*. It has the pmf

$$\begin{aligned} \Pr[X = 0] &= a_0 + \tfrac{1}{2}a_1, \\ \Pr[X = x] &= \tfrac{1}{2}a_{2x-1} + a_{2x} + \tfrac{1}{2}a_{2x+1}, \qquad x = 1, 2, \ldots, \end{aligned} \tag{5.102}$$

where

$$a_x = \frac{(k + x - 1)!P^x}{x!(k-1)!(1+P)^{k+x}}.$$

The mean and variance are

$$\mu = kP \qquad \text{and} \qquad \mu_2 = kP(P + 0.5) + \tfrac{1}{8}[1 - (1 + 4P)^{-k}]. \tag{5.103}$$

Chatfield and Goodhardt considered from their data on purchase incidence that the condensed negative binomial distribution was very little better as a model than the usual negative binomial distribution.

Murthi et al. carried out a large simulation study in which interpurchase times were assumed to be (a) exponential and (b) Erlang-2 and the heterogeneity of individuals was assumed to be (i) gamma, (ii) inverse Gaussian, (iii) lognormal, and (iv) Weibull. Seven widely differing coefficients of variation were chosen and the 56 data sets were fitted using the negative binomial, the condensed negative binomial and two other discrete distributions. The negative binomial proved robust for exponential interpurchase times. For Erlang-2 interpurchase times the condensed negative binomial was best, whatever the mixing distribution.

5.12.9 Other Related Distributions

The *hyper-negative binomial distribution* of Yousry and Srivastava (1987) belongs to the extended Katz family (see Section 2.3.1). The ratio of successive probabilities is

$$\frac{\Pr[X = x + 1]}{\Pr[X = x]} = \frac{(r + x)q}{(\theta + x)}, \qquad 0 < q < 1, \qquad r > 0, \qquad \theta > 0,$$

and $x = 0, 1, \ldots,$ giving

$$\Pr[X = x] = \frac{(r + x + 1)!(\theta - 1)!q^x}{(r-1)!(\theta + x - 1)!} \Bigg/ \sum_{x=0}^{\infty} \frac{(r + x + 1)!(\theta - 1)!q^x}{(r-1)!(\theta + x - 1)!}. \tag{5.104}$$

The pgf is

$$G(z) = \frac{{}_2F_1[1, r; \theta; qz]}{{}_2F_1[1, r; \theta; q]},$$

showing that this is an extended negative hypergeometric distribution.

If θ is a positive integer and $r > \theta - 1$, then the distribution is a left-truncated negative binomial distribution with pgf $(1-q)^k/(1-qz)^k$, where $k = r + 1 - \theta$ and the first θ probabilities, $p_0, p_1, \ldots, p_{\theta-1}$, have been truncated. Thus the distribution is related to the negative binomial in the same way that the hyper-Poisson is related to the Poisson distribution.

Yousry and Srivastava showed that the mean is

$$\mu = \frac{rq}{1-q} + \left(\frac{1-\theta}{1-q}\right)(1 - \Pr[X=0])$$

and gave a recurrence formula for the higher moments. They also gave maximum-likelihood equations. Their plots of the pmf show that the distribution is very flexible.

A *negative binomial hurdle model* with pgf

$$G(z) = \beta + (1-\beta)\left[\frac{(1+P-Pz)^{-k} - (1+P)^{-k}}{1-(1+P)^{-k}}\right], \tag{5.105}$$

$0 < \beta < 1$, $P > 0$, $k > 0$, and support $x = 0, 1, \ldots$ was applied by Pohlmeier and Ulrich (1995) to data on the demand for specialist health care in Germany. These researchers considered that demand is a two-stage process, with the hurdle (having probability of success $1 - \beta$) created by an initial visit to a general practitioner.

Mathematically this distribution is a zero-modified negative binomial distribution since (5.105) can be restated as

$$G(x) = \left[\frac{\beta - g(0)}{1 - g(0)}\right] + \left[\frac{1-\beta}{1-g(0)}\right] g(z),$$

where $g(z) = (1 + P - Pz)^{-k}$.

Bebbington and Lai's (1998) *generalized negative binomial distribution* is a waiting-time distribution given, not iid Bernoulli trials, but instead a Markov–Bernoulli sequence of successes and failures with

$$\Pr(SF) = \alpha, \qquad \Pr(FS) = \beta, \qquad \Pr(SS) = 1-\alpha, \qquad \Pr(FF) = 1-\beta.$$

They showed that the number of trials needed to obtain the first failure after a run of successes has the pgf

$$\begin{aligned} G(z) &= \frac{z(1-b-dz)}{1-cz} \\ &= z\left[(1-b) + b\left(\frac{az}{1-(1-a)z}\right)\right] \quad \text{where } d = 1-a-b,\ c = 1-a. \end{aligned} \tag{5.106}$$

This is a zero-inflated geometric distribution shifted to support $1, 2, \ldots$. The mean and variance are

$$\mu = \frac{a+b}{a} \qquad \text{and} \qquad \mu_2 = \frac{b(2-a-b)}{a^2}. \tag{5.107}$$

The pgf for the number of trials needed to obtain the kth failure is

$$G(z) = z^k \left[(1-b) + b\left(\frac{az}{1-(1-a)z}\right)\right]^k. \tag{5.108}$$

The mean and variance are now

$$\mu = \frac{k(a+b)}{a} \qquad \text{and} \qquad \mu_2 = \frac{kb(2-a-b)}{a^2}. \tag{5.109}$$

Bebbington and Lai discussed the asymptotic behavior of the distribution and gave two interesting applications. The first is a discrete-time queue, illustrated by data on years between eruptions of Mt. Sangay and by data on numbers of accesses between computer disk failures. The second application is to statistical quality control.

Jones and Lai (2002) proved that, if $(X_{n+1}|X_n)$ has a negative binomial distribution with pmf

$$\Pr[(X_{n+1}|X_n) = j] = \binom{X_n + j}{j} p^{X_n+1}(1-p)^j$$

and X_n has a geometric distribution with pmf

$$\Pr[X_n = i] = \pi(1-\pi)^i,$$

then X_{n+1} has a geometric distribution with pmf

$$\begin{aligned}
\Pr[X_{n+1} = j] &= \sum_{i=0}^{\infty} \binom{i+j}{j} p^{i+1}(1-p)^j \pi(1-\pi)^i \\
&= \frac{p(1-p)^j \pi}{[1-p(1-\pi)]^{j+1}} \\
&= \left(\frac{p\pi}{1-p+p\pi}\right)\left(\frac{1-p}{1-p+p\pi}\right)^j
\end{aligned}$$

[they say that $(X_{n+1}|X_n)$ has the parameters $(X_n+1,\ p)$ and that X_n has the parameter π; note also that summation is over i, not j]. This is their *chain negative binomial model*; it gives rise to a chain of geometric distributions with

$$G_n(z) = \frac{\theta_n}{1-(1-\theta_n)z}, \qquad n = 1, 2, \ldots, \tag{5.110}$$

where $\theta_1 = p$, $\theta_2 = p^2/(q + p^2)$, and if $\theta_n = \pi$, then $\theta_{n+1} = \pi p/(q + \pi p)$, where $q = 1 - p$. The mean and variance for X_n are

$$\mu_{n+1} = \frac{1-p}{p}(\mu_n + 1), \qquad \sigma^2_{n+1} = \frac{1-p}{p^2}(\mu_n + 1) + \frac{(1-p)^2}{p^2}\sigma_n.$$

For the stationary distribution

$$\pi = \frac{\pi p}{q + \pi p}, \quad \text{i.e., } \pi = 2 - \frac{1}{p}.$$

Jones and Lai provided an application of the model to a small business that in each time interval makes a profit with probability p and a loss with probability $1 - p$; losses can be carried forward to the next time interval according to a stated rule.

The *strict arcsine distribution* has been put forward as a competitor to the negative binomial and Poisson–inverse Gaussian distributions by Kokonendji and Khoudar (2004). It has the pmf

$$\Pr[X = x] = \frac{A(x;\alpha)}{x!}\, p^x \exp[-\alpha \arcsin(p)], \qquad x = 0, 1, 2, \ldots, \qquad (5.111)$$

where $0 < \alpha$, $0 < p < 1$, and

$$A(x;\alpha) = \begin{cases} \prod_{k=0}^{z-1}(\alpha^2 + 4k^2) & \text{if } x = 2z \text{ and } A(0;\alpha) = 1, \\ \alpha \prod_{k=0}^{z-1}(\alpha^2 + (2k+1)^2) & \text{if } x = 2z + 1 \text{ and } A(1;\alpha) = \alpha. \end{cases}$$

It is a two-parameter distribution that belongs to the Tweedie–Poisson family (Section 11.1.2). The first four cumulants are

$$\begin{aligned} \kappa_1 &= \alpha p(1 - p^2)^{-1/2}, \\ \kappa_2 &= \alpha p(1 - p^2)^{-3/2}, \\ \kappa_3 &= \alpha p(1 - p^2)^{-5/2}(1 + 2p^2), \\ \kappa_4 &= \alpha p(1 - p^2)^{-7/2}(1 + 10p^2 + 4p^4). \end{aligned}$$

The index of dispersion is therefore $(1 - p^2)^{-1} > 1$, indicating overdispersion. The distribution can also be shown to be skewed to the right and leptokurtic. The authors state that the distribution can be derived as a Poisson mixture but give no formula for this. The paper is illustrated using data on automobile claims; fitting is via the mean and the ratio of the first two observed frequencies.

CHAPTER 6

Hypergeometric Distributions

6.1 DEFINITION

The *classical hypergeometric distribution* is the distribution of the number of white balls in a sample of n balls drawn without replacement from a population of N balls, Np of which are white and $N - Np$ are black. The probability mass function (pmf) is

$$\Pr[X = x] = \binom{Np}{x}\binom{N - Np}{n - x} \bigg/ \binom{N}{n} \tag{6.1}$$

$$= \binom{n}{x}\binom{N - n}{Np - x} \bigg/ \binom{N}{Np}, \tag{6.2}$$

where $n \in \mathbb{Z}^+$, $N \in \mathbb{Z}^+$, $0 < p < 1$, and $\max(0, n - N + Np) \le x \le \min(n, Np)$. The probability generating function (pgf) is

$$G(z) = \frac{{}_2F_1[-n, -Np; N - Np - n + 1; z]}{{}_2F_1[-n, -Np; N - Np - n + 1; 1]}, \tag{6.3}$$

where

$${}_2F_1[\alpha, \beta; \gamma; z] = 1 + \frac{\alpha\beta}{\gamma} \cdot \frac{z}{1!} + \frac{\alpha(\alpha + 1)\beta(\beta + 1)}{\gamma(\gamma + 1)} \cdot \frac{z^2}{2!} + \cdots$$

is a Gaussian hypergeometric series (Section 1.1.6).

The parameters in

$$G(z) = \frac{{}_2F_1[a, b; c; z]}{{}_2F_1[a, b; c; 1]}$$

need not be restricted to $a = -n$, $b = -Np$, $c = N - Np - n + 1$. Eggenberger and Pólya (1923, 1928) studied a more general urn model leading to the pmf

$$\Pr[X = x] = \binom{n}{x}\binom{-n - (w + b)/c}{-x - w/c} \bigg/ \binom{-(w + b)/c}{-w/c}$$

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

and pgf

$$G(z) = \frac{{}_2F_1[-n, w/c; -n+1-b/c; z]}{{}_2F_1[-n, w/c; -n+1-b/c; 1]}, \tag{6.4}$$

where $w > 0$, $b > 0$, and n is a positive integer. Here it is possible for c to be either negative or positive. Taking $w/c = -Np$ leads to the classical hypergeometric distribution.

The case $w/c > 0$ arises from several models; the names *negative (inverse) hypergeometric distribution*, *hypergeometric waiting-time distribution*, and *beta–binomial distribution* all refer to the same mathematical distribution, as shown in Section 6.2.2.

The distributions mentioned so far have finite support. A distribution with infinite support, $x = 0, 1, 2, \ldots$, and pgf

$$G(z) = \frac{{}_2F_1[k, \ell; k+\ell+m; z]}{{}_2F_1[k, \ell; k+\ell+m; 1]}, \qquad k > 0, \qquad \ell > 0, \qquad m > 0, \tag{6.5}$$

arises as a beta mixture of negative binomial distributions. It is known as the *beta–negative binomial distribution* (*beta–Pascal distribution*) and also as the *generalized Waring distribution*; see Section 6.2.3.

The classical hypergeometric distribution, like the negative (inverse) hypergeometric (i.e., the beta–binomial) and the beta–negative binomial distributions, is a member of Ord's (1967a) difference-equation system of discrete distributions (see Sections 6.4 and 2.3.3). However, none of the three hypergeometric-type distributions is a power series distribution (PSD) (Section 2.2).

6.2 HISTORICAL REMARKS AND GENESIS

6.2.1 Classical Hypergeometric Distribution

The urn sampling problem giving rise to the classical hypergeometric distribution (Section 6.1) was first solved by De Moivre (1711, p. 236), when considering a generalization of a problem posed by Huygens. A multivariate version of the problem was solved by Simpson in 1740 [see Todhunter (1865, p. 206)], but little attention was given to the univariate distribution until Cournot (1843, pp. 43, 68, 69) applied it to matters concerning conscription, absent parliamentary representatives, and the selection of deputations and juries.

The reexpression of (6.1) as (6.2) shows that the distribution is unaltered when n and Np are interchanged. Note that the pmf of a hypergeometric-type distribution can always be manipulated into the form $\binom{a}{b}\binom{c}{d}\Big/\binom{a+c}{b+d}$.

The pgf (6.3) for the classical hypergeometric distribution can be restated as

$$G(z) = {}_2F_1[-n, -Np; -N; 1-z] \tag{6.6}$$

using a result from the theory of terminating Gaussian hypergeometric series. This shows that the distribution is both GHPD and GHFD; see Sections 2.4.1 and 2.4.2.

The characteristic function is

$$\frac{{}_2F_1[-n, -Np; N - Np - n + 1; e^{it}]}{{}_2F_1[-n, -Np; N - Np - n + 1; 1]}, \tag{6.7}$$

and the mean and variance are

$$\mu = E[X] = np \qquad \text{and} \qquad \mu_2 = \frac{np(1-p)(N-n)}{N-1}; \tag{6.8}$$

further moment formulas are given in Section 6.3.

The properties of the distribution were investigated in depth by Pearson (1895, 1899, 1924), who was interested in developing the system of continuous distributions that now bears his name via limiting forms of discrete distributions. Important further properties of the classical hypergeometric distribution were obtained by Romanovsky (1925).

6.2.2 Beta–Binomial Distribution, Negative (Inverse) Hypergeometric Distribution: Hypergeometric Waiting-Time Distribution

This distribution arises from a number of different models. We consider first the most widely used model.

The *beta–binomial* model gives the distribution as a mixture of binomial distributions, with the binomial parameter p having a beta distribution:

$$\Pr[X = x] = \int_0^1 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \times \frac{p^{\alpha-1}(1-p)^{\beta-1}\,dp}{B(\alpha, \beta)} \tag{6.9}$$

$$\begin{aligned} &= \binom{n}{x}\binom{-\alpha-\beta-n}{-\alpha-x}\Big/\binom{-\alpha-\beta}{-\alpha} \\ &= \binom{-\alpha}{x}\binom{-\beta}{n-x}\Big/\binom{-\alpha-\beta}{n}, \qquad n \in \mathbb{Z}^+, \\ &\qquad 0 \le \alpha, \qquad 0 \le \beta, \end{aligned} \tag{6.10}$$

where $x = 0, 1, \ldots, n$. The pgf is

$$\begin{aligned} G(z) &= \frac{{}_2F_1[-n, \alpha; -\beta - n + 1; z]}{{}_2F_1[-n, \alpha; -\beta - n + 1; 1]} \\ &= {}_2F_1[-n, \alpha; \alpha + \beta; 1 - z] \end{aligned} \tag{6.11}$$

(the distribution is therefore both GHPD and GHFD). The mean and variance are

$$\mu = \frac{n\alpha}{\alpha + \beta} \qquad \text{and} \qquad \mu_2 = \frac{n\alpha\beta(\alpha + \beta + n)}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \tag{6.12}$$

This model has been obtained and subsequently applied in many different fields by a number of research workers (see Section 6.9.2). When $\alpha = \beta = 1$, the outcome is the discrete rectangular distribution (see Section 6.10.1).

This derivation is closely related to Condorcet's *negative hypergeometric* model, which seems to have been derived for the first time by Condorcet in 1785; see Todhunter (1865, p. 383). Let A and B be two mutually exclusive events that have already occurred v and w times, respectively, in $v + w$ trials. Let $n = k + \ell$. Then the probability that in the next n trials events A and B will happen k and ℓ times, respectively (where k and ℓ are nonnegative integers), is

$$\begin{aligned}
&\frac{(k+\ell)!}{k!\ell!}\int_0^1 x^{v+k}(1-x)^{w+\ell}\,dx \Big/ \int_0^1 x^v(1-x)^w\,dx \\
&\qquad = \frac{(k+\ell)!(v+k)!(w+\ell)!(v+w+1)!}{k!\ell!(v+k+w+\ell+1)!v!w!} \\
&\qquad = \binom{n}{k}\binom{-v-w-n-2}{-v-1-k}\Big/\binom{-v-w-2}{-v-1} \\
&\qquad = \binom{-v-1}{k}\binom{-w-1}{n-k}\Big/\binom{-v-w-2}{n}, \qquad k = 0, 1, \ldots, n;
\end{aligned} \tag{6.13}$$

see also Pearson (1907). The parameters here are $n = n$, $Np = -v - 1$, and $N = -v - w - 2$. Because N is negative, the name *negative hypergeometric distribution* is used. The pgf is

$$G(z) = \frac{{}_2F_1[-n, v+1; -w-n; z]}{{}_2F_1[-n, v+1; -w-n; 1]} \tag{6.14}$$

and the mean and variance are

$$\mu = \frac{n(v+1)}{v+w+2} \qquad \text{and} \qquad \mu_2 = \frac{n(v+1)(w+1)(v+w+n+2)}{(v+w+2)^2(v+w+3)}. \tag{6.15}$$

Higher moments are discussed in Section 6.3.

Condorcet's derivation assumes that sampling takes place from an infinite population. Prevost and Lhuilier in 1799 recognized that an equivalent expression is obtained when two samples are taken in succession from a finite population, without any replacements; see Todhunter (1865, p. 454).

The derivation of the *distribution of the number of exceedances* is mathematically similar to Condorcet's result; see Gumbel and von Schelling (1950) and Sarkadi (1957b). Gumbel (1958) credits this to Thomas (1948). Consider two independent random samples of sizes m and n drawn from a population in which a measured character has a continuous distribution. The number of exceedances X is defined as the number (out of n) of the observed values in the second sample that exceed the rth largest of the m values in the first sample. The probability that in n future trials there will be x values exceeding the rth largest value in m past trials is

$$\begin{aligned}
\Pr[X = x] &= \binom{m}{r} r \binom{n}{x} (m+n)^{-1} \bigg/ \binom{m+n-1}{r+x-1} \\
&= \binom{n}{x}\binom{-m-n-1}{-r-x} \bigg/ \binom{-m-1}{-r} \\
&= \binom{-r}{x}\binom{-m+r-1}{n-x} \bigg/ \binom{-m-1}{n},
\end{aligned} \tag{6.16}$$

where $x = 0, 1, \ldots, n$. The pgf is

$$G(z) = \frac{{}_2F_1[-n, r; -n-m+r; z]}{{}_2F_1[-n, r; -n-m+r; 1]}. \tag{6.17}$$

The parameters for the exceedance model are related to the negative hypergeometric model by $r = v + 1$, $m - r = w$.

Irwin (1954) pointed out that direct sampling for a sample of fixed size n from an urn with Np white balls and $N - Np$ black balls, as in Section 6.1, *but with replacement together with an additional similarly colored ball* after each ball is drawn also gives rise to the negative hypergeometric distribution. This is a particular case of Pólya-type sampling (see Section 6.2.4).

Suppose now that sampling without replacement (as described in Section 6.1) is continued until k white balls are obtained ($0 < k \leq Np$). The distribution of the number of draws that are required is known as the *inverse hypergeometric distribution* or *hypergeometric waiting-time distribution*. (The model is analogous to the inverse binomial sampling model for the negative binomial distribution; however, the range of possible values for a negative hypergeometric distribution is finite because there is not an infinitude of black balls that might be drawn.) For this distribution

$$\begin{aligned}
\Pr[X = x] &= \binom{Np}{k-1}(Np-k+1)\binom{N-Np}{x-k}(N-x+1)^{-1} \bigg/ \binom{N}{x-1} \\
&= \binom{x-1}{x-k}\binom{N-x}{N-Np-x+k} \bigg/ \binom{N}{N-Np} \\
&= \binom{-k}{x-k}\binom{k-Np-1}{N-Np-x+k} \bigg/ \binom{-Np-1}{N-Np},
\end{aligned} \tag{6.18}$$

where $k \in \mathbb{Z}^+$, $N \in \mathbb{Z}^+$, $0 < p < 1$, and $x = k, k+1, \ldots, k+N-Np$, by manipulation of the factorials. Comparison with (6.13) shows that this is a negative hypergeometric distribution shifted k units away from the origin.

The pgf is

$$G(z) = z^k \frac{{}_2F_1[k, Np-n; k-N; z]}{{}_2F_1[k, Np-n; k-N; 1]} \tag{6.19}$$

and the mean and variance are

$$\mu = k + \frac{(N - Np)k}{Np + 1} = \frac{k(N + 1)}{Np + 1},$$
$$\mu_2 = \frac{k(N - Np)(N + 1)(Np + 1 - k)}{(Np + 1)^2(Np + 2)}. \tag{6.20}$$

The term "inverse hypergeometric distribution" can refer either to the total number of draws, as above, or to the number of unsuccessful draws, as in Kemp and Kemp (1956a) and Sarkadi (1957a). Bol'shev (1964) related the inverse sampling model to a two-dimensional random walk. Guenther (1975) has written a helpful review paper concerning the negative (inverse) hypergeometric distribution.

6.2.3 Beta–Negative Binomial Distribution: Beta–Pascal Distribution, Generalized Waring Distribution

The *beta–negative binomial distribution* was obtained analogously to the beta–binomial distribution by Kemp and Kemp (1956a), who commented that it arises both as a beta mixture of negative binomial distributions with the pgf $(1 - \lambda)^k/(1 - \lambda z)^k$ and as an F distribution mixture of the negative binomial distribution with pgf $(1 + P - Pz)^{-k}$. We have the pgf

$$G(z) = \int_0^1 \left(\frac{1-\lambda}{1-\lambda z}\right)^k \times \frac{\lambda^{\ell-1}(1-\lambda)^{m-1}\, d\lambda}{B(\ell, m)} \tag{6.21}$$
$$= \int_0^\infty (1 + P - Pz)^{-k} \times \frac{P^{\ell-1}(1+P)^{-\ell-m}\, dP}{B(\ell, m)}$$
$$= \frac{{}_2F_1[k, \ell; k + \ell + m; z]}{{}_2F_1[k, \ell; k + \ell + m; 1]}, \qquad k \geq 0, \qquad \ell \geq 0, \qquad m \geq 0. \tag{6.22}$$

The probabilities are

$$\Pr[X = x] = \binom{-k}{x}\binom{m + k - 1}{-\ell - x}\Big/\binom{m-1}{-\ell}$$
$$= \binom{-\ell}{x}\binom{\ell + m - 1}{-k - x}\Big/\binom{m-1}{-k}, \qquad x = 0, 1, \ldots, \tag{6.23}$$

and the mean and variance are

$$\mu = \frac{k\ell}{m - 1} \qquad \text{and} \qquad \mu_2 = \frac{k\ell(m + k - 1)(m + \ell - 1)}{(m - 1)^2(m - 2)}. \tag{6.24}$$

The moments exist only for $r < m$, however; see Section 6.3.

Another name that has sometimes been used for the distribution is *inverse Markov–Pólya distribution*. Kemp and Kemp (1956a) pointed out that this distribution also arises by inverse sampling from a Pólya urn *with additional replacements*; see the next section.

Unlike the classical hypergeometric and the negative hypergeometric distributions, the support of the beta–negative binomial distribution is infinite. Also, unlike those distributions, the beta–negative binomial is *not* a Kemp GHFD. It is, however, a Kemp GHPD.

The term "beta–Pascal" is often applied to a shifted form of the distribution (6.21) with support $k, k+1, \ldots$ [see, e.g., Raiffa and Schlaifer (1961, pp. 238, 270) and Dubey (1966a)]. Here k is necessarily an integer.

The unshifted distribution with pgf (6.22) and support $0, 1, \ldots$ has been studied in considerable detail by Irwin (1963, 1968, 1975a,b,c) and Xekalaki (1981, 1983a,b,c,d) under the name *generalized Waring distribution*. Irwin (1963) developed it from the following generalization of Waring's expansion (see Section 6.10.4):

$$\begin{aligned}\frac{(c-a-1)!}{(c-a+k-1)!} &= \frac{(c-1)!}{(c+k-1)!}\left[1+\frac{ak}{c+k}+\frac{a(a+1)k(k+1)}{(c+k)(c+k+1)2!}+\cdots\right]\\ &= \frac{(c-1)!}{(c+k-1)!}\,{}_2F_1[a,k;c+k;1].\end{aligned}$$

Irwin's procedure of setting $\Pr[X=x]$ proportional to the $(x+1)$th term in this series is equivalent (as he realized) to adopting the pgf

$$G(z) = \frac{{}_2F_1[a,k;c+k;z]}{{}_2F_1[a,k;c+k;1]}, \tag{6.25}$$

where $k=-n$, $a=-Np$, and $c+k=N-Np-n+1$ would give the classical hypergeometric distribution. But note that Irwin's restrictions on the parameters are $c>a>0$, $k>0$. Irwin obtained the factorial moments

$$\mu'_{[r]} = \frac{(a+r-1)!(k+r-1)!(c-a-r-1)!}{(a-1)!(k-1)!(c-a-1)!}, \tag{6.26}$$

whence, provided that they exist,

$$\mu = \frac{ak}{c-a-1} \qquad \text{and} \qquad \mu_2 = \frac{ak(c-a+k-1)(c-1)}{(c-a-1)^2(c-a-2)}. \tag{6.27}$$

Further properties of the distribution, relationships to its Pearson-type continuous analogs, tail-length behavior, and parameter estimation are the subjects of Irwin (1968, 1975a,b,c). Xekalaki (1981) has written an anthology of results concerning urn models, mixture models, conditionality models, STER (Sums successively Truncated from the Expectation of the Reciprocal) models, and some related characterizations. In Xekalaki (1983a) she studied infinite divisibility, completeness, and regression properties of the distribution and in Xekalaki (1985) she showed that the distribution can be determined uniquely from a knowledge of certain conditional distributions and some appropriately chosen regression functions. Applications of the distribution are given in Section 9.3.

Special cases of the distribution are the Yule and Waring distributions; see Sections 10.3 and 10.4.

6.2.4 Pólya Distributions

The urn models described earlier in this chapter are all particular cases of the *Pólya urn model*. This was put forward by Eggenberger and Pólya (1923) as a model for contagious distributions, that is, for situations where the occurrence of an event has an aftereffect; see also Jordan (1927) and Eggenberger and Pólya (1928).

Suppose that a finite urn initially contains w white balls and b black balls and that balls are withdrawn one at a time, with immediate replacement, together with c balls of a similar color. Then the probability that x white balls are drawn in a sample of n withdrawals is

$$\begin{aligned}\Pr[X = x] &= \binom{n}{x}\frac{w(w+c)\ldots[w+(x-1)c]b(b+c)\ldots[b+(n-x-1)c]}{(w+b)(w+b+c)\ldots[w+b+(n-1)c]}\\ &= \binom{n}{x}\frac{B(x+w/c, n-x+b/c)}{B(w/c, b/c)}\\ &= \binom{n}{x}\binom{-n-(w+b)/c}{-x-w/c}\bigg/\binom{-(w+b)/c}{-w/c}\\ &= \binom{-w/c}{x}\binom{-b/c}{n-x}\bigg/\binom{-(w+b)/c}{n}. \qquad (6.28)\end{aligned}$$

Other ways of expressing the probabilities are discussed in Bosch (1963). The pgf is

$$G(z) = \frac{{}_2F_1[-n, w/c; -n+1-b/c; z]}{{}_2F_1[-n, w/c; -n+1-b/c; 1]} \qquad (6.29)$$

and the mean and variance are

$$\mu = \frac{nw}{b+w} \qquad \text{and} \qquad \mu_2 = \frac{nwb(b+w+nc)}{(b+w)^2(b+w+c)}. \qquad (6.30)$$

Pólya (1930) pointed out the following particular cases: If c is positive, then success and failure are both contagious; if $c = 0$, then events are independent (the classical binomial situation); while if c is negative, then each withdrawal creates a reversal of fortune. When c is negative such that w/c is a negative integer (e.g., $c = -1$), the outcome is the classical hypergeometric distribution, whereas when c is positive such that w/c is a positive integer (e.g., $c = +1$), the negative hypergeometric distribution is the result. Inverse sampling for a fixed number of white balls leads to the inverse (negative) hypergeometric when w/c is a negative integer; it gives a beta–negative binomial distribution when w/c is a positive integer.

6.2.5 Hypergeometric Distributions in General

In

$$\Pr[X = x] = \binom{a}{x}\binom{b}{n-x}\bigg/\binom{a+b}{n} \tag{6.31}$$

it is clearly not essential that all the parameters n, a, b are positive; in fact, with certain restrictions, we can take any two of them negative and the remaining one positive and still obtain a valid pmf. The conditions under which (6.31) provides an honest distribution, with n, a, and b taking real values, were investigated by Davies (1933, 1934), Noack (1950), and Kemp and Kemp (1956a). Such distributions were termed generalized hypergeometric distributions by Kemp and Kemp, but the name is now used for a much wider class. *General hypergeometric distributions* have pgf's of the form

$$G(z) = \frac{{}_2F_1[-n, -a; b-n+1; z]}{{}_2F_1[-n, -a; b-n+1; 1]}; \tag{6.32}$$

they form a subset of Kemp's (1968a,b) wider class of GHPDs with pgf's of the form

$$G(z) = \frac{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda]};$$

see Section 2.4.1.

Kemp and Kemp distinguished four main types of distribution corresponding to (6.31), divided into subtypes as in Table 6.1. They adopted the following conventions:

(a) If $\Pr[X = r + 1] = 0$, then $\Pr[X = x] = 0$ for all $x \geq r + 1$.

(b) When $\alpha < 0$ and $\beta < 0$ with β an integer,

$$\frac{\alpha!}{(\alpha+\beta)!} = \frac{(-1)^{\beta}(-\alpha-\beta-1)!}{(-\alpha-1)!}.$$

They also imposed the restriction $\Pr[X = 0] \neq 0$ to ease the derivation of the moment and other properties of the distributions.

(i) The classical hypergeometric distribution belongs to type IA(i) or IA(ii) with n, a, and b all integers.

Table 6.1 Conditions for the Existence of Types I, II, III, and IV General Hypergeometric Distributions

Type	Conditions	Support
IA(i)	$n-b-1<0$; n an integer; $0 \le n-1 < a$	$x=0,1,\ldots,n$
IA(ii)	$n-b-1<0$; a an integer; $0 \le a-1 < n$	$x=0,1,\ldots,a$
IB	$n-b-1<0$; $J<a<J+1$; $J<n<J+1$	$x=0,1,\ldots$
IIA	$a<0<n$; n an integer; $b<0$; $b \neq -1$	$x=0,1,\ldots,n$
IIB	$a<0<a+b+1$; $J<n<J+1$; $J<n-b-1<J+1$	$x=0,1,\ldots$
IIIA	$n<0<a$; a an integer; $b<n-a$; $b \neq n-a-1$	$x=0,1,\ldots,a$
IIIB	$n<0<a+b+1$; $J<a<J+1$; $J<n-b-1<J+1$	$x=0,1,\ldots$
IV	$a<0$; $n<0$; $0<a+b+1$	$x=0,1,\ldots$

Note: J is a nonnegative integer (the same for any one type of distribution).

(ii) The negative (inverse) hypergeometric distribution belongs to Type IIA or IIIA with n, a, and b again all integers.

(iii) The beta–negative binomial is Type IV.

(iv) A dualism exists between types IA(i) and IA(ii), between Types IIA and IIIA, and between Types IIB and IIIB (using the substitutions $a \leftrightarrow n$ and $a+b-n \leftrightarrow b$).

(v) No meaningful models have been found for Types IB, IIB, or IIIB.

Sarkadi (1957a) extended the class of distributions corresponding to (6.31) by including the cases $b=-1, b=n-a-1$ that were excluded from types IIA and IIIA, respectively, by Kemp and Kemp. He pointed out that the sum of the probabilities over the ranges $0 \le x \le n$ and $0 \le x \le a$, respectively, is equal to unity in both cases, and so (6.31) defines a proper distribution.

By changing Kemp and Kemp's definition of $\alpha!/(\alpha+\beta)!$, Shimizu (1968) and Sibuya and Shimizu (1981) were able to include further distributions with support $[m_1, m_2]$, $[m_1, \infty)$, $[-m_2, -m_1]$, $(-\infty, -m_1]$, where m_1 and m_2 are positive integers. Their new types are distributions of the form $\pm X \pm k$, where X is a rv of one of the types in Table 6.1 and k is an integer. Table 6.2 gives the

Table 6.2 Relations between Types of Hypergeometric Distributions

Kemp and Kemp	Name	Support	Ord	Shimizu
IA(i), IA(ii)	Classical hypergeometric	Finite	I(a)	A1
IIA, IIIA	Negative (inverse) hypergeometric (beta–binomial)	Finite	I(b)	A2
IB	—	Infinite	I(e)	B1
IIB, IIIB	—	Infinite	I(e)	B2
IV	Beta–negative binomial	Infinite	VI	B3

broad relationships between Kemp and Kemp (1956a), Ord (1967a), and Shimizu (1968) hypergeometric-type distributions.

Review articles concerning general hypergeometric distributions are by Guenther (1983) and Sibuya (1983).

Kemp and Kemp's (1975) paper was concerned with models for general hypergeometric distributions. Besides urn models and models for contagion, it gave (1) models based on equilibrium stochastic processes, (2) STER models, (3) conditionality models, (4) weighting models, and (5) mixing models. It showed the following:

1. An equilibrium time-homogeneous stochastic process with birth and death rates λ_i and μ_i such that $\lambda_{i-1}/\mu_i = (a_1 + i - 1)/(b + i - 1)$ can yield a type IIA/IIIA or a type IV distribution by a suitable choice of parameters. Similarly
$$\frac{\lambda_{i-1}}{\mu_i} = \frac{(a_1 + i - 1)(a_2 + i - 1)}{(b + i - 1)i} \tag{6.33}$$
can lead to any one of type IA, IIA/IIIA, or IV by a suitable choice of parameters.
2. The STER distributions arise in connection with an inventory decision problem. If demand is a discrete rv with pgf $G(z) = \sum_{i\geq 0} p_i z^i$, then the corresponding STER distribution has probabilities that are Sums successively Truncated from the Expectation of the Reciprocal of the demand variable, giving the STER pgf
$$H(z) = (1 - z)^{-1}(1 - p_0)^{-1} \int_z^1 \frac{[G(z) - p_0]\, dz}{z}; \tag{6.34}$$
see Bissinger (1965) and Section 11.2.13. Kemp and Kemp found that, if a type IIA/IIIA demand distribution has the support $1, 2, \ldots, \min(n, a)$, it gives rise to a STER distribution that is also type IIA/IIIA; see also Kemp and Kemp (1969b).
3. Let X and Y be mutually independent discrete rv's. If X and Y are both binomial with parameters (n, p) and (m, p), then the conditional distribution of $X|(X + Y)$ is hypergeometric (type IA). If X and Y are both negative binomial with parameters (u, λ) and (v, λ), then $X|(X + Y)$ has a negative hypergeometric (type IIA/IIB) distribution (Kemp, 1968a). If
$$G_X(z) = \frac{{}_1F_1[-n; c; -\lambda z]}{{}_1F_1[-n; c; \lambda]} \qquad \text{and} \qquad G_Y(z) = \exp \lambda(z - 1),$$
then the distributions of $X|(X + Y)$ and $Y|(X + Y)$ are both type IA. Discrete Bessel distributions for X and Y can also lead to a type IA distribution for $X|(X + Y)$. Similarly, binomial distributions with parameters (n, p) and $(m, 1 - p)$ for X and Y lead to a type IA distribution for $X|(Y - X)$. Kemp and Kemp (1975) gave further models of this kind; see Kemp (1968a) for the general theory.

4. Weighting models give rise to distributions that have been modified by the method of ascertainment. When the weights (sampling chances) w_x are proportional to the value of the observation (i.e., to x), the distribution with pgf ${}_2F_1[a_1, a_2; b; z]/{}_2F_1[a_1, a_2; b; 1]$ is ascertained as the distribution with pgf

$$G(z) = z\frac{{}_2F_1[a_1+1, a_2+1; b+1; z]}{{}_2F_1[a_1+1, a_2+1; b+1; 1]}; \tag{6.35}$$

if w_x is proportional to $x!/(x-k)!$, then the same initial distribution is ascertained as the distribution with pgf

$$G(z) = z^k \frac{{}_2F_1[a_1+k, a_2+k; b+k; z]}{{}_2F_1[a_1+k, a_2+k; b+k; 1]} \tag{6.36}$$

(Kemp, 1968a).

5. Kemp and Kemp (1975) pointed out that a beta mixture of extended beta–binomial distributions can, under certain circumstances, give rise to a beta–binomial distribution. Two possibilities are as follows:

$$\begin{aligned} G_1(z) &= \int_0^1 {}_2F_1[-n, a; c; y(z-1)]\frac{y^{c-1}(1-y)^{d-c-1}\,dy}{B(c, d-c)} \\ &= {}_2F_1[-n, a; d; z-1]; \end{aligned} \tag{6.37}$$

$$\begin{aligned} G_2(z) &= \int_0^1 {}_2F_1[-n, d; b; y(z-1)]\frac{y^{c-1}(1-y)^{d-c-1}\,dy}{B(c, d-c)} \\ &= {}_2F_1[-n, c; b; z-1]. \end{aligned} \tag{6.38}$$

They also commented that a type IIA/IIIA distribution can be obtained as a gamma mixture of restricted Laplace–Haag distributions, and they pointed out that a type IV distribution can be derived as a mixture of Poisson distributions.

6.3 MOMENTS

The moment properties of the general hypergeometric distribution can be obtained from the factorial moments and indeed exist only when the factorial moments exist. The general form for the rth factorial moment (if it exists) for the distribution with pgf (6.31) is

$$\mu'_{[r]} = \frac{n!a!(a+b-r)!}{(n-r)!(a-r)!(a+b)!}, \tag{6.39}$$

and so

$${}_2F_1[-a, -n; -a-b; -t] \tag{6.40}$$

can be treated as the factorial moment generating function (fmgf) for the factorial moments.

Moments exist for the following:

Type IA(i)	Always (but are zero if $r > n$)
Type IA(ii)	Always (but are zero if $r > a$)
Type IB	When $r < a + b + 1$
Type IIA	Always (but are zero if $r > n$)
Type IIB	Never
Type IIIA	Always (but are zero if $r > a$)
Type IIIB	Never
Type IV	When $r < a + b + 1$

In other words,

$\mu'_{[r]}$ is finite for all r for types IA(i), IA(ii), IIA, and IIIA;
$\mu'_{[r]}$ is finite for $r < a + b + 1$ for types IB and IV; and
types IIB and IIIB have no moments.

Provided that the specified moment exists, it is straightforward (though tedious) to show via the factorial moments that

$$
\begin{aligned}
E[X] = \mu &= \frac{na}{a+b}, \\
\mathrm{Var}(X) = \mu_2 &= \frac{nab(a+b-n)}{(a+b)^2(a+b-1)}, \\
\mu_3 &= \frac{\mu_2(b-a)(a+b-2n)}{(a+b)(a+b-2)}, \\
\mu_4 &= \frac{\mu_2}{(a+b-2)(a+b-3)}\Bigg\{(a+b)(a+b+1-6n) \\
&\quad + 3ab(n-2) + 6n^2 + \frac{3abn(6-n)}{a+b} - \frac{18abn^2}{(a+b)^2}\Bigg\}.
\end{aligned}
\tag{6.41}
$$

The moment ratios are

$$
\sqrt{\beta_1} = \left[\frac{(a+b-1)}{abn(a+b-n)}\right]^{1/2} \frac{(b-a)(a+b-2n)}{(a+b-2)}, \tag{6.42}
$$

$$
\begin{aligned}
\beta_2 &= \frac{(a+b)^2(a+b-1)}{nab(a+b-n)(a+b-2)(a+b-3)} \\
&\quad \times \Bigg((a+b)(a+b+1-6n) + 3ab(n-2) \\
&\quad + 6n^2 + \frac{3abn(6-n)}{a+b} - \frac{18abn^2}{(a+b)^2}\Bigg) + 3.
\end{aligned}
\tag{6.43}
$$

Table 6.3 Comparison of Hypergeometric Types IA(ii), IIA, IIIA, and IV, Binomial, Poisson, and Negative Binomial Distributions

$\Pr[X = x]$

x	IA(ii)	Binomial	IIA	Poisson	IIIA	Negative Binomial	IV
0	0.076	0.107	0.137	0.135	0.123	0.162	0.197
1	0.265	0.269	0.266	0.271	0.265	0.269	0.267
2	0.348	0.302	0.270	0.271	0.284	0.247	0.220
3	0.222	0.201	0.184	0.180	0.195	0.164	0.144
4	0.075	0.088	0.093	0.090	0.093	0.089	0.083
5	0.013	0.026	0.036	0.036	0.032	0.042	0.044
6	0.001	0.006	0.011	0.012	0.007	0.017	0.023
7	0.000	0.001	0.003	0.003	0.001	0.007	0.011
8	0.000	0.000	0.000	0.001	0.000	0.002	0.005
9	—	0.000	0.000	0.000	—	0.001	0.003
10	—	0.000	0.000	0.000	—	0.000	0.001
11	—	—	—	0.000	—	0.000	0.001
$\geq$12	—	—	—	0.000	—	0.000	0.001

Note: For each distribution the mean is 2, $|n| = 10$, $|a/(a+b)| = 0.2$, and $|a+b| = 40$. A dash means that the probability is zero and 0.000 means that the probability is less than 0.0005.
Source: Adapted from Kemp and Kemp (1956a).

As $a \to \infty$, $b \to \infty$ such that $a/(a+b) = p$ (constant), the moment properties tend to those of the binomial distribution, provided that n is a positive integer. When n is negative and $a/(a+b)$ tends to λ, $\lambda < 0$, then the moment properties tend to those of the negative binomial distribution; see Table 6.3.

An alternative approach to the moment properties is via the differential equation for the mgf. The pgf is

$$G(z) = K\,{}_2F_1[-n, -a; b-n+1; z],$$

where K is a normalizing constant; this satisfies

$$\theta(\theta + b - n)G(z) = z(\theta - n)(\theta - a)G(z), \tag{6.44}$$

where θ is the differential operator $z\, d/dz$. The moment generating function (mgf) is $G(e^t)$; from the relationship between the θ and $D = d/dz$ operators (Section 1.1.4) it follows that $G(e^t)$ satisfies

$$D(D + b - n)G(e^t) = e^t(D - n)(D - a)G(e^t), \tag{6.45}$$

where D is the differential operator d/dt. The central mgf is $M(t) = e^{-\mu t}G(e^t)$, and this satisfies

$$(D + \mu)(D + \mu + b - n)M(t) = e^t(D + \mu - n)(D + \mu - a)M(t). \tag{6.46}$$

Identifying the coefficients of t^0, t^1, t^2, and t^3 in (6.46) gives expressions for the first four central moments that are equivalent to (6.41). Higher moments may be obtained similarly. This is essentially the method of Pearson (1899).

Lessing (1973) has shown that the uncorrected moments can be obtained from the following expression for the mgf:

$$G(e^t) = \frac{(a+b-n)!}{(a+b)!} \frac{\partial^n}{\partial y^n} \left[(1+ye^t)^a(1+y)^b\right]_{y=0}. \tag{6.47}$$

Janardan (1973b) commented that this result is a special case of an expression obtained by Janardan and Patil (1972).

The following finite difference relation holds among the central moments $\{\mu_j\}$:

$$(a+b)\mu_{r+1} = [(1+E)^r - E^r][\mu_2 + \alpha\mu_1 + \beta\mu_0], \tag{6.48}$$

where E is the displacement operator (i.e., $E^p[\mu_s] \equiv \mu_{s+p}$),

$$\alpha = -a + \frac{n(a-b)}{a+b}, \qquad \beta = \frac{nab(a+b-n)}{(a+b)^2},$$

and $\mu_0 = 1, \mu_1 = 0$ (Pearson, 1924).

The mean deviation is

$$\begin{aligned}\nu_1 &= E\left[\left|\frac{X-na}{a+b}\right|\right]\\ &= \frac{2m(b-n+m)}{a+b}\binom{a}{m}\binom{b}{n-m}\bigg/\binom{a+b}{n},\end{aligned} \tag{6.49}$$

where m is the greatest integer not exceeding $\mu + 1$ (Kamat, 1965).

Matuszewski (1962) and Chahine (1965) have studied the ascending factorial moments of the negative (inverse) hypergeometric distribution.

6.4 PROPERTIES

Let

$$f(x|n,a,b) = \Pr[X=x] = \binom{a}{x}\binom{b}{n-x}\bigg/\binom{a+b}{n}, \tag{6.50}$$

$$F(x|n,a,b) = \sum_j \Pr[X=j] = \sum_j \binom{a}{j}\binom{b}{n-j}\bigg/\binom{a+b}{n}, \tag{6.51}$$

where the range of summation for j is $\max(0, n-b) \le j \le x$ [or $0 \le j \le \min(x, n, a)$ when n and a are positive]. Then the following probability relationships hold:

$$f(x+1|n,a,b) = \frac{(a-x)(n-x)}{(x+1)(b-n+x+1)} f(x|n,a,b), \tag{6.52}$$

$$f(x|n, a+1, b-1) = \frac{(a+1)(b-n+x)}{(a+1-x)b} f(x|n, a, b), \tag{6.53}$$

$$f(x|n+1, a, b) = \frac{(b-n+x)(n+1)}{(n+1-x)(a+b-n)} f(x|n, a, b), \tag{6.54}$$

$$f(x|n, a, b+1) = \frac{(a+b-n+1)(b+1)}{(b-n+x+1)(a+b+1)} f(x|n, a, b). \tag{6.55}$$

Also

$$\begin{aligned} f(x|n, a, b) &= f(n-x|n, b, a) \\ &= f(a-x|a+b-n, a, b) \\ &= f(b-n+x|a+b-n, b, a); \end{aligned} \tag{6.56}$$

see Lieberman and Owen (1961). Furthermore

$$\begin{aligned} F(x|n, a, b) &= 1 - F(n-x-1|n, b, a) \\ &= F(b-n+x|a+b-n, b, a) && (6.57) \\ &= 1 - F(a-x-1|a+b-n, a, b). && (6.58) \end{aligned}$$

Raiffa and Schlaifer (1961) obtained relationships between the tails of the hypergeometric, beta–binomial (negative hypergeometric), and beta–negative binomial distributions. These authors used a different notation from Lieberman and Owen (1961). Let $F_h(\cdot)$ and $G_h(\cdot)$ denote the lower and upper tails of a classical hypergeometric distribution; then

$$\begin{aligned} &G_h(k|n, \ell+m-1, k+m-1) \\ &\quad = \sum_{x \geq k} \binom{k+m-1}{x} \binom{\ell+n-k}{n-x} \Big/ \binom{\ell+m+n-1}{n} \\ &\quad = \sum_{x \leq n-k} \binom{\ell+n-k}{x} \binom{k+m-1}{n-x} \Big/ \binom{\ell+m+n-1}{n} \\ &\quad = F_h(n-k|n, \ell+m-1, \ell+n-k). \end{aligned} \tag{6.59}$$

Furthermore, let

$$\begin{aligned} G_{\beta b}(k|m, \ell+m, n) &= \sum_{x \geq k} \binom{-m}{x} \binom{-\ell}{n-x} \Big/ \binom{-\ell-m}{n} \\ &= \sum_{x \geq k} \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \times \frac{p^{m-1}(1-p)^{\ell-1}\, dp}{B(\ell, m)} \end{aligned} \tag{6.60}$$

be the upper tail of a beta–binomial (negative hypergeometric) distribution. Also let

$$\begin{aligned} F_{\beta Pa}&(n|m, \ell + m, k) \\ &= \sum_{x \le n-k} \binom{-k}{x}\binom{k+m-1}{-\ell - x} \Big/ \binom{m-1}{-\ell} \\ &= \sum_{x \le n-k} \int_0^1 \binom{k+x-1}{x}(1-\lambda)^k \lambda^x \times \frac{\lambda^{\ell-1}(1-\lambda)^{m-1}\,d\lambda}{B(\ell, m)} \end{aligned} \tag{6.61}$$

denote the lower tail of a beta–negative binomial (beta–Pascal) distribution. Then from the relationship between the tails of a binomial and a negative binomial distribution (Section 5.6), Raiffa and Schlaifer proved that

$$G_{\beta b}(k|m, \ell + m, n) = F_{\beta Pa}(n|m, \ell + m, k)$$

and hence that

$$G_{\beta b}(k|m, \ell + m, n) = F_{\beta Pa}(n|m, \ell + m, k) \tag{6.62}$$

$$= G_h(k|n, \ell + m - 1, k + m - 1) \tag{6.63}$$

$$= F_h(n - k - 1|n, \ell + m - 1, \ell + n - k) \tag{6.64}$$

by a probabilistic argument; see Raiffa and Schlaifer (1961, pp. 238–239).

From (6.52), $\Pr[X = x + 1]$ is greater or less than $\Pr[X = x]$ according as

$$\frac{(a - x)(n - x)}{(x + 1)(b - n + x + 1)} \gtrless 1,$$

that is, according as

$$x \gtrless \frac{(n + 1)(a + 1)}{(a + b + 2)} - 1. \tag{6.65}$$

Let $c = (n + 1)(a + 1)/(a + b + 2)$. Then $\Pr[X = x]$ increases with x, reaching a maximum at the greatest integer that does not exceed c, and then decreases. The mode of the distribution is therefore at $[c]$, where $[\cdot]$ denotes the integer part. If c is an integer, then there are two equal maxima at $c - 1$ and c. [Note that, if a and b are large, then the mode is very close to the mean, since $\mu = na/(a + b)$.]

The classical hypergeometric distribution is known to have a monotone likelihood ratio in x for known values of n and $a + b$ (Ferguson, 1967).

The following limiting results hold:

(i) The classical hypergeometric distribution tends to a Poisson distribution with mean μ as $n \to \infty$, $(a + b) \to \infty$ such that $na/(a + b) = \mu$, μ constant. Feller (1957), Nicholson (1956), Molenaar (1970a), and Lieberman and Owen (1961) have examined conditions under which it tends to a normal distribution.

(ii) As $a \to \infty$, $b \to \infty$ such that $a/(a+b) = p$, p constant, $0 < p < 1$, a type IA(i) distribution tends to a binomial distribution with parameters n, p.

(iii) If a is a positive integer, then as $n \to \infty$, $b \to \infty$ such that $n/b = p$, p constant, $0 < p < 1$, a type IA(ii) distribution tends to a binomial distribution with parameters a, p.

(iv) As a and b both tend to $-\infty$ in such a way that $a/(a+b) = p$, p constant, $0 < p < 1$, a type IIA distribution tends to a binomial distribution with parameters n, p.

(v) As $a \to \infty, b \to -\infty$ such that $a/b = -\lambda, \lambda$ constant, $0 < \lambda < 1$, a type IIIA distribution tends to a negative binomial distribution with pgf $(1-\lambda)^k/(1-\lambda z)^k$, where $k = -n$.

(vi) As $a \to -\infty, b \to \infty$ such that $a/b = -\lambda, \lambda$ constant, $0 < \lambda < 1$, a type IV distribution similarly tends to a negative binomial distribution.

(vii) From the duality relationship between type IIA and type IIIA, a type IIA distribution can also tend to a negative binomial and a type IIIA can tend to a positive binomial distribution.

A comparison between hypergeometric types IA(i), IIA, IIIA, and IV, binomial, Poisson, and negative binomial distributions [from Kemp and Kemp (1956a)] is presented in Table 6.3; see also Table 3.1.

6.5 APPROXIMATIONS AND BOUNDS

There is a considerable variety of approximations to the individual probabilities, and also to cumulative sums of probabilities, for the classical hypergeometric distribution. Many of these are based on the approximation of the hypergeometric distribution (6.1) by a binomial distribution with parameters n, p.

Sródka (1963) obtained some very good bounds on the probabilities:

$$\binom{n}{x}\left(\frac{Np-x}{N}\right)^x\left(\frac{N-Np-n+x}{N}\right)^{n-x}\left(1+\frac{6n^2-6n-1}{12N}\right)$$
$$< \Pr[X=x] < \binom{n}{x}p^x(1-p)^{n-x}\left(1-\frac{n}{N}\right)^{-n}\left(1+\frac{6n^2+6n-1}{12N}\right)^{-1}. \tag{6.66}$$

For sufficiently large N these can be simplified to

$$\binom{n}{x}\left(\frac{Np-x}{N}\right)^x\left(\frac{N-Np-n+x}{N}\right)^{n-x}$$
$$< \Pr[X=x] < \binom{n}{x}p^x(1-p)^{n-x}\left(1-\frac{n}{N}\right)^{-n}. \tag{6.67}$$

It is often adequate to use the simple binomial approximation

$$\Pr[X = x] \approx \binom{n}{x} p^x (1-p)^{n-x} \qquad (6.68)$$

when $n < 0.1N$.

There is a marked improvement if n and p are replaced by n^* and p^*, where

$$p^* = \frac{(n-1) + (N-n)p}{N-1} \qquad \text{and} \qquad n^* = \frac{np}{p^*}, \qquad (6.69)$$

that is, if $n^* p^*$ and $n^* p^*(1-p^*)$ are set equal to the theoretical mean and variance of the hypergeometric distribution (Sandiford, 1960).

Greater accuracy still may be obtained by using the following modification suggested by Ord (1968a):

$$\Pr[X = x] \approx \binom{n}{x} p^x (1-p)^{n-x} \left[1 + \frac{x(1-2p) + np^2 - (x-np)^2}{2Np(1-p)}\right]. \qquad (6.70)$$

Burr (1973) showed that

$$\Pr[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \left[1 + \frac{x - (x-np)^2}{2Np} + O\left(\frac{1}{N^2 p^2}\right)\right], \qquad (6.71)$$

and that for $n > Np$ a closer approximation is obtained by interchanging the roles of n and Np.

Ma (1982) independently derived an approximation for $n \leq Np$ that is equivalent to Ord's approximation. He also showed that when $n > Np$ interchanging the roles of n and Np gives a better approximation.

Since, as already noted, the hypergeometric distribution is unchanged by interchanging n and Np, it is clear that a binomial with parameters Np, n/N has a claim equal to that of a binomial with parameters n, p as an approximating distribution for (6.1). In addition, the distribution of $n - x$ could be approximated by a binomial with parameters $N - Np$, n/N; similarly the distribution of $Np - x$ could be approximated by a binomial with parameters $N - n$, p. Brunk et al. (1968) compared these approximations. Their investigations support the opinion of Lieberman and Owen (1961) that it is best to use the binomial with smallest power parameter, that is, $\min(n, Np, N - Np, N - n)$.

The following binomial-type approximation for the cumulative probabilities was obtained by Wise (1954):

$$\sum_{j=0}^{x} \Pr[X = j] \approx \sum_{j=0}^{x} \binom{n}{x} w^j (1-w)^{n-j}, \qquad (6.72)$$

where $w = \left(Np - \frac{1}{2}x\right) / \left(N - \frac{1}{2}n + \frac{1}{2}\right)$; that is, he showed that the distribution (6.1) is approximated by a binomial distribution with parameters n and

$$\left(Np - \tfrac{1}{2}x\right) / \left(N - \tfrac{1}{2}n - \tfrac{1}{2}\right).$$

A more complicated approximation of a similar type was constructed by Bennett (1965).

Molenaar (1970a) found that the use of

$$p^{\ddagger} = \frac{Np - x/2}{N - (n-1)/2} - \frac{n\left(x - np - \frac{1}{2}\right)}{6[N - (n-1)/2]^2} \tag{6.73}$$

gives very accurate results even when $n/N > 0.1$.

Uhlmann (1966) made a systematic comparison between the hypergeometric distribution (with parameters n, Np, N, where $0 < p < 1$) and the binomial distribution (with parameters n, p). Denoting $\Pr[X \le c]$ for the two distributions by $L_{N,n,c}(p)$ and $L_{n,c}(p)$, respectively, he showed that in general

$$L_{N,n,c}(p) - L_{n,c}(p) \begin{cases} =0 & \text{for } p = 0, \\ >0 & \text{for } 0 < p \le c(n-1)^{-1}N(N+1)^{-1}, \\ <0 & \text{for } c(n-1)^{-1}N(N+1)^{-1} + (N+1)^{-1} \le p < 1, \\ =0 & \text{for } p = 1; \end{cases}$$

these results simplify when n is odd.

Subsequently, in a study of the relationship between hypergeometric and binomial pmf's, Ahrens (1987) used simple majorizing functions (upper bounds) for their ratio. He found that the ratio of the hypergeometric to the binomial pmf can always be kept below $\sqrt{2}$ by a suitable choice of approximating binomial.

If the binomial approximation to the hypergeometric distribution can itself be approximated by a Poisson or normal approximation (see Section 3.6.1), then there is a corresponding Poisson or normal approximation to the hypergeometric. Thus when p is small but n is large, the Poisson approximation

$$\Pr[X = x] \approx \frac{e^{-np}(np)^x}{x!} \tag{6.74}$$

may be used. Burr (1973) sharpened this approximation.

The χ^2-test for association in a 2×2 contingency table uses a χ^2 approximation for the tail of a hypergeometric distribution; this is

$$\Pr[X \le x] \approx \Pr[\chi^2_{[1]} \ge T], \tag{6.75}$$

where

$$T = \frac{(N-1)(x - np)^2}{p(1-p)n(N-n)}.$$

The relationship between the tails of a χ^2 distribution (Johnson et al., 1994, Chapter 17) and a Poisson distribution means that this is a Poisson-type approximation for the cumulative hypergeometric probabilities. It can be improved by the use of Yates' correction, giving "the usual $\frac{1}{2}$-corrected chi-statistic" of Ling and Pratt (1984). It is appropriate for n large, provided that p is not unduly small.

From the relationship between the χ^2 distribution with one degree of freedom and the normal distribution, we have, when p is not small and n is large,

$$\Pr[X \leq x] \approx (2\pi)^{-1/2} \int_{-\infty}^{y} \exp\left(-\frac{u^2}{2}\right) du, \tag{6.76}$$

with

$$y = \frac{x - np + \frac{1}{2}}{[(N-n)np(1-p)/(N-1)]^{1/2}}.$$

Hemelrijk (1967) reported that, unless the tail probability is less than about 0.07 *and* $Np + n \leq N/2$, some improvement is effected by replacing $(N-1)^{-1}$ by N^{-1} under the square-root sign. A more refined normal approximation was proposed by Feller (1957) and Nicholson (1956).

Pearson (1906) approximated hypergeometric distributions by (continuous) Pearson-type distributions. This work was continued by Davies (1933, 1934). The Pearson distributions that appeared most promising were type VI or III. Bol'shev (1964) also proposed an approximation of this kind that gives good results for $N \geq 25$.

Normal approximations would, however, seem to be the most successful. Ling and Pratt (1984) carried out an extensive empirical study of 12 normal and 3 binomial approximations for cumulative hypergeometric probabilities, including two relatively simple normal approximations put forward by Molenaar (1970a, 1973). Ling and Pratt considered that binomial approximations are not appropriate as competitors to normal approximations because of the computational problems with the tails. The four normal approximations that they found best originated from an unpublished paper; this was submitted to the *Journal of the American Statistical Association* by D. B. Peizer in 1968 but was never revised or resubmitted. Peizer's approximations are extremely good, but they are considerably more complicated than those above of Molenaar. See Ling and Pratt (1984) for details of the approximations that they studied.

Some new binomial approximations to the hypergeometric distribution have recently been obtained by López-Blázquez and Salamanca Miño (2000).

6.6 TABLES, COMPUTATION, AND COMPUTER GENERATION

An extensive set of tables of individual and cumulative probabilities for the classical hypergeometric distribution was prepared by Lieberman and Owen (1961). They give values for individual and cumulative probabilities to six decimal places for

$$N = 2(1)50(10)100, \qquad Np = 1(1)N-1, \qquad n = 1(1)Np.$$

Less extensive tables were published earlier by Chung and DeLury (1950). Graphs based on hypergeometric probabilities were given by Clark and Koopmans (1959).

Guenther (1983) thought that the best way to evaluate $\Pr[X \leq x]$ is from the Lieberman and Owen tables or by means of a packaged computer program. Computer algorithms have been provided by Freeman (1973) and Lund (1980). Lund's algorithm was improved by Shea (1989) and by Berger (1991).

Little serious attention seems to have been given to the use of Stirling's expansion for the computation of individual hypergeometric probabilities.

Computer generation of classical hypergeometric random variables has been discussed in detail by Kachitvichyanukul and Schmeiser (1985). When the parameters remain constant, the alias method of Walker (1977) and Kronmal and Peterson (1979) is a good choice. Kachitvichyanukul and Schmeiser gave an appropriate program with safeguards to avoid underflow.

The simplest of all algorithms needs a very fast uniform generator. It is based on a sequence of trials in which the probability of success depends on the number of previous successes; that is, it uses the model of finite sampling without replacement. The number of successes in a fixed number of trials is counted. Fishman (1973) and McGrath and Irving (1973) have given details.

Fishman's (1978) algorithm requires a search of the cdf. Kachitvichyanukul and Schmeiser (1985) suggest ways in which the speed of the method can be improved.

Devroye (1986) indicated in an exercise how hypergeometric rv's can be generated by rejection from a binomial envelope distribution.

For large-scale simulations with changing parameters, Kachitvichyanukul and Schmeiser's algorithm H2PE uses acceptance/rejection from an envelope consisting of a uniform with exponential tails. The execution time is bounded over the range of parameter values for which the algorithm is intended, that is, over the range $M - \max(0, n - N + Np) \geq 10$, where M is the mode of the distribution. Kachitvichyanukul and Schmeiser recommended inversion of the cumulative distribution function for other parameter values.

While the negative hypergeometric distribution could be generated by inverse sampling without replacement for a fixed number of successes, it would seem preferable to generate it using its beta–binomial model with the good extant beta and binomial generators.

Similarly a hypergeometric type IV distribution could be generated as a beta mixture of negative binomials; see Section 6.2.5.

6.7 ESTIMATION

Most papers on estimation for hypergeometric-type distributions have concentrated on particular distributions (e.g., beta–binomial distribution). Rodríguez-Avi et al. (2003) have recently studied a variety of estimation methods for distributions with pgf's of the general form

$$G(z) = \frac{{}_2F_1[\alpha, \beta; \lambda; z]}{{}_2F_1[\alpha, \beta; \lambda; 1]}.$$

These include (i) methods based on relations between moments and frequencies and the observed values; (ii) the minimum χ^2 procedure; and (iii) maximum likelihood. Two applications to real data are provided.

6.7.1 Classical Hypergeometric Distribution

In one of the most common situations, inspection sampling (see Section 6.9.1), there is a single observation of r defectives (successes!) in a sample of size n taken from a lot of size N. Both N and n are known, and the hypergeometric parameter Np denotes the number of defectives in a lot; an estimate of Np is required.

The maximum-likelihood estimator $\widehat{Np}$ is the integer maximizing

$$\binom{\widehat{Np}}{r}\binom{N-\widehat{Np}}{n-r} \tag{6.77}$$

for the observed value r. From the relationship between successive probabilities,

$$\Pr[X = r|n, Np+1, N] \gtrless \Pr[X = r|n, Np, N]$$

according as $Np \gtrless n^{-1}r(N+1) - 1$. Hence $\widehat{Np}$ is the greatest integer not exceeding $r(N+1)/n$; if $r(N+1)/n$ is an integer, then $[r(N+1)/n] - 1$ and $r(N+1)/n$ both maximize the likelihood. The variance of $r(N+1)/n$ is, from (6.8),

$$\frac{(N+1)^2(N-n)p(1-p)}{n(N-1)}.$$

Neyman confidence intervals for Np have been tabulated extensively, notably by Chung and DeLury (1950) and Owen (1962), and have been used widely. Steck and Zimmer (1968) outlined how these may be obtained; see also Guenther (1983). Steck and Zimmer also derived Bayes confidence intervals for Np based on various special cases of a Pólya prior distribution. They related these to Neyman confidence intervals. It would seem that Bayes confidence intervals are highly sensitive to choice of prior distribution.

A test of $Np = a$ against $Np = a_0$ is sometimes required. Guenther (1977) discussed hypothesis testing in this context, giving numerical examples.

In the simplest capture–recapture application (again see Section 6.9.1) we want to estimate the total size of a population N, with both $n_1 = Np$ (the number caught on the first occasion) and $n_2 = n$ (the number caught on the second occasion) known, given a single observation r. Here

$$\Pr[X = r|n_2, n_1, N+1] \gtrless \Pr[X = r|n_2, n_1, N] \tag{6.78}$$

according as $N \gtrless (n_1n_2/r) - 1$. Hence the MLE $\hat{N}$ of N is the greatest integer not exceeding n_1n_2/r; if n_1n_2/r is an integer, then $(n_1n_2/r) - 1$ and n_1n_2/r both maximize the likelihood.

The properties of the estimator $\hat{N}$ have been discussed in detail by Chapman (1951). Usually $n_1 + n_2 \not> N$, in which case the moments of $\hat{N}$ are infinite. Because of the problems of bias and variability concerning $\hat{N}$, Chapman suggested instead the use of the estimator

$$N^* = \frac{(n_1+1)(n_2+1)}{r+1} - 1. \tag{6.79}$$

He found that

$$E[N^* - N] = \frac{(n_1+1)(n_2+1)(N-n_1)!(N-n_2)!}{(N+1)!(N-n_1-n_2-1)!}, \tag{6.80}$$

which is less than 1 when $N > 10^4$ and $n_1n_2/N > 9.2$. The variance of N^* is approximately

$$N^2(m^{-1} + 2m^{-2} + 6m^{-3}), \tag{6.81}$$

and its coefficient of variation is approximately $m^{-1/2}$, where $m = n_1n_2/N$. Chapman (1951, p. 148) concluded that "sample census programs in which the expected number of tagged members is much smaller than 10 may fail to give even the order of magnitude of the population correctly."

Robson and Regier (1964) discussed the choice of n_1 and n_2. Chapman (1948, 1951) showed how large-sample confidence intervals for N^* can be constructed; see also Seber (1982b).

The estimator

$$N^{**} = \frac{(n_1+2)(n_2+2)}{r+2} \tag{6.82}$$

has also been suggested for n_2 sufficiently large. Here

$$\begin{aligned} E(N^{**}) &\approx N(1 - m^{-1}), \\ \operatorname{Var}(N^{**}) &\approx N^2(m^{-1} - m^{-2} - m^{-3}). \end{aligned} \tag{6.83}$$

In epidemiological studies the estimation of a target population size N is quite often achieved by merging two lists of sizes n_1 and n_2; see Section 9.1. Here it is usual to have $n_1 + n_2 > N$, in which case $\hat{N}$ is unbiased, and

$$s^2 = \frac{(n_1+1)(n_2+1)(n_1-r)(n_2-r)}{(r+1)^2(r+2)}$$

(where r is the number of items in common on the two lists) is an unbiased estimator of $\operatorname{Var}(\hat{N})$ (Wittes, 1972).

6.7.2 Negative (Inverse) Hypergeometric Distribution: Beta–Binomial Distribution

The *beta–binomial distribution* is the most widely used of all the general hypergeometric distributions. It is particularly useful for regression situations involving binary data.

Consider the beta–binomial parameterization

$$\Pr[X = x] = \binom{-\alpha}{x}\binom{-\beta}{n-x}\Big/\binom{-\alpha-\beta}{n} \tag{6.84}$$

from Section 6.2.2, with $\alpha, \beta, n > 0$, n an integer. Moment and maximum-likelihood estimation procedures for the parameters α and β were devised by Skellam (1948) and Kemp and Kemp (1956b).

The moment estimators are obtained by setting

$$\overline{x} = \frac{\tilde{\alpha} n}{\tilde{\alpha} + \tilde{\beta}}, \qquad s^2 = \frac{n\tilde{\alpha}\tilde{\beta}(\tilde{\alpha} + \tilde{\beta} + n)}{(\tilde{\alpha} + \tilde{\beta})^2(\tilde{\alpha} + \tilde{\beta} + 1)}, \tag{6.85}$$

that is,

$$\tilde{\alpha} = \frac{(n - \overline{x} - s^2/\overline{x})\overline{x}}{(s^2/\overline{x} + \overline{x}/n - 1)n}, \qquad \tilde{\beta} = \frac{(n - \overline{x} - s^2/\overline{x})(n - \overline{x})}{(s^2/\overline{x} + \overline{x}/n - 1)n}. \tag{6.86}$$

Maximum-likelihood estimation is reminiscent of maximum-likelihood estimation for the negative binomial distribution. Let the observed frequencies be f_x, $x = 0, 1, \ldots, n$, and set

$$A_x = f_{x+1} + f_{x+2} + \cdots + f_n, \qquad B_x = f_0 + f_1 + \cdots + f_x;$$

then the total number of observations is $A_{-1} = B_n$.

The maximum-likelihood equations are

$$\begin{aligned} 0 = F &\equiv \sum_{x=0}^{n-1} \frac{A_x}{\hat{\alpha} + x} - \sum_{x=0}^{n-1} \frac{A_{-1}}{\hat{\alpha} + \hat{\beta} + x}, \\ 0 = G &\equiv \sum_{x=0}^{n-1} \frac{B_x}{\hat{\beta} + x} - \sum_{x=0}^{n-1} \frac{A_{-1}}{\hat{\alpha} + \hat{\beta} + x}. \end{aligned} \tag{6.87}$$

Iteration is required for their solution. Given initial estimates α_1 and β_1 (e.g., the moment estimates), corresponding values of F_1 and G_1 can be computed; better estimates, α_2 and β_2, can then be obtained by solving the simultaneous linear equations

$$\begin{aligned} F_1 &= (\alpha_2 - \alpha_1) \sum_{x=0}^{n-1} \frac{A_x}{(\alpha_1 + x)^2} - (\alpha_2 - \alpha_1 + \beta_2 - \beta_1) \sum_{x=0}^{n-1} \frac{A_{-1}}{(\alpha_1 + \beta_1 + x)^2}, \\ G_1 &= (\beta_2 - \beta_1) \sum_{x=0}^{n-1} \frac{B_x}{(\beta_1 + x)^2} - (\alpha_2 - \alpha_1 + \beta_2 - \beta_1) \sum_{x=0}^{n-1} \frac{A_{-1}}{(\alpha_1 + \beta_1 + x)^2}. \end{aligned} \tag{6.88}$$

The next cycle is then begun by calculating F_2 and G_2. Kemp and Kemp (1956b, p. 174) reported that "on average about five cycles were needed to stabilize the estimates to three decimal places."

Chatfield and Goodhardt (1970) put forward an estimation method based on the mean and zero frequency. For highly J-shaped distributions (e.g., distributions of numbers of items purchased) this method has good efficiency; however, it does require iteration.

Griffiths (1973) remarked that the MLEs can be obtained by the use of a computer algorithm to maximize the log-likelihood. Williams (1975), like Griffiths, considered it advantageous to reparameterize, taking $\pi = \alpha/(\alpha + \beta)$ (the mean of the beta distribution) and $\theta = 1/(\alpha + \beta)$ (a shape parameter). Williams was hopeful that convergence of a likelihood maximization algorithm would be more rapid with this parameterization.

Qu et al. (1990) pointed out that $\alpha > 0$, $\beta > 0$, that is, $\theta > 0$, gives the beta–binomial distribution, while $\alpha < 0$, $\beta < 0$, that is, $\theta < 0$, gives the hypergeometric distribution. When $\alpha \to \infty$, $\beta \to \infty$, that is, $\theta \to 0$, the distribution tends to the binomial. Hence maximum-likelihood estimation with this parameterization enables an appropriate distribution from within this group to be fitted to data without assuming one particular distribution. Qu et al. also proposed methods of testing $H_0 : \theta = 0$ using (1) a Wald statistic and (2) the likelihood ratio. He showed how the homogeneity of the parameters can be tested using the deviance.

Bowman, Kastenbaum, and Shenton (1992) have shown that the joint efficiency for the method of moments estimation of α and β is very high over much of the parameter space. Series were derived for the first four moments of the moment estimators; simulation approaches were used for validation.

Various methods of estimation for the beta–binomial and zero-truncated beta–binomial distributions are explored in Tripathi, Gupta, and Gurland (1994). They are compared with maximum likelihood on the basis of asymptotic relative efficiency. Examples of their use and recommendations are provided.

Moment and maximum-likelihood estimation for the case where all three parameters α, β, and n are unknown was discussed in outline by Kemp and Kemp (1956b).

6.7.3 Beta–Pascal Distribution

Dubey (1966a) studied estimation for the beta–Pascal distribution assuming that k is known.

Given the beta–Pascal distribution with parameterization

$$\Pr[X = y] = \int_0^1 \binom{y-1}{k-1} (1-\lambda)^k \lambda^{y-k} \times \frac{\lambda^{\ell-1}(1-\lambda)^{m-1}\,d\lambda}{B(\ell, m)}, \qquad y = k, k+1, \ldots, \tag{6.89}$$

the mean and variance are

$$\mu = k + \frac{k\ell}{m-1} = \frac{k(\ell+m-1)}{m-1}, \quad \text{provided that } m > 1,$$
$$\mu_2 = \frac{k\ell(k+m-1)(\ell+m-1)}{(m-1)^2(m-2)}, \quad \text{provided that } m > 2 \tag{6.90}$$

(note that the support is $k, k+1, \ldots$). Hence the moment estimates are

$$\tilde{m} = 2 + \frac{\overline{x}(\overline{x}-k)(k+1)}{(s^2k - \overline{x}^2 + k\overline{x})}, \qquad \tilde{\ell} = \frac{(\tilde{m}-1)(\overline{x}-k)}{k}. \tag{6.91}$$

Dubey also discussed maximum-likelihood estimation (assuming that k is known). There are close parallels with maximum-likelihood estimation for the beta–binomial distribution. Irwin (1975b) described briefly maximum-likelihood estimation procedures for the three-parameter generalized Waring distribution (i.e., k unknown).

6.8 CHARACTERIZATIONS

There are several characterizations for hypergeometric-type distributions.

Patil and Seshadri's (1964) very general result for discrete distributions has the following corollaries:

1. Iff the conditional distribution of X given $X+Y$ is hypergeometric with parameters a and b, then X and Y have binomial distributions with parameters of the form (a, θ) and (b, θ), respectively.
2. Iff the conditional distribution of X given $X+Y$ is negative hypergeometric with parameters α and β for all values of $X+Y$, then X and Y have negative binomial distributions with parameters of the form (α, θ) and (β, θ), respectively.

Further details are in Kagan, Linnik, and Rao (1973).

Consider now a family of $N+1$ distributions indexed by $j = 0, 1, \ldots, N$, each supported on a subset of $\{0, 1, \ldots, n\}, n \le N$. Skibinsky (1970) showed that this is the hypergeometric family with parameters N, n, j iff for each θ, $0 \le \theta \le 1$, the mixture of the family with binomial (N, θ) mixing distribution is the binomial (n, θ) distribution. Skibinsky restated this characterization as follows: Let $h_0, h_1, \ldots, h_N$ denote $N+1$ functions on $\{0, 1, \ldots, n\}$; then

$$h_j(i) = \binom{j}{i}\binom{N-j}{n-i}\bigg/\binom{N}{n}, \tag{6.92}$$

$i = 0, 1, \ldots, n$, $j = 0, 1, \ldots, N$, iff the h_i are independent of θ and

$$\sum_{j=0}^{N} h_j b(j; N, \theta) = b(\cdot; n, \theta), \qquad 0 \le \theta \le 1, \tag{6.93}$$

where $b(\cdot; n, \theta)$ is the binomial pmf with parameters n and θ. He reinterpreted this characterization as follows: Suppose that X has a binomial distribution with parameters N and θ and that Y is distributed on $0, 1, \ldots, n$; suppose also that h_j is defined as in (6.92). Then Y has a binomial distribution with parameters n and θ if the conditional distributions of $Y|X$ are specified by h_x. Conversely, if the conditional distributions of $Y|X$ are independent of θ, $0 \leq \theta \leq 1$, and Y is binomially distributed with parameters (n, θ), then the conditional distributions of $Y|X$ are specified by h_x.

Skibinsky noted the connection between his characterization and certain results of Mood (1943) and Hald (1960) concerning acceptance sampling. Neville and Kemp (1975) elaborated on this. Given a population of lots of fixed size N such that the distribution of the number of defectives per lot is binomial, suppose that a sample of size n is drawn without replacement from one of these lots and that the distribution of the number of defectives in a single sample is hypergeometric. Then the overall distribution of the number of defectives in such samples is binomial. In Hald's terminology, the binomial distribution is reproducible with respect to sampling without replacement.

Neville and Kemp (1975), also Janardan (1973a) in a departmental report, found that a similar characterization holds if the initial distribution of the number of defectives per lot is hypergeometric; the classical hypergeometric distribution is also reproducible with respect to sampling without replacement.

Patil and Ratnaparkhi (1977) obtained characterization results for the Poisson, binomial, and negative binomial distributions, given a bivariate observation with the second component subject to damage and linearity of regression of the first component on the second under the transition of the second component from the original to the damaged state. They described how these results can be extended to provide characterizations for the hypergeometric and negative hypergeometric distributions.

Qu et al. (1990) have proved the following characterization theorem for distributions with pmf of the form

$$\binom{n}{x}\prod_{i=0}^{x-1}(A+iS)\prod_{i=0}^{n-x-1}(B+iS)\Big/\prod_{i=0}^{n-1}(A+B+iS)$$
$$= \binom{n}{x}\binom{-1/\theta - n}{-\pi/\theta - x}\Big/\binom{-1/\theta}{-\pi/\theta}, \qquad (6.94)$$

where $\pi = A/(A+B)$ and $\theta = S/(A+B)$ (i.e., for beta–binomial, binomial, and hypergeometric distributions, according as θ is greater, equal, or less than zero): Let $Y_j (j = 1, \ldots, n)$ be n binary variables taking the values 0 and 1. Then $Y = \sum_{j=1}^{n} Y_i$ has the above distribution iff the following conditions are satisfied:

1. The conditional expectation of Y_k, $k = 2, 3, \ldots, n$, with respect to $Y_1, \ldots, Y_{k-1}$ is a linear function of the observed values $Y_1, \ldots, Y_{k-1}$.

2. The expectation for each variable is the same; that is, $E[Y_j] = \pi,\ j = 1, \ldots, n$.
3. The correlation between each pair of variables Y_i and Y_j is the same and is equal to

$$\frac{E[(Y_i - E[Y_i])(Y_j - E[Y_j])]}{\sqrt{E[(Y_i - E[Y_i])^2]E[(Y_j - E[Y_j])^2]}} = \frac{\theta}{1+\theta} = r \qquad (6.95)$$

(the interclass correlation), where $i = 1, \ldots, n,\ j > i$.

Xekalaki (1981) gave two characterizations for the generalized Waring distribution based on conditionality properties. Papageorgiou (1985) has obtained characterizations of the hypergeometric and negative hypergeometric distributions using the linear regression of one random variable on another and the conditional distribution of the latter given the first variable; he suggested an application in genetics. Further characterizations of this kind are in Kyriakoussis and Papageorgiou (1991a).

6.9 APPLICATIONS

6.9.1 Classical Hypergeometric Distribution

Besides providing a fund of teaching examples, the classical hypergeometric distribution has a number of important practical applications.

In industrial quality control, lots of size N containing a proportion p of defectives are sampled using samples of fixed size n. The number of defectives X per sample is then a hypergeometric rv. If $x \le c$ (the acceptance number), the lot is accepted; otherwise it is rejected. The design of suitable sampling plans requires the calculation of confidence intervals for Np, given c, N, and n. Tables of these have been published by Chung and DeLury (1950) and Owen (1962). In many cases binomial or Poisson approximations to the hypergeometric distribution suffice.

Another useful application is in the estimation of the size of animal and other populations from "capture–recapture" data. This kind of application dates back at least to Peterson (1896), quoted by Chapman (1952a). Consider, for example, the estimation of the number N of fish in a pond. First a known number m of fish are netted, marked (tagged), and returned to the pond. A short time later, long enough to ensure (hopefully) random dispersion of the tagged fish but not long enough for natural changes to affect the population size, a sample of size n is taken from the pond and the number X of tagged fish in the second sample is observed. The assumptions that the samples have fixed sizes m and n and the fish behave randomly and independently lead to a hypergeometric distribution for X.

An alternative viewpoint is to regard the N organisms in the population as N independent trials each with the same probability of belonging to a given one of the four capture–recapture categories. This gives rise to a multinomial model where the probability of an observed pattern of captures and recaptures is proportional to the hypergeometric pmf (Cormack, 1979). The feature that the sample sizes can be regarded as either fixed or random, with little change in the analysis of the data, occurs in other, more elaborate capture–recapture sampling schemes, such as schemes with multiple recaptures at successive points of time and those with open (that is, changeable) populations rather than closed populations (Seber, 1982a).

There are variations on the situations leading to the basic hypergeometric capture–recapture model. As mentioned in Section 6.7.1, the estimation of a target population N in epidemiological studies (e.g., the number of anencephalics born in Boston Lying-In Hospital over several decades) can be achieved by counting the number of cases that appear on both of two lists of sizes n and m (a diagnostic registry list and a list of discharge summaries) (Wittes, 1972).

In opinion surveys a random sample of respondents of size n is drawn without replacement from a finite population of size N. From the proportion in the sample who answer a particular question positively, it is desired to estimate the proportion in the whole population who would answer positively.

A further important use of the hypergeometric distribution is in the analysis of 2×2 contingency tables with both sets of marginal frequencies fixed. The probability of a result as extreme as the observed result is the tail probability for a classical hypergeometric distribution. When testing whether or not the two dichotomies are independent (i.e., testing homogeneity), the test based on the computation of a hypergeometric cumulative probability is known as *Fisher's exact test* (see, e.g., Gibbons, 1983). This test was suggested in the mid-1930s by Fisher, Irwin, and Yates; Gibbons has provided relevant references. This test can also be used in the case where only one set of marginal frequencies is fixed; here the test is conditional on the observed values of the unfixed marginal frequencies. Similarly it can be used when neither set of marginal frequencies is fixed.

Cochran (1954) recommended the use of Fisher's exact test, in preference to the approximate χ^2-test with continuity correction, (1) when $N \leq 20$ and (2) when $20 < N < 40$ and the smallest expected frequency is less than 5, where N is the total frequency. Tables for the test are provided in Pearson and Hartley (1976); see also Finney et al. (1963). Significance tests and confidence intervals for the odds ratio are described in Cox and Snell (1989) along with useful bibliographic notes.

The test has been used widely. Yates (1934) illustrated its use with data on breast- and bottle-fed children with normal and maloccluded teeth. Cox and Snell (1989) have examined data on physicians dichotomized as smokers/nonsmokers and as cancer patients/controls. We note that 2×2 tables can arise in two distinct contexts: one sample with two dichotomously observed variables (as in the two preceding examples) and two independent samples with one dichotomous variable. A sample of pregnant women who smoke and a second sample of pregnant nonsmokers, with both samples dichotomized by low/normal birthweight of infant, provides an example of the second context. A discussion of an application of Fisher's exact test in linguistics is given in the first edition of this book.

Suppose with Yu, Waller, and Zelterman (1998) that there are N twin pairs and that at time t we find m of these individuals still alive and $2N - m$ dead ($0 \le m \le 2N$). Let the original $2N$ individuals be matched to form N pairs (without replacement) and suppose that in X_1 pairs both individuals are still alive, in X_2 pairs exactly one individual is still alive, and in X_3 pairs both are dead ($X_1 + X_2 + X_3 = N$). Then

$$\Pr[X_1 = k | 2X_1 + X_2 = m] \propto \frac{N!2^{m-2k}}{k!(m-2k)!(N-m+k)!}.$$

The authors showed that the pgf for this distribution is

$$G(z) = \frac{{}_2F_1[-m/2, (1-m)/2; N-m+1; z]}{{}_2F_1[-m/2, (1-m)/2; N-m+1; 1]}.$$

[This is a IA(i)-type hypergeometric distribution since either $m/2$ or $(m-1)/2$ is a positive integer.] The mean and variance are

$$\mu = \frac{m(m-1)}{2(2N-1)}, \qquad \mu_2 = \frac{m(m-1)(2N-m)(2N-m-1)}{2(2N-1)^2(2N-3)}.$$

Yu et al. (1998) applied the distribution to the Danish Twin Registry data on monozygotic female twins born between 1870 and 1880. They found that an excessive number of the twin pairs were both alive at older ages than one would expect.

Further references concerning properties and applications of the classical hypergeometric distribution are in Patil, Boswell, Joshi, and Ratnaparkhi (1984) and Wimmer and Altmann (1999).

6.9.2 Negative (Inverse) Hypergeometric Distribution: Beta–Binomial Distribution

The waiting-time model for the inverse hypergeometric distribution has a number of practical uses. For instance, in inspection sampling, instead of taking a sample of fixed size n from a batch of items and then accepting the batch if the observed number of defectives is less than or equal to some predetermined value c (otherwise rejecting), a form of "curtailed sampling" can be adopted (see, e.g., Guenther, 1969). Here items are drawn one at a time until either $c + 1$ defectives are observed (at which point the batch is rejected) or $n - c$ nondefectives are observed (and the batch is accepted). The number of observations required to reach a decision then has an inverse hypergeometric distribution.

The use of inverse hypergeometric sampling to estimate the size of a biological population has been studied by Bailey (1951) and Chapman (1952a). A sample of k individuals is caught, marked, and released; afterward individuals are sampled one at a time until a predetermined number c of marked individuals are recaptured.

The negative hypergeometric as the distribution of the number of exceedances has been mentioned in Section 6.2.2.

Distribution-free prediction intervals can be constructed by taking two samples of sizes n_1 and n_2. Then the probability that there are w items in the second sample that are greater than the rth-order statistic of the first sample is a negative hypergeometric tail probability. Guenther (1975) gave details.

Suppose now that $X_1, X_2, \ldots, X_{2s+1}$ and $Y_1, Y_2, \ldots, Y_m$ are independent random samples from continuous distributions with cdf's $F_1(x)$ and $F_2(x)$, respectively. Then the null hypothesis H_0: $F_1(x) = F_2(x)$ for all x can be tested using the statistic V equal to the number of observations in the second sample that are less than the sample median of the first sample; see Gart (1963) and Mood (1950, pp. 395–398). The pmf of the statistic V is negative hypergeometric.

The beta–binomial model for the distribution is used very widely. Muench (1936, 1938) was interested in applications to medical trials. Skellam (1948) applied the model to the association of chromosomes and to traffic clusters. Barnard (1954), Hopkins (1955), and Hald (1960) recognized its importance in inspection sampling. Irwin (1954) and Griffiths (1973) used it for disease incidence. Kemp and Kemp (1956b) demonstrated its relevance for point quadrat data. Ishii and Hayakawa (1960) examined it as a model for the sex composition of families and for absences of students. Chatfield and Goodhardt (1970) have used it for analyzing market purchases, Williams (1975) and Haseman and Kupper (1979) have employed it in toxicology, and Crowder (1978) has used it for seed germination data.

The distribution has also arisen in a Bayesian context from binomial sampling with a beta prior. An early discussion is in Pearson (1925). Guenther (1971) showed that the average cost per lot for the Hald linear cost model with a beta prior has a beta–binomial distribution. The beta–binomial distribution has itself been used as a prior distribution by Steck and Zimmer (1968).

Consider the effect of a vector of covariates $\mathbf{x}_j = 1, 2, \ldots, n$ on the number of successes in a series of N independent Bernoulli trials. If there is little evidence of overdispersion, such data are usually analyzed using logistic regression. Here

$$\Pr[Y = y_j | p_j] = \binom{N_j}{y_j} p_j^{y_j} (1 - p_j)^{N_j - y_j},$$
$$y_j = 0, 1, \ldots, N_j, \quad j = 1, \ldots, n,$$

where $\sum_{j=1}^{n} N_j = N$ and p_j is postulated to depend on the vector $\mathbf{x}_j$ of covariates via the logistic function

$$\ln\left(\frac{p_j}{1 - p_j}\right) = \boldsymbol{\beta}^T \mathbf{x}, \qquad j = 1, 2, \ldots, n.$$

When data of this kind are overdispersed, a beta–binomial regression model is a common choice. Here

$$\Pr[Y = y_j | a_j, b_j] = \binom{N_j}{y_j} \frac{B(a_j + y_j, b_j + N_j - y_j)}{B(a_j, b_j)},$$

where $y_j = 0, 1, \ldots, N_j$, $j = 1, \ldots, n$, and $\sum_{j=1}^{n} N_j = N$. We have

$$E(Y_j) = N_j\theta_j \qquad \text{and} \qquad \text{Var}(Y_j) = N_j\theta_j(1-\theta_j)[1-(N_j-1)\lambda_j]$$

where $\theta_j = a_j/(a_j+b_j)$ and $\lambda_j = 1/(a_j+b_j+1)$. Parametric forms can be postulated for θ_j and λ_j (McLachlan and Peel, 2000). Examples of fitted beta–binomial regression models are in Griffiths (1973) (diseases in households), Williams (1975) (malformed fetuses), Prentice (1986) (aberrant chromosomal cells), Lindsey (1995) (faults in rolls of fabric), Brooks et al. (1997) (dead fetuses), and Collett (2003) (seed germination).

The usefulness of this distribution has stimulated an extensive literature. For a classified bibliography see Patil, Boswell, Joshi, and Ratnaparkhi (1984). For a more recent unclassified list of references see Wimmer and Altmann (1999).

6.9.3 Beta–Negative Binomial Distribution: Beta–Pascal Distribution, Generalized Waring Distribution

In-depth studies of the application of the generalized Waring distribution to accident theory have been made by Irwin (1968) and Xekalaki (1983b). Irwin derived the distribution as the distribution of accidents in an accident-prone community exposed to variable risk, whereas Xekalaki's two derivations were based on a "contagion" hypothesis and on a "spells" hypothesis.

In a Bayesian context the beta–negative binomial distribution can arise from negative binomial sampling with a beta prior.

Xekalaki has studied the generalized Waring distribution in depth; see Xekalaki (1981, 1983a,b,c, 1985) and Panaretos and Xekalaki (1986a). Further references are in Wimmer and Altmann (1999).

6.10 SPECIAL CASES

6.10.1 Discrete Rectangular Distribution

The *discrete rectangular distribution* (sometimes called the *discrete uniform distribution*) is defined in its most general form by

$$\Pr[X = a + xh] = \frac{1}{n+1}, \qquad x = 0, 1, 2, \ldots, n. \tag{6.96}$$

Various standard forms are in use. One that is frequently used is obtained by putting $a = 0$, $h = 1$, so that the values taken by X are $0, 1, \ldots, n$. Other forms can be obtained from this standard form by a linear transformation.

The pgf corresponding to $\Pr[X = x] = 1/(n+1)$, $x = 0, 1, \ldots, n$, is

$$G(z) = (n+1)^{-1}(1 + z + z^2 + \cdots + z^n) \tag{6.97}$$

$$= (n+1)^{-1}\,{}_2F_1[-n, 1; -n; z], \tag{6.98}$$

showing that the distribution is a special case of the negative (inverse) hypergeometric (beta–binomial) distribution.

From (6.97) the mgf is

$$G(e^t) = \frac{1 + e^t + e^{2t} + \cdots + e^{nt}}{n+1}, \tag{6.99}$$

whence

$$\mu_r{}' = \sum_{j=1}^{n} \frac{j^r}{n+1}. \tag{6.100}$$

The cf is

$$G(e^{it}) = \frac{e^{(n+1)it} - 1}{(n+1)(e^{it} - 1)}, \tag{6.101}$$

and the central mgf is

$$e^{-\mu t} G(e^t) = \frac{e^{-nt/2}(e^{(n+1)t} - 1)}{(n+1)(e^t - 1)} = \frac{\sinh(nt/2)}{(n+1)\sinh(t/2)}, \tag{6.102}$$

whence

$$\begin{aligned} \mu_{2r+1} &= 0, \\ \mu_{2r} &= (n+1)^{-1} \sum_{j=0}^{n} \left(j - \frac{n}{2}\right)^{2r}, \qquad r = 1, 2, \ldots. \end{aligned} \tag{6.103}$$

The factorial moment generating function is

$$\begin{aligned} G(1+t) &= \frac{(1+t)^{n+1} - 1}{(n+1)t} = {}_2F_1[-n, 1; 2; -t] \\ &= \sum_{r=0}^{n} \frac{n!}{(n-r)!(r+1)} \times \frac{t^r}{r!}. \end{aligned} \tag{6.104}$$

The first four moments are

$$\mu = \frac{n}{2}, \qquad \mu_2 = \frac{n(n+2)}{12}, \qquad \mu_3 = 0, \qquad \mu_4 = \frac{n(n+2)(3n^2 + 6n - 4)}{240}. \tag{6.105}$$

The cumulants of the distribution are

$$\begin{aligned} \kappa_{2r+1} &= 0, \\ \kappa_{2r} &= (2r)^{-1}[(n+1)^{2r} - 1]B_{2r}, \qquad r = 1, 2, \ldots, \end{aligned} \tag{6.106}$$

where the B_{2r} are the Bernoulli numbers (Section 1.1.9). The distribution has no mode. The median is $n/2$ for n even, and it lies between $(n-1)/2$ and $(n+1)/2$ for n odd. The coefficient of variation is $[(n+2)/(3n)]^{1/2}$.

The hazard rate (failure rate) is increasing, with $r_x = 1/(n+1-x)$ (Patel, 1973).

The beta–binomial model for the discrete rectangular distribution involves a beta distribution with parameters $\alpha = 1$, $\beta = 1$, that is, a uniform distribution on the interval [0, 1]. The somewhat surprising implication is that a continuous uniform mixture of binomials gives rise to a discrete rectangular (discrete uniform) distribution.

The exceedance model is as follows: Whatever the distribution of Z, in a future random sample of size n of values of Z, the number of values that exceed a previously observed value has a rectangular distribution.

Irwin's (1954) urn model (a sample of fixed size with additional replacements) implies that the distribution arises if a fixed number of balls are drawn from an urn initially containing one white ball and one black ball, provided that each drawn ball is replaced together with another similar ball.

The inverse sampling model is often presented as a problem concerning keys. If a key ring contains one correct and N incorrect keys, then the number of unsuccessful keys that must be tried in order to find the correct key is a discrete rectangular rv, assuming that no key is tried more than once.

The largest order statistic T in a sample of size k from a discrete rectangular distribution with support $0, 1, \ldots, n$ is a complete and sufficient statistic for n. Also

$$T^* = \frac{T^{k+1} - (T-1)^{k+1}}{T^k - (T-1)^k} \tag{6.107}$$

is the uniformly minimum variance unbiased estimator for n.

Patil and Seshadri's (1964) general characterization result for discrete distributions implies that iff the conditional distribution of X given $X + Y$ is discrete rectangular, then X and Y have identical geometric distributions.

The following authors have also studied the discrete rectangular distribution: Srivastava and Kashyap (1982) (in a queueing context), Nagaraja (1988b) (in a study of order statistics for discrete distributions), and Nair and Hitha (1989) (in a partial-sums context).

Ericksen (2002) has investigated the sum X_k of k iid discrete rectangular variables. From (6.97) the pgf for X_k is

$$G_k(z;n) = \frac{(1-z^{n+1})^k}{(n+1)^k(1-z)^k} = \sum_{j=0}^{nk} \frac{C_j(k, n+1)z^j}{(n+1)^k}, \tag{6.108}$$

where $C_j(k, n+1)$ are the Pascal–De Moivre coefficients. Ericksen's expression for $C_j(k, n+1)$ is

$$C_j(k, n+1) = \sum_{i=0}^{[j/(n+1)]} (-1)^i \binom{j - ni - i + k - 1}{k-1}\binom{k}{i}$$

where $[u]$ is the largest integer not exceeding u. The distribution is symmetric about $nk/2$, so it is unnecessary to compute $\Pr[X_k = j]$ beyond $j = [nk/2]$. The cumulants of a sum of independent rv's are equal to the sums of the individual cumulants. The cumulants of X_k are therefore

$$\kappa_{2r} = \frac{k[(n+1)^{2r} - 1]B_{2r}}{2r}, \qquad \kappa_{2r+1} = 0, \qquad r = 1, 2, \ldots,$$

and

$$\mu = \tfrac{1}{2}kn, \qquad \mu_2 = \kappa_2 = \tfrac{1}{12}kn(n+2). \tag{6.109}$$

The special case $n = 1$ is the symmetric binomial distribution. Merlini, Sprugnoli, and Verri (2002) have studied the trinomial case, $n = 2$, for which the pgf is $G_2(z) = (1 + z + z^2)^k/3^k$.

6.10.2 Distribution of Leads in Coin Tossing

The *distribution of leads in coin tossing* (Chung and Feller, 1949; Feller, 1957, p. 77) has the pmf

$$\Pr[X = x] = \frac{(2x)!(2n-2x)!2^{-2n}}{x!x!(n-x)!(n-x)!}, \qquad x = 0, 1, 2, \ldots, n. \tag{6.110}$$

Use of Legendre's duplication formula (Section 1.1.2) gives

$$\begin{aligned}
\Pr[X = x] &= \frac{\left(-\frac{1}{2}\right)!\left(-\frac{1}{2}\right)!(-1)^n}{\left(-x-\frac{1}{2}\right)!\left(-n+x-\frac{1}{2}\right)!(n-x)!x!} \\
&= \binom{n}{x}\binom{-n-1}{-\frac{1}{2}-x}\Big/\binom{-1}{-\frac{1}{2}} \\
&= \binom{-\frac{1}{2}}{x}\binom{-\frac{1}{2}}{n-x}\Big/\binom{-1}{n}, \qquad x = 0, 1, 2, \ldots, n,
\end{aligned} \tag{6.111}$$

that is, the particular negative hypergeometric distribution with pgf

$$G(z) = \frac{\left(n-\frac{1}{2}\right)!}{n!\left(-\frac{1}{2}\right)!}\, {}_2F_1\left[-n, \tfrac{1}{2}; -n+\tfrac{1}{2}; z\right] \tag{6.112}$$

(Kemp, 1968b). We find that

$$\frac{\Pr[X = x] - \Pr[X = x-1]}{\Pr[X = x-1]} = \frac{x - \frac{1}{2}(n+1)}{x\left(n - x + \frac{1}{2}\right)}. \tag{6.113}$$

Hence $\Pr[X = x] \gtrless \Pr[X = x - 1]$ according as $x \gtrless \frac{1}{2}(n+1)$; that is, the distribution is U shaped. The mean and variance are

$$\mu = \tfrac{1}{2}n \qquad \text{and} \qquad \mu_2 = \tfrac{1}{8}n(n+1); \tag{6.114}$$

the mean is the same as that for a discrete rectangular distribution with the same support but the variance is always greater.

Feller (1950, p. 71) has said, "The picturesque language of gambling should not detract from the general importance of the coin-tossing model. In fact, the model may serve as a first approximation to many more complicated chance-dependent processes in physics, economics, and learning-theory." He elaborated on this theme.

6.10.3 Yule Distribution

The *Yule distribution* was developed by Yule (1925) as a descriptor of the number of species of biological organisms per family using a mixture of shifted geometric distributions with pmf

$$e^{-\beta u}(1 - e^{-\beta u})^{y-1}, \qquad y = 1, 2, \ldots, \tag{6.115}$$

where u has an exponential distribution with pdf

$$f(u) = \theta^{-1}\exp\left(-\frac{u}{\theta}\right), \qquad \theta > 0, \qquad 0 \le u. \tag{6.116}$$

The resultant mixture distribution can be represented as

$$\text{Geometric}(e^{-\beta U}) \bigwedge_{U} \text{Exponential}(\theta);$$

the probabilities are

$$\Pr[Y = y] = \int_0^\infty e^{-\beta u}(1 - e^{-\beta u})^{y-1}\theta^{-1}e^{-u/\theta}\,du. \tag{6.117}$$

Putting $v = e^{-\beta u}$ and $\rho = (\beta\theta)^{-1}$ (so that $\rho > 0$), this becomes

$$\begin{aligned}\Pr[Y = y] &= \int_0^1 (1 - v)^{y-1}v^{\rho}\rho\,dv \\ &= \frac{\rho(\rho!)(y-1)!}{(y+\rho)!}, \qquad y = 1, 2, \ldots.\end{aligned} \tag{6.118}$$

The pgf is

$$H(z) = \frac{\rho\ z}{\rho + 1}\,{}_2F_1[1, 1; \rho + 2; z], \qquad \rho > 0, \tag{6.119}$$

showing that the distribution is a special case of a beta–negative binomial distribution shifted to support $1, 2, \ldots$.

The integral in (6.118) demonstrates that the distribution is also the outcome of a beta mixture of shifted geometric distributions with pmf $v(1-v)^{y-1}$, $y = 1, 2, \ldots$; this was the mode of genesis underlying its use by Miller (1961) for the distribution of sizes of traffic clusters and by Pielou (1962) for runs of plant species.

The mean and variance of the Yule distribution are

$$\begin{aligned} \mu &= \frac{\rho}{\rho-1} && \text{provided that } \rho > 1, \\ \mu_2 &= \frac{\rho^2}{(\rho-1)^2(\rho-2)} && \text{provided that } \rho > 2. \end{aligned} \tag{6.120}$$

The name Yule has also been used for the distribution when it is shifted to the support $0, 1, \ldots$, that is, with pmf

$$\Pr[X = x] = \frac{\rho(\rho!)x!}{(x+\rho+1)!}, \qquad x = 0, 1, \ldots. \tag{6.121}$$

The corresponding pgf is

$$G(z) = \frac{\rho}{\rho+1}\, {}_2F_1[1, 1; \rho+2; z]; \tag{6.122}$$

the mean is now

$$\mu = \frac{1}{\rho-1} \quad \text{provided that } \; \rho > 1 \tag{6.123}$$

(the variance is of course unchanged). More generally, the rth descending factorial moment for (6.121) is

$$\mu'_{[r]} = \frac{r!r!(\rho-r-1)!}{(\rho-1)!} \quad \text{provided that } \rho > r, \tag{6.124}$$

and the rth ascending factorial moment is

$$E[X(X+1)\cdots(X+r-1)] = \frac{r!\rho}{(\rho-r)(\rho-r+1)} \quad \text{provided that } \rho > r. \tag{6.125}$$

The Yule distribution has been used to model word frequency data by Simon (1955, 1960) and Haight (1966). Herdan (1961), however, was doubtful about the fit in the upper tail. Mandelbrot (1959) was especially critical about the use of values of ρ less than unity (i.e., the use of distributions with an infinite mean).

In his paper on natural law in the social sciences, Kendall (1961) applied the distribution to certain kinds of bibliographic data; see also Kendall (1960).

Haight (1966) fitted the logarithmic, Yule, and Borel (see Section 9.11) distributions to four distributions of responses in psychological tests. In each case three methods of fitting were used:

1. Equating the sample and population means.
2. Equating the sample and population first frequencies.
3. Equating sample and population tail frequencies.

He found that generally the Yule distribution gave the best fit and the logarithmic distribution the worst fit. Furthermore, fitting by moments (method 1) generally gave a worse fit, for a given distribution, than did methods 2 and 3.

Xekalaki (1984) has shown how the distribution can arise in an econometric context. In Xekalaki (1983c) she applied it to an income underreporting model and to an inventory control problem. The applications in the latter paper involved the following result: Let X and Y have discrete distributions on the nonnegative integers with the property that

$$\Pr[Y = y] = c^{-1} \sum_{x \geq y+1} \frac{\Pr[X = x]}{x}, \tag{6.126}$$

where $c = 1 - \Pr[X = 0]$. Then $\Pr[Y = y] = \Pr[X = x]$ iff X has a Yule distribution with pmf (6.118). Singh and Vasudeva (1984) characterized the exponential distribution via the Yule distribution. Xekalaki and Dimaki (2004) gave characterizations of the Yule distribution based on reliability measures.

The Yule distribution can be regarded as a discrete analog of the Pareto distribution; see Xekalaki and Panaretos (1988). Also, when $\rho = 1$, it can be regarded as an approximation to the zeta distribution (Section 11.20). Panaretos (1989a,b) has given an application to surname frequency data.

Prasad (1957) has described a generalized form of the Yule distribution with the following pmf:

$$\Pr[Y = y] = \frac{2\lambda(\lambda + 1)}{(\lambda + y - 1)(\lambda + y)(\lambda + y + 1)}, \qquad y = 1, 2, \ldots, \tag{6.127}$$

where $\lambda > 0$. The pgf is

$$H(z) = \frac{2z}{\lambda + 2}\, {}_2F_1[1, \lambda; \lambda + 3; z]. \tag{6.128}$$

(Putting $\lambda = 1$, we obtain a Yule distribution with $\rho = 2$.) The expected value of Y is $1 + \lambda$; the second and higher moments are, however, infinite. The cdf is

$$\Pr[Y \leq y] = 1 - \frac{\lambda(\lambda + 1)}{(\lambda + y)(\lambda + y + 1)}. \tag{6.129}$$

6.10.4 Waring Distribution

The *Waring distribution* is a generalization of the Yule distribution that was developed by Irwin (1963) using the Waring expansion

$$\frac{1}{c - a} = \frac{1}{c} + \frac{a}{c(c + 1)} + \frac{a(a + 1)}{c(c + 1)(c + 2)} + \cdots; \tag{6.130}$$

this converges for $c > a$. Taking $\Pr[X = x]$ proportional to the $(x + 1)$th term in the series gives

$$\Pr[X = x] = \frac{(c-a)(a+x-1)!c!}{c(a-1)!(c+x)!}, \qquad x = 0, 1, 2, \ldots; \tag{6.131}$$

the corresponding pgf is

$$G(z) = \frac{c-a}{c} {}_2F_1[1, a; c+1; z]. \tag{6.132}$$

The Yule is the special case $a = 1$.

The Waring distribution is the special case $k = 1$ of the generalized Waring distribution (see Section 6.2.3); its moments and properties follow therefrom. In particular

$$\begin{aligned} \mu &= \frac{a}{c-a-1} && \text{provided } c - a > 1, \\ \mu_2 &= \frac{a(c-a)(c-1)}{(c-a-1)^2(c-a-2)} && \text{provided } c - a > 2. \end{aligned} \tag{6.133}$$

Irwin (1963, p. 30) commented that the series converges much more slowly than the geometric distribution and said that "we can make the tail as long as we please, as c and $a \to 0$." He demonstrated how the distribution can be fitted, both by the mean-and-zero-frequency method and by maximum likelihood, using one of Kendall's (1961) bibliographic data sets.

The distribution arises as a special case of the beta–negative binomial distribution in two ways. First,

$$G(z) = \int_0^1 \left(\frac{1-\lambda}{1-\lambda z}\right) \frac{\lambda^{a-1}(1-\lambda)^{c-a-1}\, d\lambda}{B(a, c-a)}, \tag{6.134}$$

showing that it is a mixture of geometric distributions (Miller, 1961; Pielou, 1962; Weinberg and Gladen, 1986).

Second,

$$G(z) = \int_0^1 \left(\frac{1-\lambda}{1-\lambda z}\right)^a \frac{(1-\lambda)^{c-a-1}\, d\lambda}{B(1, c-a)}; \tag{6.135}$$

this demonstrates that it is also a mixture of negative binomial distributions using a particular type of beta mixing distribution.

The hazard function of the Waring distribution has been investigated by Xekalaki (1983d).

Marlow's (1965) *factorial distribution* has the pgf

$$G(z) = \frac{{}_2F_1[1, \lambda - n + 1; \lambda + 1; z]}{{}_2F_1[1, \lambda - n + 1; \lambda + 1; 1]}, \tag{6.136}$$

where $n = 2, 3, 4, \ldots$ and $n - 1 < \lambda$, and hence it is a Waring distribution with $a = \lambda - n + 1, c = \lambda$.

The Salvia and Bollinger (1982) distribution with pgf

$$G(z) = \frac{{}_2F_1[1, 1-\alpha; 2; z]}{{}_2F_1[1, 1-\alpha; 2; 1]}, \qquad 0 < \alpha < 1, \tag{6.137}$$

is another Waring distribution. It is a special case of Engen's extended negative binomial distribution; see Section 5.12.2.

Smith and Diaconis (1988) have studied the Waring distribution with $a = \frac{1}{2}$, $c = 1$. This is called the Feller–Shreve distribution by Wimmer and Altmann (1999); see McNally (1990) for an application to fecundability in cows and Altmann (1993) for an application to phoneme counts.

Another generalization of the Yule distribution appears in Johnson and Kotz (1989). These authors extended Singh and Vasudeva's (1984) characterization theorem of the exponential distribution by the Yule distribution as follows: If X and Y have the same (nonnegative) support and

$$\Pr[Z = k|X = t] = \Pr[Z = k|Y = t] = g(t)\{h(t)\}^k,$$

$(g(t),\ h(t) > 0)$ for all $k = 0, 1, 2, \ldots$ and all t in the common support of X and Y, and $h(t)$ is a strictly monotonic function of t, then X and Y have identical distributions. If Z takes only the values $0, 1, 2, \ldots$, then $g(t) = 1 - h(t)$.

When $h(t) = t/(1+t)$ and the density function of X is

$$f_X(t) = \frac{1}{B(\alpha, \beta)} \cdot \frac{t^{\beta-1}}{(1+t)^{\alpha+\beta}}, \qquad 0 < t, \qquad \alpha > 0, \qquad \beta > 0,$$

then

$$\Pr[Z = k] = \frac{B(\alpha+1, \beta+k)}{B(\alpha, \beta)}, \qquad k = 0, 1, \ldots.$$

This is a Waring distribution with pgf $\alpha(\alpha+\beta)^{-1}\,{}_2F_1[1, \beta; \alpha+\beta+1; z]$. The Yule distribution is the special case, $\beta = 1$. If $h(t) = [t/(1+t)]^\gamma$, $\gamma > 0$, $t > 0$, with X having the same density function as above, then

$$\Pr[Z = k] = \frac{B(\alpha, k\gamma+\beta) - B(\alpha, k\gamma+\gamma+\beta)}{B(\alpha, \beta)}, \qquad k = 0, 1, \ldots.$$

Johnson and Kotz suggested the name "generalized" Yule distribution.

6.10.5 Narayana Distribution

The *Narayana distribution* gives the proportions of paths, from $(0, 0)$ to (n, n) in a two-dimensional lattice, that have X turns, where the steps are $(i, j) \to (i, j+1)$ and $(i, j) \to (i+1, j)$ and the paths do not cross the diagonal from $(0, 0)$ to (n, n); see Narayana (1959, 1979).

The number of such paths from $(0, 0)$ to (n, n) is given by the Narayana number

$$N_{n,x} = \frac{1}{n}\binom{n}{x-1}\binom{n}{x}, \qquad x = 1, 2, \ldots, n,$$

and

$$\sum_{x=1}^{n} N_{n,x} = \frac{1}{n+1}\binom{2n}{n} = C_n,$$

where C_n is a Catalan number. The distribution therefore has the pmf

$$\Pr[X = x] = \frac{(n-1)!n!n!(n+1)!}{(x-1)!x!(n-x)!(n-x+1)!(2n)!}, \qquad x = 1, 2, \ldots, n;$$

that is,

$$\Pr[Y = y] = \frac{(n-1)!n!n!(n+1)!}{y!(y+1)!(n-y-1)!(n-y)!(2n)!}, \qquad y = 0, 1, \ldots, n-1,$$

where $Y = X - 1$. The pgf for X can be expressed as

$$G_X(z) = \frac{z\,n!(n+1)!}{(2n)!}\,{}_2F_1[-n, 1-n; 2; z],$$

showing that the distribution is a classical hypergeometric distribution with $N = 2n$ and $Np = (n-1)$ shifted to support $1, 2, \ldots, n$. Thus the mean and variance are

$$\mu = \frac{n-1}{2} + 1 \qquad \text{and} \qquad \mu_2 = \frac{(n-1)(n+1)}{4(2n-1)}.$$

The distribution is unimodal and symmetric.

Stanley (1999, Examples 6.19 and 6.36) listed, with illustrations, 66 sets of combinatorial objects which have Catalan numbers of elements; these are called Catalan structures. He stated that in each case there is a decomposition into subsets counted by the Narayana numbers. Sulanke (1998, 1999) catalogued over 200 Narayana rv's; see also Sulanke (2002).

Sulanke (2002) also examined the Kirkman (1857) distribution in a combinatorial context by relating it to the numbers of dissections of an $(n + 2)$-sided convex polygon with a distinguished base into r interior regions [as in Stanley (1999, Example 6.33c)]. Here the pgf is

$$\begin{aligned} G^*(z) &= \sum_{j=0}^{n-1} \frac{1}{n}\binom{n}{j}\binom{2n-j}{n+1} z^j \Bigg/ \sum_{j=0}^{n-1} \frac{1}{n}\binom{n}{j}\binom{2n-j}{n+1} \\ &= \sum_{k=1}^{n} N_{n,k}(1+z)^{k-1} \Bigg/ \sum_{k=1}^{n} N_{n,k} 2^{k-1}, \end{aligned}$$

where $N_{n,k}$ are the Narayana numbers. The distribution can easily be shown to be unimodal; it is not symmetric. Sulanke (2002) provided a useful list of references for both distributions.

6.11 RELATED DISTRIBUTIONS

6.11.1 Extended Hypergeometric Distributions

The name "extended" when applied to a hypergeometric distribution with pgf of the form ${}_pF_q[z]/{}_pF_q[1]$ refers to a distribution with pgf of the form

$$G(z) = \frac{{}_pF_q[\lambda z]}{{}_pF_q[\lambda]} = \frac{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda]}, \tag{6.138}$$

where $\lambda \neq 1$.

Harkness (1965) used the term *extended hypergeometric distribution* in this way to describe the conditional distribution of one of two binomial rv's, given that their sum is fixed. The distribution had arisen in an investigation of the power function for Fisher's (1934) exact test of independence in a 2×2 contingency table. If X_i has parameters n_i and $p_i = 1 - q_i$ for $i = 1, 2$, then

$$\begin{aligned}
&\Pr[X_1 = x | X_1 + X_2 = m] \\
&\quad = \binom{n_1}{x}\binom{n_2}{m-x}\left(\frac{p_1q_2}{q_1p_2}\right)^x \Bigg/ \left[\sum_x \binom{n_1}{x}\binom{n_2}{m-x}\left(\frac{p_1q_2}{q_1p_2}\right)^x\right] &(6.139)\\
&\quad = \binom{n_1}{x}\binom{n_2}{m-x}\theta^x \Bigg/ \binom{n_2}{m} {}_2F_1[-n_1, -m; n_2 + 1 - m; \theta], &(6.140)
\end{aligned}$$

where $\theta = p_1q_2/(q_1p_2)$ and $\max(0, m - n_2) \leq x \leq \min(n_1, m)$; the same limits apply to the summation in (6.139).

This is a four-parameter extension of the classical hypergeometric distribution of Section 6.2.1; the classical hypergeometric is of course the outcome when $p_1 = p_2$ (and $q_1 = q_2$). Unlike the classical hypergeometric, however, it is a GPSD (Section 2.2.1). The pgf is

$$G(z) = \frac{{}_2F_1[-n_1, -m; n_2 + 1 - m; \theta z]}{{}_2F_1[-n_1, -m; n_2 + 1 - m; \theta]}; \tag{6.141}$$

the distribution is therefore also GHPD. Its properties can be obtained as in Section 2.4.1. For example, as $n_1 \to \infty$ and $n_2 \to \infty$ with $n_1/(n_1 + n_2) = c$, c constant, it tends to a binomial distribution with parameters m and $\theta c/(\theta c + 1 - c)$. Hannan and Harkness (1963) derived a normal limiting form, and Harkness (1965) obtained a Poisson limiting form.

The rth factorial moment of the distribution is

$$\mu'_{[r]} = \frac{n_1!m!(n_2-m)!\ {}_2F_1[r-n_1, r-m; r+n_2+1-m; \theta]}{(n_1-r)!(m-r)!(n_2-m+r)!\ {}_2F_1[-n_1, -m; n_2+1-m; \theta]}; \tag{6.142}$$

in particular

$$\mu = \frac{n_1 m\ {}_2F_1[1-n_1, 1-m; n_2+2-m; \theta]}{(n_2-m+1)\ {}_2F_1[-n_1, -m; n_2+1-m; \theta]}. \tag{6.143}$$

There are no simple explicit expressions for the moments, although there are recurrence relations such as

$$(1-\theta)\mu'_2 = n_1 m\theta - [n_1+n_2-(n_1+m)(1-\theta)]\mu. \tag{6.144}$$

Harkness (1965) studied maximum-likelihood estimation for θ, assuming that the other three parameters n_1, n_2, and m are known. Let $y_1, y_2, \ldots, y_N$ be a sample of N independent observations from distribution (6.139). Then the MLE of θ satisfies the equation $\overline{y} = \mu$; this does not have an explicit solution. Using the "natural" estimator

$$\tilde{\theta} = \frac{\overline{y}(n_2-m+\overline{y})}{(n_1-\overline{y})(m-\overline{y})}, \tag{6.145}$$

Harkness obtained lower and upper bounds for $\hat{\theta}$; these are given in the second edition of this book.

There has been much biostatistical interest in the properties of the log-odds ratios in 2×2 tables. This has led to a long series of papers involving the extended (classical) hypergeometric distribution, many in the journal *Biometrics*. They include Heilbron (1981), Hauck (1984), Gastwirth and Greenhouse (1987), Holland (1989), Sato (1990), Satten and Kupper (1990), Vollset, Hirji, and Elashoff (1991), and Fu and Arnold (1992).

The extended (classical) hypergeometric distribution with pgf of the form (6.141) is also the outcome when the weighting factor $w_x = \theta^x$ is applied to the classical hypergeometric distribution.

Seneta (1994) has discussed the work of Liebermeister (1877) on the problem of inferring equality or nonequality of the success probabilities in two series of binomial trials. Liebermeister used a Bayesian-type argument to obtain a precursor of Fisher's exact test. The cdf is similar to the cdf for Fisher's test but has slightly different parameters.

Applying the weighting factor $w_x = \theta^x$ to the negative hypergeometric distribution gives the *extended negative hypergeometric distribution* with pgf

$$G(z) = \frac{{}_2F_1[-n, v+l; -w-n; \theta z]}{{}_2F_1[-n, v+1; -w-n; \theta]}. \tag{6.146}$$

Replacement of both of the binomial rv's in (6.139) by negative binomial rv's also leads to this distribution. The negative binomial distribution truncated above $n > 0$ has the pgf (6.146) with $w = 0$.

The pgf of the doubly truncated geometric distribution with support $L, L+1, \ldots, R-1, R$, when shifted to support $0, 1, \ldots$, has the form

$$G(z) = \frac{{}_2F_1[L-R, 1; L-R; \theta z]}{{}_2F_1[L-R, 1; L-r; \theta]}. \tag{6.147}$$

The solution to Banach's matchbox problem (Feller, 1968) has the pmf

$$\Pr[X = x] = \binom{2N-r}{N-r} 2^{r-2n},$$

where N is the number of matches in each matchbox. The pgf is

$$G(z) = \binom{2N}{N} 2^{-2n} \, {}_2F_1[-N, 1; -2N; 2z]. \tag{6.148}$$

Morton (1991) has examined the use of an extended negative hypergeometric distribution in the analysis of data with extra-multinomial variation.

The *extended generalized Waring distribution* arises when the generalized Waring probabilities are weighted using the weighting factor $w_x = \theta^x$. The pgf is

$$G(z) = \frac{{}_2F_1[k, \ell; k+\ell+m; \theta z]}{{}_2F_1[k, \ell; k+\ell+m; \theta]}. \tag{6.149}$$

Because this distribution has infinite support, an important restriction on the parameters is $0 < \theta < 1$; also $k > 0$, $\ell > 0$ are needed, but it is no longer necessary for m to be positive.

Taking $k = \ell = 1$, $m = 0$ gives a logarithmic distribution (Chapter 5) shifted to the origin; for a left-truncated logarithmic distribution shifted to the origin, $k = 1$, $\ell = L$, $m = 0$.

Taking $k = \ell = m = 1$ gives Good's (1953) species frequency distribution; taking $k = \ell = 1$, $m = n - 1$ gives the *hyperlogarithmic distribution* of Tripathi and Gurland (1977), Tripathi (1983), Tripathi and Gupta (1985), and Tripathi, Gupta, and White (1987); and taking $k = 1$, $\ell = 2$, $m = a - 2$ gives the generalized geometric distribution of Tripathi and Gupta (1987) and Tripathi, Gupta, and White (1987).

A negative value of m occurs when $k = \ell = 2$, $m = -3$ [the *Moran–Gani distribution* of Moran (1955) and Gani (1957)].

Engen's (1974, 1978) extended negative binomial distribution (Section 5.12.2) has the form of a zero-truncated (positive) negative binomial distribution. When shifted to support $0, 1, \ldots$, it becomes an extended Waring distribution with pgf

$$G(z) = \frac{{}_2F_1[1; \kappa + 1; 2; \theta z]}{(1-\theta)^{-\kappa} - 1}, \tag{6.150}$$

where κ is the negative binomial exponent. Engen was able to relax the negative binomial restriction $\kappa > 0$ to $\kappa > -1$. The special case $-1 < k < 0$, $\theta = 1$, is the Salvia and Bollinger (1982) distribution (6.137); see also Section 5.12.2.

Further examples of distributions with pgf's of the form $C\,{}_2F_1[a, b; c; \theta z]$ are the distributions of X and of Y, given $X + Y = c$, where X is a binomial rv, Y is a negative binomial rv, and c is a fixed positive integer. The *positive binomial* and the *hyperbinomial distributions* of Shanmugam (1986) and the right-truncated binomial are other instances.

It is easy to construct various ad hoc distributions with pgf's of the form $C\,{}_2F_1[a, b; c; \theta z]$. However, fitting a four-parameter distribution such as this may be troublesome—a small change in the estimate of one parameter may lead to big changes in the estimates of the other parameters. When a statistical model leads to a distribution of this kind, however, very often one or more of the parameters will be known (as in the Harkness model above). The lost-games distribution (Section 11.2.10) is an example of an extended generalized Waring distribution that arises naturally and has only two parameters that need to be estimated.

6.11.2 Generalized Hypergeometric Probability Distributions

This section deals with certain distributions with pgf's of the form

$$G(z) = z^k \frac{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda z]}{{}_pF_q[a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; \lambda]},$$

where $p = 2$, $q = 1$ do not hold simultaneously and $k = 0, 1$. General comments about such GHPDs are in Section 2.4.1.

The *positive hypergeometric distribution* is formed from the classical hypergeometric distribution by omitting the zero class. (If n exceeds $N - Np$, then there is no zero class.) If $n \le N - Np$, then the pmf is

$$\Pr[X = x] = \binom{Np}{x}\binom{N - Np}{n - x} \Bigg/ \left[\binom{N}{n} - \binom{N - Np}{n}\right], \qquad x = 1, 2, \ldots, \min(n, Np). \tag{6.151}$$

The pgf is

$$G(z) = z\frac{{}_3F_2[1 - n, 1 - Np, 1; 2, N - Np - n + 2; z]}{{}_3F_2[1 - n, 1 - Np, 1; 2, N - Np - n + 2; 1]} \tag{6.152}$$

$(n, Np, N \in \mathbb{Z}^+)$. Govindarajulu (1962) has made a detailed study of the inverse moments of this distribution and has tabulated values of $E(X^{-r})$ for $r = 1,\ 2$.

For the positive (i.e., zero-truncated) negative hypergeometric distribution the pgf is

$$G(z) = z\frac{{}_3F_2[1-n, 1+M, 1; 2, M-n-K+2; z]}{{}_3F_2[1-n, 1+M, 1; 2, M-n-K+2; 1]},$$
$$K > M > 0, \qquad n \in \mathbb{Z}^+. \tag{6.153}$$

Good's (1957) study of word frequencies included consideration of the distribution with pgf

$$G(z) = z\frac{{}_3F_2[1-n, a+1, 1; 1-n, a+2; z]}{{}_3F_2[1-n, a+1, 1; 1-n, a+2; 1]}, \qquad a > 1, \qquad n \in \mathbb{Z}^+. \tag{6.154}$$

Darwin's (1960) ecological distribution (see Section 7.1.12) has a very similar pgf:

$$G(z) = z\frac{{}_3F_2[1-n, 1, 1; 2, 2-n-b; z]}{{}_3F_2[1-n, 1, 1; 2, 2-n-b; 1]}, \qquad b > 0, \qquad n \in \mathbb{Z}^+. \tag{6.155}$$

The pgf's for the digamma and trigamma distributions in Section 11.2.2 and for the Lotka distribution of population size [see Wimmer and Altmann (1999) for references] can also all be stated in the form $z\,{}_3F_2[z]/{}_3F_2[1]$.

Dacey (1972) obtained a distribution with pgf of the form ${}_4F_3[z]/{}_4F_3[1]$. The pgf of the Yousry–Srivastava negative hypergeometric distribution (Yousry and Srivastava, 1987) also has this form.

The remainder of this section mentions generalized hypergeometric probability distributions with argument parameter $\theta \neq 1$. Certain others that appear in Table 2.4 are dealt with elsewhere in this book.

A pgf of the form ${}_0F_1\left[-; \frac{1}{2}; \theta z\right]/{}_0F_1\left[-; \frac{1}{2}; \theta\right]$ is derived in Hall's (1983) consideration of the roles of the Bessel and Poisson distributions in chemical kinetics.

The Palm–Poisson distribution (Knessel, 1991) has the pgf

$$G(z) = \frac{{}_2F_0[1, -R; -; -\theta z]}{{}_2F_0[1, -R; -; \theta]}.$$

For Abakuks' (1979) distribution, the pgf is

$$G(z) = \frac{{}_3F_1[1, r-1, r-n-1; r; -\theta z]}{{}_3F_1[1, r-1, r-n-1; r; -\theta]};$$

for Ehrenfest's heat exchange model (Bingham, 1991) it is

$$G(z) = \frac{{}_3F_1[1, b+1, -a; b; -z/(a+b)]}{{}_3F_1[1, b+1, -a; b; -1/(a+b)]}.$$

Gutiérrez-Jáimez and Rodríguez Avi (1997) have made a detailed study of distributions with pgf's of the form

$$G(z) = \frac{{}_3F_2[\alpha_1, \alpha_2, \alpha_3; \gamma_1, \gamma_2; \lambda z]}{{}_3F_2[\alpha_1, \alpha_2, \alpha_3; \gamma_1, \gamma_2; \lambda]}.$$

6.11.3 Generalized Hypergeometric Factorial Moment Distributions

This section deals with some distributions of the type that were discussed in Section 2.4.2.

Consider a binomial distribution with parameters n and θp, where p has a beta distribution on (0, 1). This gives the following extension of the beta–binomial (≡negative hypergeometric) distribution of Section 6.2.2:

$$\begin{aligned} G(z) &= \int_0^1 (1 - \theta p + \theta p z)^n \times \frac{p^{\alpha-1}(1-p)^{\beta-1}\,dp}{B(\alpha, \beta)} \\ &= {}_2F_1[-n, \alpha; \alpha+\beta; \theta(1-z)], \qquad 0 < \theta < 1; \end{aligned} \tag{6.156}$$

this is a GHFD and hence the pmf can be shown to be

$$\begin{aligned} &\Pr[X = x] \\ &= \binom{n}{x} \frac{\theta^x (\alpha + x - 1)!(\alpha+\beta-1)!}{(\alpha-1)!(\alpha+\beta+x-1)!}\, {}_2F_1[-n + x, \alpha + x; \alpha + \beta + x; \theta], \end{aligned} \tag{6.157}$$

where $x = 0, 1, \ldots, n$.

Similarly, given a negative binomial distribution with pgf $g(z) = (1 + \theta P - \theta P z)^{-k}$, if P has a beta distribution on (0, 1), the resultant mixture has the pgf

$$\begin{aligned} G(z) &= \int_0^1 (1 + \theta P - \theta P z)^{-k} \times \frac{P^{\alpha-1}(1-P)^{\beta-1}\,dP}{B(\alpha, \beta)} \\ &= {}_2F_1[k, \alpha; \alpha+\beta; \theta(z-1)], \qquad 0 < \theta, \end{aligned} \tag{6.158}$$

and the pmf

$$\begin{aligned} \Pr[X = x] = &\binom{k + x - 1}{x} \frac{\theta^x (\alpha + x - 1)!(\alpha+\beta-1)!}{(\alpha-1)!(\alpha+\beta+x-1)!} \\ &\times {}_2F_1[k + x, \alpha + x; \alpha + \beta + x; -\theta], \end{aligned} \tag{6.159}$$

where $x = 0, 1, \ldots$. This is *not* a direct extension of the beta–negative binomial distribution. It is the type H_2 distribution of Gurland (1958), Katti (1966), and Tripathi and Gurland (1979).

The beta–Poisson distribution has the pgf $G(z) = {}_1F_1[a; a + b; \phi(z - 1)]$; see Section 8.3.3. Wimmer and Altmann (1999) call this the *Quinkert distribution*.

Taking $a = b = 1$ gives a mixture of Poisson distributions using a (continuous) uniform mixing distribution on (0, 1).

There are also matching distributions that are GHFD; see Table 2.5 and Section 10.3.

6.11.4 Other Related Distributions

The *noncentral hypergeometric distribution* is the name given by Wallenius (1963) to a distribution constructed by supposing that in sampling without replacement (as in Section 6.1) the probability of drawing a white ball given that there are Np white and $N - Np$ black balls is not p but $p/[p + \theta(1 - p)]$ with $\theta \neq 1$. The mathematical analysis following from this assumption is complicated. Starting from the recurrence relationship

$$\begin{aligned}\Pr[X = x|Np, n, N] = {} & \frac{p \Pr[X = x - 1|Np - 1, n - 1, N - 1]}{p + \theta(1 - p)} \\ & + \frac{\theta(1 - p) \Pr[X = x|Np, n - 1, N - 1]}{p + \theta(1 - p)}\end{aligned} \tag{6.160}$$

Wallenius obtained the formula

$$\Pr[X = x] = \binom{Np}{x}\binom{N - Np}{n - x} \int_0^1 (1 - t^c)^x (1 - t^{\theta c})^{n-x}\, dt \tag{6.161}$$

with $c = [Np - x + \theta(N - Np - n + x)]^{-1}$. He gave bounds for $\Pr[X = x]$, and he showed that, for n small compared with Np and $N - Np$, X is distributed approximately binomially with parameters n, $[1 + \theta p/(1 - p)]^{-1}$.

Janardan (1978) developed a *generalized Markov–Pólya* (generalized negative hypergeometric) distribution based on a voting model. The pmf is

$$\Pr[X = k] = \binom{N}{k} \frac{J_k(a, c, t) J_{N-k}(b, c, t)}{J_N(a + b, c, t)}, \tag{6.162}$$

where

$$J_k(a, c, t) = a(a + kt + c)(a + kt + 2c) \cdots (a + kt + (k - 1)c).$$

Janardan gave recurrence relations for the probabilities as well as expressions for the mean and variance; he also considered certain special and limiting cases and maximum-likelihood estimation. A generalized inverse Markov–Pólya distribution was also introduced. The use of the generalized Markov–Pólya distribution as a random damage model was studied by B. R. Rao and Janardan (1984).

Consider the distribution of the number of items observed to be defective in samples from a finite population when the detection of a defective is not certain. Johnson, Kotz, and Sorkin (1980) examined this problem in relation to audit sampling of financial accounts. If the probability of selecting y erroneous accounts is hypergeometric, that is, if

$$\Pr[Y = y] = \binom{Np}{y}\binom{N-Np}{n-y}\Big/\binom{N}{n},$$
$$\max(0, n - N + Np) \le y \le \min(n, Np),$$

and if the conditional probability of detecting x among these y erroneous accounts is

$$\binom{y}{x}\pi^x(1-\pi)^{y-x}, \qquad x = 0, 1, \ldots, y,$$

then the unconditional probability of detecting x erroneous accounts is

$$\Pr[X = x] = \sum_{y \ge x}\binom{Np}{y}\binom{N-Np}{n-y}\binom{N}{n}^{-1}\binom{y}{x}\pi^x(1-\pi)^{y-x}, \tag{6.163}$$

where $\max(0, n - N + Np) \le x \le \min(n, Np)$. Symbolically the distribution is

$$\text{Binomial}(Y, \pi) \bigwedge_{Y} \text{Hypergeometric}(n, Np, N). \tag{6.164}$$

Distributions of this kind that arise in inspection sampling with imperfect inspection have been studied in depth by Johnson and Kotz and their co-workers; a full bibliography of their work is given in Johnson, Kotz, and Wu (1991). For an exposition of the types of distributions that can arise from particular screening procedures see Johnson and Kotz (1985). Johnson, Kotz, and Rodriguez (1985, 1986) developed an *imperfect inspection hypergeometric distribution* that takes into account misclassification of nondefectives as well as defectives. This distribution can be represented symbolically as

$$\text{Binomial}(Y, p) * \text{Binomial}(n - Y, p') \bigwedge_{Y} \text{Hypergeometric } (n, D, N), \tag{6.165}$$

where the asterisk denotes the convolution operation; see Sections 1.2.11 and 3.12.3.

The pmf is

$$\Pr[X = k|n, D, N; p, p']$$
$$= \binom{N}{n}^{-1}\sum_{y}\left[\binom{y}{w}\binom{n-y}{k-w}p^w(1-p)^{y-w}(p')^{k-w}(1-p')^{n-y-k+w}\right], \tag{6.166}$$

where $\max(0, n - N + D) \le y \le \min(n, D)$ and $\max(0, k - n + y) \le w \le \min(n, D)$. Perfect inspection corresponds to $p = 1$ and $p' = 0$. Johnson, Kotz, and Rodriguez provided tables of the pmf given k and the number of apparently nonconforming items for $N = 100, 200, \infty$ and described how to interpolate for other values of N. The extension to double and multiple sampling schemes was investigated.

Hypergeometric, negative hypergeometric, and generalized Waring (Pólya and inverse Pólya) distributions of order k are discussed in Section 10.7.

The *intrinsic hypergeometric distribution* of Baldessari and Weber (1987) is the distribution of $\sum_{i=1}^{N} X_i$, where the X_i are dependent zero–one rv's.

Gurland and Tripathi (1975) have derived a distribution with pgf

$$G(z) = \frac{{}_1F_1[a; b; c + dz]}{{}_1F_1[a; b; c + d]}. \tag{6.167}$$

Kumar (2002) has investigated similar generalizations of generalized hypergeometric probability distributions. These have pgf's of the form

$$G(z) = \frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta_1 z + \theta_2 z^m]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta_1 + \theta_2]} \tag{6.168}$$

and include the Hermite, generalized Hermite, Gegenbauer, and generalized Gegenbauer distributions.

CHAPTER 7

Logarithmic and Lagrangian Distributions

7.1 LOGARITHMIC DISTRIBUTION

7.1.1 Definition

The random variable X has a *logarithmic distribution* if

$$\Pr[X = x] = \begin{cases} \dfrac{a\theta^x}{x}, & x = 1, 2, \ldots, \quad (7.1) \\[2ex] \dfrac{(x-1)\theta}{x} \Pr[X = x-1], & x = 2, 3, \ldots, \quad (7.2) \end{cases}$$

where $0 < \theta < 1$ and $a = -[\ln(1-\theta)]^{-1}$. It is a one-parameter generalized power series distribution (GPSD) with infinite support on the positive integers.

The characteristic function is

$$\varphi(t) = \frac{\ln(1-\theta e^{it})}{\ln(1-\theta)}, \tag{7.3}$$

and the probability generating function (pgf) is

$$G(z) = \frac{\ln(1-\theta z)}{\ln(1-\theta)} = \frac{z\ {}_2F_1[1, 1; 2; \theta z]}{{}_2F_1[1, 1; 2; \theta]}. \tag{7.4}$$

when shifted to the origin, it is a Kemp generalized hypergeometric probability distribution; see Section 2.4.1.

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

The mean and variance are

$$\mu = \frac{a\theta}{1-\theta}, \tag{7.5}$$

$$\mu_2 = \frac{a\theta(1-a\theta)}{(1-\theta)^2} = \mu[(1-\theta)^{-1} - \mu], \tag{7.6}$$

where $a = -1/\ln(1-\theta)$.

Confusion has sometimes arisen concerning the name *log-series distribution*. The term originated in the literature on species abundance data, in particular in Fisher, Corbet, and Williams (1943) and Williams (1964). Consider a catch of individuals containing S species and N individuals. If the expected *frequency* distribution of species represented r times is given by the terms of the series $-\alpha \ln(1-\theta)$, that is, $E[f_r] = \alpha\theta^r/r$, then $E[S] = -\alpha \ln(1-\theta)$ and $E[N] = \alpha\theta/(1-\theta)$. This is a two-parameter distribution; the parameter α is known as the index of diversity. Fisher's derivation of the log-series distribution was not well worded. Rao (1971), Boswell and Patil (1971), and Lo and Wani (1983) have endeavored to make his argument more rigorous; see also Wani (1978) concerning the interpretation of the parameters.

Since the late-1970s the terms "logarithmic distribution" and "log-series distribution" have generally been used interchangeably to mean the one-parameter distribution with pmf (7.1).

Further confusion occurred in the literature on floating-point arithmetic where the term "logarithmic distribution" has been used to denote the distribution of X where $\log_\beta X (\text{mod } 1)$ is uniformly distributed on [0, 1] (see, e.g., Bustoz et al., 1979).

7.1.2 Historical Remarks and Genesis

In a little-known German paper, Lüders (1934) used the terms of a logarithmic series as the parameters for a convolution of distributions that are Poisson, Poisson doublet, Poisson triplet, and so on, thereby obtaining a negative binomial distribution. Quenouille (1949) reinterpreted this relationship, showing that the negative binomial arises as the sum of n independent logarithmic variables where n has a Poisson distribution; see Section 5.3.

The log-series distribution has been used extensively by Williams [see, e.g., Williams (1947, 1964) and Section 7.1.9]. This led to Fisher's derivation of the logarithmic distribution (Fisher et al., 1943) as the limit as $k \to 0$ of the zero-truncated negative binomial distribution. A more rigorous proof is

$$\begin{aligned}\lim_{k\to 0} \frac{[(1-\theta)/(1-\theta z)]^k - (1-\theta)^k}{1-(1-\theta)^k} &= \lim_{k\to 0} \frac{(1-\theta z)^{-k} - 1}{(1-\theta)^{-k} - 1} \\ &= \frac{\theta z + \theta^2 z^2/2 + \cdots}{\theta + \theta^2/2 + \cdots};\end{aligned} \tag{7.7}$$

see also Stuart and Ord (1987, p. 177). The formal mathematical relationship between Fisher's derivation and Quenouille's result was explained by Kemp (1978a).

An alternative model for the log-series distribution with two parameters was put forward by Anscombe (1950, p. 360) who envisaged a sample of species from a finite but unknown number of species. He wrote, "It is a multivariate distribution, consisting of a set of independent Poisson distributions with mean values $\alpha X, \alpha X^2/2, \alpha X^3/3, \ldots$. A 'sample' comprises one reading from each distribution." Boswell and Patil (1971) discussed this model at length and showed that it implies that the conditional distribution of the number of individuals behaves like the sum of S_0 independent logarithmic distributions (i.e., has a Stirling distribution of the first kind); see Section 7.1.12.

The negative binomial distribution is a gamma–mixed Poisson distribution (see Section 5.3); this implies that there are two mixed Poisson models for the logarithmic distribution, corresponding to mixing after or mixing before truncation of the zero frequency. In the first case we have

$$\int_0^\infty \left(\frac{e^{\lambda z}-1}{e^\lambda - 1}\right) \frac{(e^\lambda - 1)e^{-\lambda/\theta}\, d\lambda}{(-\lambda)\ln(1-\theta)} = \frac{\ln(1-\theta z)}{\ln(1-\theta)} \tag{7.8}$$

(this can be shown by expanding $e^{\lambda z}-1$ as an infinite series in λz). In the second case we have

$$\int_\epsilon^\infty e^{\lambda(z-1)} \frac{e^{\lambda(\theta-1)/\theta}\, d\lambda}{\lambda E_1[\epsilon(1-\theta)/\theta]} = \frac{E_1[\epsilon/\theta - \epsilon z]}{E_1[\epsilon/\theta - \epsilon]}, \tag{7.9}$$

where $E_1[\omega]$ is the exponential integral

$$E_1[\omega] = \int_\omega^\infty e^{-t} t^{-1}\, dt = \Gamma(0, \omega);$$

the logarithmic distribution is obtained when the zero class is truncated and $\epsilon \to 0$. Such models were first suggested by Kendall (1948) in his paper on some modes of growth leading to the distribution.

The negative binomial distribution is the outcome of a number of stochastic processes, for example, the Yule–Furry process, the linear birth–death process, and the Pólya process. Kendall also studied the logarithmic distribution as a limiting form of such processes. He showed moreover that

$$\int_0^\theta \frac{z(1-v)}{1-vz} \frac{dv}{(v-1)\ln(1-\theta)} = \frac{\ln(1-\theta z)}{\ln(1-\theta)}, \tag{7.10}$$

that is, that the logarithmic distribution is a mixed shifted geometric distribution. A computer generation algorithm for the distribution has been based on this property; see Section 7.1.6. The logarithmic distribution shifted to support $0, 1, 2, \ldots$ is similarly a mixed geometric distribution.

The steady-state birth–death process with linearly state-dependent birth and death rates appears in Caraco (1979) in the context of animal group-size dynamics. For the steady-state distribution to be logarithmic, it is necessary to assume that $\mu_i = \mu i$ for $i > 1$ and $\mu_i = 0$ when $i = 1$.

When Rao's damage process (Section 9.2) is applied to the logarithmic distribution with parameter θ, the result is another logarithmic distribution with parameter $\theta p/(1-\theta+\theta p)$. Useful reviews of models for the logarithmic distribution are given by Nelson and David (1967), Boswell and Patil (1971), and Kemp (1981a). A functional equation for the pgf of the logarithmic distribution is presented in Panaretos (1987b).

7.1.3 Moments

The rth factorial moment of the logarithmic distribution is

$$\begin{aligned}\mu'_{[r]} &= a\theta^r \sum_{k=r}^{\infty}(k-1)(k-2)\cdots(k-r+1)\theta^{k-r}\\ &= a\theta^r \frac{d^{r-1}}{d\theta^{r-1}}\left(\sum_{k=1}^{\infty}\theta^{k-1}\right)\\ &= a\theta^r(r-1)!(1-\theta)^{-r}. \end{aligned} \tag{7.11}$$

The mgf is

$$E[e^{tX}] = \frac{\ln(1-\theta e^t)}{\ln(1-\theta)}. \tag{7.12}$$

The first four uncorrected moments about the origin are

$$\begin{aligned}\mu'_1 &= \mu = a\theta(1-\theta)^{-1},\\ \mu'_2 &= a\theta^2(1-\theta)^{-2} + a\theta(1-\theta)^{-1} = a\theta(1-\theta)^{-2},\\ \mu'_3 &= a\theta(1+\theta)(1-\theta)^{-3},\\ \mu'_4 &= a\theta(1+4\theta+\theta^2)(1-\theta)^{-4},\end{aligned} \tag{7.13}$$

where $a = -1/\ln(1-\theta)$, and the corresponding moments about the mean are

$$\begin{aligned}\mu_2 &= a\theta(1-a\theta)(1-\theta)^{-2} = \mu'_1[(1-\theta)^{-1} - \mu'_1],\\ \mu_3 &= a\theta(1+\theta-3a\theta+2a^2\theta^2)(1-\theta)^{-3},\\ \mu_4 &= a\theta[1+4\theta+\theta^2-4a\theta(1+\theta)+6a^2\theta^2-3a^3\theta^3](1-\theta)^{-4}.\end{aligned} \tag{7.14}$$

Table 7.1 (β_1, β_2) Points of Logarithmic Distributions

θ	β_1	β_2	θ	β_1	β_2
0.05	43.46	49.86	0.55	9.01	16.46
0.10	23.59	30.06	0.60	9.03	16.66
0.15	17.06	23.61	0.65	9.14	16.97
0.20	13.88	20.51	0.70	9.36	17.42
0.25	12.04	18.76	0.75	9.70	18.04
0.30	10.88	17.70	0.80	10.20	18.89
0.35	10.12	17.04	0.85	10.93	20.09
0.40	9.61	16.65	0.90	12.08	21.91
0.45	9.17	16.36	0.95	14.25	25.28
0.50	9.09	16.39			

The coefficient of variation is

$$\frac{\mu_2^{1/2}}{\mu} = (a^{-1}\theta^{-1} - 1)^{1/2}.$$

The index of dispersion is

$$\frac{\mu_2}{\mu} = \frac{1 - a\theta}{1 - \theta} \gtrless 1 \quad \text{as } a \lessgtr 1, \quad \text{i.e., as } \theta \gtrless 1 - e^{-1}. \tag{7.15}$$

The moment ratios $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$ both tend to infinity as $\theta \to 0$ and as $\theta \to 1$, with

$$\lim_{\theta \to 0} \left(\frac{\beta_2}{\beta_1} \right) = 1 \qquad \text{and} \qquad \lim_{\theta \to 1} \left(\frac{\beta_2}{\beta_1} \right) = \frac{3}{2}.$$

The (β_1, β_2) points of the logarithmic distribution are given in Table 7.1.

The moments about zero satisfy the relation

$$\mu'_{r+1} = \theta \frac{d\mu'_r}{d\theta} + \frac{a\theta}{1 - \theta} \mu'_r. \tag{7.16}$$

The central moments satisfy the relation

$$\mu_{r+1} = \theta \frac{d\mu_r}{d\theta} + r\mu_2\mu_{r-1}. \tag{7.17}$$

Since the logarithmic distribution is a GPSD, it follows that the cumulants satisfy the recurrence relations (Khatri, 1959)

$$\kappa_r = \theta \frac{d\kappa_{r-1}}{d\theta} \tag{7.18}$$

and that the factorial cumulants satisfy

$$\kappa_{[r]} = \theta \frac{d\kappa_{[r-1]}}{d\theta} - (r-1)\kappa_{[r-1]}. \tag{7.19}$$

The mean deviation is

$$\nu_1 = 2a \sum_{k=1}^{[\mu]} \frac{(\mu - k)\theta^k}{k} = \frac{2a\theta(\theta^{[\mu]} - \Pr[X > [\mu]])}{1-\theta}, \tag{7.20}$$

where $\mu = a\theta(1-\theta)^{-1}$ and $[\mu]$ denotes the integer part of μ; see Kamat (1965).

7.1.4 Properties

From (7.1)

$$\frac{\Pr[X = x+1]}{\Pr[X = x]} = \frac{x\theta}{x+1}, \qquad x = 1, 2, \ldots. \tag{7.21}$$

This ratio is less than 1 for all values of $x = 1, 2, \ldots$, since $\theta < 1$. Hence the value of $\Pr[X = x]$ decreases as x increases.

The failure rate r_x is given by

$$\begin{aligned} \frac{1}{r_x} &= \sum_{i \geq x} \frac{\Pr[X = i]}{\Pr[X = x]} \\ &= 1 + \frac{x\theta}{x+1} + \frac{x\theta^2}{x+2} + \cdots < \frac{1}{r_{x+1}}, \end{aligned} \tag{7.22}$$

and hence the distribution has a decreasing failure rate (Patel, 1973); see also Gupta, Gupta, and Tripathi (1997).

Methods of approximating $\Pr[X \leq [\mu]]$ are described in Section 7.1.5. Approximation (7.25) leads to the following approximate formula for the median M of the distribution:

$$M \approx e^{-\gamma}(1-\theta)^{-1/2} + 0.5 + e^{-2\gamma} \doteqdot 0.56146(1-\theta)^{-1/2} + 0.81524, \tag{7.23}$$

where γ is Euler's constant; see Section 1.1.2. This formula has been attributed to Grundy (Williams, 1964; Gower, 1961) and it seems to give a good approximation if θ is not too small.

The distribution has a long positive tail. For large values of x the shape of the tail is similar to that of a geometric distribution (Section 5.2) with parameter θ.

Table 7.2 Properties of the Logarithmic Distribution for Various Values of θ

Table 7.2 from Kemp (1981b) summarizes various properties of the logarithmic distribution.

The entropy of the distribution is

$$a \sum_{j=1}^{\infty} \theta^j j^{-1} \log j - a\theta(1-\theta)^{-1} \log \theta - \log a.$$

Siromoney (1962) has shown that this is an increasing function of θ. Siromoney also gave an expression for the entropy of an individual with respect to a species for the log-series distribution.

Another property of the logarithmic distribution shifted to the support $0, 1, 2, \ldots$ is infinite divisibility, a result that was first proved by Katti (1967). Infinite divisibility follows from the logconvexity of the probabilities [see Steutel (1970)]. The shifted logarithmic distribution arises therefore from a Poisson distribution of clusters of various sizes. Explicit formulas for the cluster size probabilities were given by Kemp (1978a).

Patil and Seshadri (1975) showed that, if X has a logarithmic distribution with parameter θ, then the conditional distribution of X/k, given that X/k is a positive integer, is logarithmic with parameter θ^k.

The size-biased logarithmic distribution has the pmf

$$\Pr[X = x] = \frac{x\theta^x/x}{\sum_{x=1}^{\infty} \theta^x} = (1-\theta)\theta^{x-1}.$$

and hence is a geometric distribution on $1, 2, \ldots$. The weighted logarithmic distribution with weight function α^x, $0 < \alpha < 1$, is another logarithmic distribution with parameter $\alpha\theta$.

7.1.5 Approximations and Bounds

The cumulative probability

$$\Pr[X \le x] = \frac{-\sum_{j=1}^{x} \theta^j/j}{\ln(1-\theta)} = 1 + \frac{\sum_{j=x+1}^{\infty} \theta^j/j}{\ln(1-\theta)}$$

can be approximated using

$$\begin{aligned}\sum_{j=x+1}^{\infty} \frac{\theta^j}{j} &= -\ln(1-\theta) - \sum_{j=1}^{x} j^{-1} + \int_0^{1-\theta} \phi^{-1}[1-(1-\phi)^x]\,d\phi \\ &\approx -\ln(1-\theta) - \left(\gamma + \tfrac{1}{2}x^{-1} - \tfrac{1}{12}x^{-2} + \log x\right) \\ &\quad + \{-Ei(-x(1-\theta)) + \gamma + \ln[x(1-\theta)]\} \\ &\approx -Ei(-x(1-\theta)) - \tfrac{1}{2}x^{-1} + \tfrac{1}{12}x^{-2} \end{aligned} \tag{7.24}$$

with an error less than $x(1-\theta)^2$ for $x \ge 10$; $Ei(u) = \int_{-\infty}^{u} z^{-1}e^z\,dz$ is called the exponential integral.

If $x(1-\theta)$ is small, then the approximation

$$\sum_{j=x+1}^{\infty} \frac{\theta^j}{j} \approx -\ln[x(1-\theta)] + x(1-\theta) - \gamma \tag{7.25}$$

has an error less than $[x(1-\theta)]^2/4$ for $x \ge 10$. Also

$$\sum_{j=x+1}^{\infty} \frac{\theta^j}{j} \approx Ei(x\ln\theta) + \frac{\theta^x}{2x}. \tag{7.26}$$

These formulas are given in Gower (1961), and they lead to an approximation for the median of the distribution; see Section 7.1.4.

Owen (1965) gave the following formula, which is suitable when $x(1-\theta)$ is large:

$$\begin{aligned}\sum_{j=x+1}^{\infty} \frac{\theta^j}{j} \approx \theta^{x+1}(1-\theta)^{-1}\big\{&x^{-1} - 1![x(x-1)]^{-1}(1-\theta)^{-1} \\ &+ 2![x(x-1)(x-2)]^{-1}(1-\theta)^{-2} - \cdots \\ &+ (-1)^r r![x(x-1)\cdots(x-r)]^{-1}(1-\theta)^{-r}\big\}.\end{aligned} \tag{7.27}$$

The inequalities

$$\theta^{x+1}(x+1)^{-1}\left[1 - \frac{(x+1)\theta}{(x+2)}\right] < \sum_{j=x+1}^{\infty} \frac{\theta^j}{j} < \theta^{x+1}(x+1)^{-1}(1-\theta)^{-1} \tag{7.28}$$

provide simple bounds for the tail probability and are useful for large x.

7.1.6 Computation, Tables, and Computer Generation

The one-term recurrence relationship

$$\Pr[X = x + 1] = \frac{\theta x \Pr[X = x]}{x + 1}, \qquad x = 1, 2, \ldots,$$

together with $\Pr[X = 1] = -\theta / \ln(1 - \theta)$, facilitates the computation of the probabilities and the cumulative probabilities. When $\theta > 0.99$, the tail length is great, and it may be advisable to use high-precision arithmetic on a computer.

Kemp (1981b) suggested and discussed various algorithms for the computer generation of logarithmic pseudorandom variables.

Tables of the individual probabilities (7.1) and cumulative probabilities were given by Williamson and Bretherton (1964). The argument of the tables is the expected value $\mu = -\theta[(1 - \theta)\ln(1 - \theta)]^{-1}$ and not θ; values of the second and fourth differences were also given.

Extensive tables of the probabilities as a function of θ were provided by Patil et al. (1964) and Patil and Wani (1965a). A table for μ as a function of θ for values of θ from 0.01 to 0.99 appeared in Patil (1962d). The availability of tables for minimum-variance unbiased estimation was described by Patil (1985).

Fisher et al. (1943) gave a table which enables the "index of diversity" α to be estimated from observed values of N and S using $e^{E[S]/\alpha} = 1 + E[N]/\alpha$; see Section 7.1. Alternatively, the tables of Barton et al. (1963) may be used.

Efficient ways of generating rv's from a logarithmic distribution with a fixed value of θ by searching a stored table of the cumulative probabilities are discussed in Devroye (1986). Devroye also (in his Chapter 3) describes other very efficient general methods suitable for fixed θ, such as the alias and acceptance-complement methods.

When θ varies, a distribution-specific method is required. Kemp, Kemp, and Loukas (1979) gave a build-up unstored-search procedure similar to Fishman's method for the Poisson distribution; Kemp (1981b) presented an algorithm for an unstored chop-down search procedure. Very high values of θ (>0.99) are commonplace in ecological applications of the logarithmic distribution. With this in mind, Kemp (1981b) also gave a generation algorithm based on the mixed shifted-geometric model for the distribution (see Section 7.1.2). She showed that two variants of this mixed shifted-geometric method make enormous savings in computer time for very high values of θ.

Devroye (1986) gave details of all these algorithms and also presented two other, seemingly less attractive, algorithms. Shanthikumar's (1985) *discrete thinning method* (for distributions with hazard rate bounded below unity) and *dynamic thinning method* (for decreasing-failure-rate distributions) are interesting; Devroye showed how the discrete thinning method can be used for the logarithmic distribution. He also gave an algorithm based on rejection from an exponential distribution.

7.1.7 Estimation

Model Selection Before considering numerical estimation of the parameter θ of the logarithmic distribution, we first give details of graphical methods of model selection.

Ord (1967a, 1972) has shown that plotting $u_x = xf_x/f_{x-1}$ against x can be expected to give a straight line with intercept $-\theta$ and slope θ, where f_x is the observed frequency of an observation x (see Section 2.3.3). Clearly successive u_x are dependent; to smooth the data, Ord suggested the use of the statistic $v_x = (u_x + u_{x-1})/2$. In Ord (1967a) he applied the method to Corbet's butterfly data (Fisher et al., 1943), and in Ord (1972) to product-purchasing data.

Gart (1970) noted that $x\Pr[X = x]/\{(x-1)\Pr[X = x-1]\} = \theta$ and proposed the use of $xf_x/[(x-1)f_{x-1}]$; when plotted against x, this can be expected to give a horizontal line with intercept θ.

A third method, proposed by Hoaglin, Mosteller, and Tukey (1985), is to plot $\log n_x^* + \log x - \log N$ against x, where N is the total frequency and n_x^* is obtained by smoothing f_x. This can be expected to give a straight line with intercept $-\ln[-\ln(1-\theta)]$ and slope θ. Hoaglin and Tukey (1985) considered their method to be superior to Ord's. They too used Corbet's butterfly data for illustration.

These graphical methods would seem to have much to commend them with regard to distribution selection, but their usefulness for parameter estimation may be dubious.

Point Estimation and Confidence Intervals Consider now the problem of the numerical estimation of θ given values of n independent rv's $x_1, x_2, \ldots, x_n$ each having distribution (7.1). The maximum-likelihood estimator $\hat{\theta}$ is given by

$$\overline{x} = n^{-1}\sum_{j=1}^{n} x_j = \frac{\hat{\theta}}{-(1-\hat{\theta})\ln(1-\hat{\theta})}; \tag{7.29}$$

the asymptotic variance is $n^{-1}(\theta^2/\mu_2)$, where μ_2 is the variance of the logarithmic distribution (Patil, 1962d). To estimate the variance of $\hat{\theta}$, the usual practice is to replace θ by $\hat{\theta}$. Because (7.1) is a generalized PSD, maximum-likelihood and moment estimation are equivalent.

Whereas Patil (1962d) gave a table of μ as a function of θ, Williamson and Bretherton (1964) provided values of θ corresponding to μ (they also tabulated the bias and standard error of $\hat{\theta}$ in some cases). Patil and Wani (1965a) gave a more extensive table of θ for values of μ. Barton et al. (1963) gave solutions of b (to seven decimal places) for the equation $e^b = 1 + b\overline{x}$, where $1 - (\overline{x})^{-1} = 0(0.001)0.999$. The MLE $\hat{\theta}$ can then be calculated as

$$\hat{\theta} = \frac{b\overline{x}}{1 + b\overline{x}}. \tag{7.30}$$

Chatfield (1969) noted that θ changes very slowly for large values of μ and advocated reparameterization with $\psi = \theta/(1-\theta)$; he tabulated ψ against μ for $\mu = 0.0(0.1)20(1)69$.

A detailed study of the MLE $\hat{\theta}$ was carried out by Bowman and Shenton (1970). They gave values for the mean, variance, skewness ($\sqrt{\beta_1}$), and kurtosis (β_2) of $\hat{\theta}$ for various combinations of $\hat{\theta}$ and the sample size n, where $0.1 \le \hat{\theta} \le 0.9$ and $n \ge 8$; they commented on the computational difficulties when $0.9 < \hat{\theta} < 1.0$. Bowman and Shenton (1970, p. 136) made the general statement that "the distribution of $\hat{\theta}$ is not too far removed from the normal distribution; however, departures from normality become serious when $\hat{\theta}$ exceeds about 0.9 [or] when the sample size is less than about nine." Bowman and Shenton remarked that the convergence of the Taylor expansion for $\hat{\theta}$ poses difficult problems.

Earlier, Birch (1963) had provided a computer algorithm for $\hat{\theta}$. This uses $\xi = (1-\hat{\theta})^{-1}$ and solves $\xi - 1 - \overline{x}\ln\xi = 0$ by Newton–Raphson iteration, taking

$$\xi_{i+1} = \frac{1 - \overline{x}(\log\xi_i - 1)}{1 - \overline{x}/\xi_i}, \qquad i = 1, 2, \ldots.$$

The recommended first approximation is $\xi_1 = 1 + [k(\overline{x}-1)+2]\ln\overline{x}$, with k some value between 1 and $\frac{5}{3}$. When $\overline{x}$ is close to unity, Birch recommended, instead, the use of the first few terms of the Taylor series

$$\xi = 1 + 2(\overline{x}-1) + \frac{2(\overline{x}-1)^2}{3} - \frac{2(\overline{x}-1)^3}{9} + \frac{14(\overline{x}-1)^4}{135} - \cdots.$$

Böhning (1983a,b) treated maximum-likelihood estimation for the logarithmic distribution as a problem in numerical analysis. He referred to Birch (1963), but not to Bowman and Shenton (1970). Bowman and Shenton stated, without proof, the uniqueness of $\hat{\theta}$. Böhning has proved this. He has also shown that $\hat{\theta}$ is the MLE of θ iff $\hat{\theta}$ is a fixed point of

$$\Psi(\hat{\theta}) = \frac{\overline{x}\log(1-\hat{\theta})}{\overline{x}\log(1-\hat{\theta}) - 1}. \tag{7.31}$$

He found that $\hat{\theta}_{i+1} = \Psi(\hat{\theta}_i)$ converges monotonically to $\hat{\theta}$ and that for this method, unlike Newton–Raphson, the choice of $\hat{\theta}_1$ is not critical. His computer timings for his fixed-point method, for Newton-Raphson, and for a secant method indicate that his fixed-point method sometimes converges much more slowly. Nevertheless, it converges more surely.

An alternative maximum-likelihood approach is to use a computer optimization package to solve the maximum-likelihood equation. Such packages often require bounds for the parameter estimates to be specified. Kemp and Kemp (1988) gave

Table 7.3 Asymptotic Efficiencies of Estimators of θ for the Logarithmic Distribution

	Efficiency (%)		
Estimator	$\theta = 0.1$	$\theta = 0.5$	$\theta = 0.9$
1	98.3	89.7	73.9
2	22.8	44.9	48.8
3	89.5	44.7	5.7

the following bounds for $\hat{\theta}$:

$$\frac{(9\overline{x} - 6) - (9\overline{x}^2 - 12\overline{x} + 12)^{1/2}}{6\overline{x} - 2} < \hat{\theta} < \frac{(6\overline{x} - 3) - (24\overline{x}^2 - 24\overline{x} + 9)^{1/2}}{\overline{x}}. \tag{7.32}$$

Patil (1962d) suggested the following explicit estimators of θ (based on relationships that are exactly true in the population):

1. $1 - f_1/\overline{x}$.

2. $1 - \sum_{j\geq 1}(jf_j)/\sum_{j\geq 1}(j^2 f_j)$.

3. $\theta^{\dagger} = \sum_{j\geq 2}[j/(j-1)]f_j$, where f_j, $j = 1, 2, \ldots,$ is the proportion of observations that are equal to j.

Estimators 1 and 2 are asymptotically unbiased. Estimator 3, $\theta^{\dagger}$, is unbiased for all sizes of sample and has variance

$$\text{Var}(\theta^{\dagger}) = n^{-1}\left[\sum_{j=2}^{\infty}\left(\frac{j}{j-1}\right)^2 \Pr[X = j] - \theta^2\right].$$

The asymptotic efficiencies of these estimators (relative to $\hat{\theta}$) for certain values of θ are shown in Table 7.3.

It appears that estimation method 1 is preferable to estimation methods 2 or 3 on grounds of both accuracy and simplicity. A. W. Kemp (1986) showed that the equation for estimator 1 is an approximation to the maximum-likelihood equation.

From (1993) has suggested a family of estimators of θ given by $\theta^{\ddagger} = A/B$, where

$$A = \sum_{j=1}^{\infty}(j+1)E_j f_{k+1}, \qquad B = \sum_{j=1}^{\infty} jE_j f_j. \tag{7.33}$$

The choice $E_j = 1$ gives Patil's estimator 1 while $E_j = 1/j$ gives Patil's estimator 3. From took $E_j = j^{\delta}$ and calculated the values of δ that maximize the asymptotic relative efficiency (ARE) for different values of θ; see Table 7.4. When $\delta = 0.40$, the ARE falls from 0.999 (for $\theta = 0.01$) to 0.872 (when $\theta = 0.99$).

Table 7.4 Values of δ That Maximize the ARE

From's first-stage estimator θ_1 uses $\delta = 0.40$. His second-stage estimator θ_2 uses the value of δ corresponding to θ_1 in Table 7.4. From carried out a large simulation study and came to the following paraphrased conclusions:

(i) The MLE $\hat{\theta}$ performs best for $\theta \approx 1$, although it is only slightly superior to θ_1 and θ_2.

(ii) If $\theta \approx 0$, then $\hat{\theta}$ has a large mean-square error (MSE) and bias; here θ_1 and θ_2 are greatly superior.

(iii) θ_2 is better than θ_1 when $\theta \approx 1$.

(iv) Overall θ_2 is best.

(v) The biases are comparable for large samples ($n \geq 500$).

(vi) For $n < 500$ and $\theta \approx 0$, the bias is much greater for $\hat{\theta}$ than for θ_1 or θ_2.

(vii) For $n < 500$ and $\theta \approx 1$, the bias is less for $\hat{\theta}$ than for θ_1 or θ_2.

(viii) $\hat{\theta}$, θ_1, and θ_2 all have negative biases and underestimate θ.

The minimum variance unbiased estimator (MVUE) θ^* of θ is given by the following equation:

$$\theta^* = \begin{cases} \dfrac{b(\sum_{j=1}^n x_j - 1)}{b(\sum_{j=1}^n x_j)} & \text{if } \displaystyle\sum_{j=1}^n x_j > n, \\ 0 & \text{if } \displaystyle\sum_{j=1}^n x_j = n, \end{cases} \tag{7.34}$$

where $b(m)$ is the coefficient of θ^m in $[-\ln(1-\theta)]^n$; see Patil, Boswell, Joshi, and Ratnaparkhi (1984) for relevant references. Tables of θ^* given $\sum_{j=1}^n x_j$ for small n were given to four decimal places in Patil et al. (1964).

Wani and Lo (1975a,b) provided tables and charts for obtaining confidence intervals for θ given the values of X in a random sample of size n. They used the observed value of $\sum_{j=1}^{n} x_j$ and the fact that the sum of independent logarithmic rv's has a Stirling distribution of the first kind; see Patil and Wani (1965b) and Section 7.1.12. These confidence intervals were compared with ones obtained by other methods in Wani and Lo (1977).

Estimation for the early two-parameter log-series distribution was studied by Anscombe (1950), Engen (1974), Lo and Wani (1983), and Rao (1971). Rao also discussed the special problems of testing goodness of fit for this distribution.

7.1.8 Characterizations

The following characterization of the logarithmic distribution has been given by Patil and Wani (1965b, pp. 276–277): "Let X and Y be two independent discrete random variables each taking the value 1 with nonzero probability; then if

$$\Pr[X = x|X + Y = z] = \frac{[x^{-1} + (z - x)^{-1}]\beta^x}{\sum_{j=1}^{z-1}[j^{-1} + (z - j)^{-1}]\beta^j}, \tag{7.35}$$

for $0 < \beta < \infty$, $z = 2, 3, \ldots$, then X and Y each has the logarithmic distribution with parameters in the ratio β." Wani (1967) has also shown that a distribution is a logarithmic distribution with parameter θ iff its moments are $\mu_1' = a\theta/(1 - \theta)$, $\mu_s = a\theta c_s(\theta)/(1 - \theta)^{s+2}$, $s = 0, 1, 2, \ldots$, with $a = -[\ln(1 - \theta)]^{-1}$ and $c_s(\theta) = \sum_{i=0}^{s} c(s, i)\theta^i$, where

$$c(s, i) = (i + 1)c(s - 1, i) + (s - i + 1)c(s - 1, i - 1)$$

and $c(0, 0) = 1$, $c(s, i) = 0$ for $i < 0$.

Kyriakoussis and Papageorgiou (1991b) have obtained the following two characterizations: First, suppose that X is a rv with nonzero pmf for every nonnegative integer value of X and finite mean and also that Y is another nonnegative integer-valued rv such that the conditional distribution of $Y|X = x$ is binomial. Then the regression function of X on Y has the form

$$E[X|Y = y] = \begin{cases} b & \text{if } y = 0, \\ ay & \text{if } y = 1, 2, \ldots, \end{cases} \tag{7.36}$$

where a and b are constants iff X has a logarithmic distribution with added zeroes (a modified logarithmic distribution; see Section 7.1.10).

Their second characterization is a characterization of the logarithmic distribution itself, and it involves the assumption that $Y|X = x$ is a positive integer-valued rv with pmf

$$\Pr[Y = y|X = x] = \binom{y-1}{x-1} p^x (1-p)^{y-x}, \tag{7.37}$$

where $y = x, x+1, \ldots, x = 1, 2, \ldots, 0 < p < 1$. Then the regression function $E[X|Y = y]$ has a specified form (more complicated than above) iff X has a logarithmic distribution.

7.1.9 Applications

In Fisher, Corbet, and Williams (1943), the log-series distribution was applied to the results of sampling butterflies (Corbet's data) and also to data obtained in connection with the collection of moths by means of a light-trap (Williams's data). In these experiments it was found that, if the number of species represented by exactly one individual is n_1, then the numbers of species represented by 2, 3, ... individuals are approximately $(n_1/\theta)\theta^2/2$, $(n_1/\theta)\theta^3/3, \ldots$, respectively, where θ is a positive number less than unity; see Section 7.1. The quantity $\alpha = n_1/\theta$ is sometimes called an "index of diversity" and is thought to be independent of the size of catch.

Rowe (1942) fitted data on the numbers of mosquitoes, N, caught in light traps in 10 cities in Iowa, together with the numbers of species, S, observed. For each city he estimated $\hat{\alpha}$ by eliminating $\hat{\theta}$ from $S = -\hat{\alpha}\ln(1-\hat{\theta})$, $N = \hat{\alpha}\hat{\theta}/(1-\hat{\theta})$, that is, by solving $\exp(S/\hat{\alpha}) = 1 + N/\hat{\alpha}$. The values of $\hat{\alpha}$ ranged from 1.71 to 2.24; Rowe considered that this was good evidence for a constant value of α (of about 2) for mosquito data.

Williams (1964) gave numerous other sets of ecological data on species abundance where the log-series distribution fits very well, such as Blackman's (1935) data on the average number of plant species found on quadrats of various sizes in a grassland formation, and also the distribution of species of British nesting birds by numbers of individuals [Witherby et al. (1941)]. Williams (1964) also fitted the logarithmic distribution to many instances of data concerning the distribution of parasites per host, such as the number of head lice per host (Buxton, 1940). In Williams (1944), he had applied it to numbers of publications by entomologists. An application to an inventory control problem in the steel industry appears in Williamson and Bretherton (1964).

In a long series of papers, including Chatfield et al. (1966) and Chatfield (1970, 1986), the authors have described the use of the logarithmic distribution to represent the distribution of numbers of items of a product purchased by a buyer in a given period of time. They remarked that the logarithmic distribution is likely to be a useful approximation to a negative binomial distribution with a low value of the parameter k (e.g., less than 0.1). The logarithmic distribution has the advantage of depending on only one parameter, θ, instead of the two, k and P, that are needed for the negative binomial distribution.

Winkelmann (2000) has commented that the logarithmic distribution has been an ineffective competitor to the Poisson and negative binomial distributions in a regression framework; he attributes this to the practical problems associated with the mean function.

7.1.10 Truncated and Modified Logarithmic Distributions

The left (head) truncated logarithmic distribution with support $k, k+1, \ldots$ has the pgf

$$G_{\mathrm{LT}}(z) = \frac{\theta^{k-1} z^k [\theta/k + \theta^2 z/(k+1) \cdots]}{\sum_{i=k}^{\infty} \theta^i / i}$$
$$= z^k \, \frac{{}_2F_1[1, k; k+1; \theta z]}{{}_2F_1[1, k; k+1; \theta]} \tag{7.38}$$

and therefore is an extended beta–binomial (i.e., an extended generalized Waring) distribution; see Section 6.11.1. Minimum-variance unbiased estimation of the left-truncated distribution has been examined by Ahuja and Enneking (1972b) for k known and by Enneking and Ahuja (1978) for k unknown. Enneking and Ahuja gave further relevant references.

The most common form of truncation, however, is by exclusion of values greater than a specified value r. This kind of truncation is likely to be encountered with distributions of logarithmic type, as they have long positive tails and it is not always practicable to evaluate each large observed value. The pgf for the right-truncated logarithmic distribution is

$$G_{\mathrm{RT}}(z) = z \, \frac{{}_3F_2[1-r, 1, 1; 1-r, 2; \theta z]}{{}_3F_2[1-r, 1, 1; 1-r, 2; \theta]}. \tag{7.39}$$

Paloheimo (1963) describes circumstances under which a right-truncated logarithmic distribution might be appropriate.

If $X_1, X_2, \ldots, X_n$ are independent rv's each with the distribution

$$\Pr[X = x] = \frac{\theta^x}{x} \left(\sum_{j=1}^{r} \frac{\theta^j}{j} \right)^{-1}, \qquad x = 1, 2, \ldots, r, \tag{7.40}$$

then the MLE $\hat{\theta}$ of θ satisfies the equation

$$\overline{x} = n^{-1} \sum_{j=1}^{n} x_j = \frac{\hat{\theta}(1 - \hat{\theta}^r)}{(1-\hat{\theta})[\sum_{j=1}^{r} (\hat{\theta}^j / j)]}. \tag{7.41}$$

Patil and Wani (1965a) provided tables giving the value of $\hat{\theta}$ corresponding to selected values of $\overline{x}$ and r. They also gave some numerical values of the bias and standard error of $\hat{\theta}$.

Alternatively, a method based on equating sample and population moments may be used. This leads to

$$\theta^* = \frac{m_3' - (r+2)m_2' + (r+1)m_1'}{m_3' - rm_2'}, \tag{7.42}$$

where $m_s' = n^{-1}\sum_{j=1}^{n} x_j^s$.

A modified logarithmic distribution, called the *logarithmic-with-zeroes distribution*, or *log-zero distribution*, with pgf

$$G(z) = c + (1-c)\frac{\ln(1-\theta z)}{\ln(1-\theta)} \tag{7.43}$$

was introduced by Williams (1947). Derivations of the distribution via models of population growth were given by Feller (1957, p. 276), Kendall (1948), and Bartlett (1960, p. 9). Its use as a model for stationary purchasing behavior was suggested by Chatfield et al. (1966). Khatri (1961), also Patil (1964b), obtained it by mixing binomials, the binomial index having a logarithmic distribution; we have

$$G(z) = \frac{\sum_{n=1}^{\infty}(q+pz)^n\theta^n/n}{-\ln(1-\theta)} = c + (1-c)\frac{\ln(1-\phi z)}{\ln(1-\phi)}, \tag{7.44}$$

with $c = \ln(1-q\theta)/\ln(1-\theta)$ and $\phi = p\theta/(1-q\theta)$. The distribution is discussed in greater detail in Section 8.2.3.

Lwin (1981) studied a more complicated modification. Here the logarithmic distribution is shifted from support $1, 2, \ldots$ to support $k, k+1, \ldots$ and the probabilities for $x = 0, 1, \ldots, k-1$ are arbitrary. The pmf is

$$\Pr[X = x] = \begin{cases} \theta_x & \text{for } x = 0, 1, \ldots, k-1, \quad (7.45) \\ \dfrac{a\theta^{x-k+1}}{(k-x+1)\log(1-\theta)} & \text{for } x = k, k+1, \ldots, \quad (7.46) \end{cases}$$

where $a = 1 - \sum_{i=0}^{k-1}\theta_i$, $0 < \theta < 1$, $0 < a < 1$ and k is a positive integer. The pgf is

$$G_W(z) = \sum_{x=0}^{k-1}\theta_x z^x + az^k\,\frac{{}_2F_1[1, 1; 2; \theta z]}{{}_2F_1[1, 1; 2; \theta]}. \tag{7.47}$$

It becomes the logarithmic distribution when $k = 1$, $\theta_0 = 0$.

7.1.11 Generalizations of the Logarithmic Distribution

The *"generalized" logarithmic distribution* of Jain and Gupta (1973) is the *Lagrangian logarithmic distribution.* It is formed by letting $n \to \infty$ in a zero-truncated Lagrangian negative binomial distribution and is dealt with in Section 7.2.8.

Kempton's (1975) full beta generalization of the log-series distribution has three independent parameters, b, p, and q. It is a Poisson distribution mixed using a beta distribution of the second kind; see Holla and Bhattacharya (1965). The pmf is

$$\Pr[X = x] = \frac{1}{B(p,q)} \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} \frac{b^p \lambda^{p-1}\, d\lambda}{(1+b\lambda)^{p+q}} \tag{7.48}$$

$$= \frac{\Gamma(p+x)\, b^p}{x! B(p,q)}\, {}_2F_0[p+q, p+x; -; -b], \qquad b,\ p,\ q > 0, \tag{7.49}$$

where $x = 0, 1, \ldots$. The distribution is J shaped, like the log-series distribution; however, it has an even longer tail.

Kempton's (1975) "generalized" log-series distribution is obtained when $p \to 0$. The pmf is

$$\Pr[X = x] = C \int_0^\infty \frac{e^{-t} t^{x-1}\, dt}{x!(1+bt)^q}, \qquad b,\ q > 0, \tag{7.50}$$

where $x = 1, 2, \ldots$ and C is a normalizing constant. If $b \to 0$ and $q \to \infty$ in such a way that $\rho = bq$ remains finite and positive, then it tends to a log-series distribution with

$$\Pr[X = x] = \frac{x-1}{x(1+\rho)} \Pr[X = x-1].$$

Tripathi, Gupta, and White (1987) studied a length-biased version of this distribution. It is a generalization of the geometric distribution and is known as the GG1 distribution.

Ong (1995b) has given expansions and three-term recurrence relations for the probabilities for (i) the full beta model, (ii) Kempton's generalized log-series distribution, and (iii) the GG1 distribution. He made a special study of the stability of these recurrence relations. The three distributions are compared in Ong and Muthaloo (1995) with a new four-parameter, very long-tailed distribution. Its pmf involves a Bessel (equivalently a Whittaker) function. Ong and Muthaloo called it the modified Bessel function distribution of the third kind mixed Poisson (BF3-P) distribution. The distribution, known as Ong's distribution, is a convolution of a negative binomial and a Poisson m-tuplet distribution (Ong, 1987).

A different generalization of the logarithmic distribution has been obtained by Tripathi and Gupta (1985) as a limiting form of the distribution with pgf

$$g(z) = z\,\frac{{}_2F_1[a/\beta + c, 1; \lambda + c; \beta z]}{{}_2F_1[a/\beta + c, 1; \lambda + c; \beta]}.$$

When $c = 1$ and $a \to 0$, we have

$$G(z) = \lim_{a\to 0} z\,\frac{{}_2F_1[a/\beta + 1, 1; \lambda + 1; \beta z]}{{}_2F_1[a/\beta + 1, 1; \lambda + 1; \beta]} = z\,\frac{{}_2F_1[1, 1; \lambda + 1; \beta z]}{{}_2F_1[1, 1; \lambda + 1; \beta]}, \qquad (7.51)$$

where either (i) $0 < \lambda \le 1$, $0 < \beta < 1$ or (ii) $1 < \lambda$, $0 < \beta \le 1$. It is a shifted generalized hypergeometric distribution, with the properties of such distributions. Tripathi and Gupta called this the *hyperlogarithmic distribution*. The pmf is

$$\Pr[X = x] = \frac{\lambda!(x-1)!\beta^x}{(\lambda + x - 1)!\,{}_2F_1[1, 1; \lambda + 1; \beta]}, \qquad x = 1, 2, \ldots. \qquad (7.52)$$

It tends to the ordinary logarithmic distribution as $\lambda \to 1$. Tripathi and Gupta gave recurrence formulas for the probabilities and factorial moments, some simple estimators of the parameters, minimum χ^2 estimators, and MLEs. They also suggested a graphical selection procedure.

A quite different "generalized" logarithmic distribution was developed by Tripathi and Gupta (1988). It is a limiting form of their "generalized" negative binomial distribution and has a complicated pmf. The authors did not give moment properties. They refitted the data on entomologists' papers using a form of maximum likelihood.

Many researchers have used the word *generalized* to describe a distribution with a pgf of the form $G_1(G_2(z))$. More precisely, it is said to be an $\mathcal{F}_1$ *distribution generalized* by the *generalizing* $\mathcal{F}_2$ *distribution* where $G_1(z)$ is the pgf for $\mathcal{F}_1$ and $G_2(z)$ is the pgf for $\mathcal{F}_2$. Distributions with pgf's of the form

$$G(z) = \frac{\ln[1 - \theta g(z)]}{\ln(1 - \theta)} \qquad (7.53)$$

are generalized logarithmic distributions in this sense.

Panaretos (1983a) used this terminology to describe the distribution with pgf (7.53), where $g(z)$ is the pgf of a Pascal distribution. Medhi and Borah (1984) took $g(z) = (z + cz^m)/(1 + c)$, where $m > 1$ is an integer. In Xekalaki and Panaretos (1989) the generalizing distribution has the pgf $\sum_{i=1}^{\infty} q_i z^i / \sum_{i=1}^{\infty} q_i$, giving

$$G(z) = \frac{\ln[1 - \lambda(q_1 z + q_2 z^2 + q_3 z^3 + \cdots)]}{\ln[1 - \lambda(q_1 + q_2 + q_3 + \cdots)]}; \qquad (7.54)$$

The authors used an inverse cluster sampling model to derive the distribution.

7.1.12 Other Related Distributions

The distribution of the sum (X_n) of n independent rv's each having the same logarithmic distribution (7.1) has the pgf

$$G(z) = \left[\frac{\ln(1-\theta z)}{\ln(1-\theta)}\right]^n. \tag{7.55}$$

From this it follows that $\Pr[X_n = x]$ is proportional to the coefficient of z^x in $[-\ln(1-\theta z)]^n$, giving

$$\Pr[X_n = x] = \frac{n!|s(x,n)|\theta^x}{x![-\ln(1-\theta)]^n}, \qquad x = n, n+1, \ldots, \tag{7.56}$$

where $s(x,n)$ is the Stirling number of the first kind, with arguments x and n; see Section 1.1.3. The distribution of X_n is called the *Stirling distribution of the first kind*. The mean and variance are

$$\mu = \frac{na\theta}{1-\theta} \qquad \text{and} \qquad \mu_2 = \frac{na\theta(1-a\theta)}{(1-\theta)^2}, \tag{7.57}$$

where $a = [-\ln(1-\theta)]^{-1}$. Patil and Wani (1965b) state the following properties of the distribution:

1. If $\theta < 2n^{-1}$, the distribution has a unique mode at $x = n$ (the smallest possible value), and the values of $\Pr[X_n = x]$ decrease as x increases.
2. If $\theta = 2n^{-1}$, there are two equal modal values at $x = n$ and $x = n + 1$.
3. If $\theta > 2n^{-1}$, the value of $\Pr[X_n = x]$ increases with x to a maximum (or pair of equal maxima) and thereafter decreases as x increases; see also Sibuya (1988).

From the definition of this distribution, it is clear that as n tends to infinity the standardized distribution corresponding to (7.56) tends to the unit-normal distribution. Douglas (1971) and Shanmugam and Singh (1981) have studied properties and estimation for the distribution. A good general discussion is in Berg (1988a).

Cacoullos (1975) introduced a multiparameter extension that was obtained from the convolution of a number of logarithmic distributions that had been left truncated at different points.

The multivariate Ewens distribution arises in genetics, where it provides the null-hypothesis distribution of allele frequencies for a non-Darwinian theory of evolution and is known as Ewens sampling formula. It gives the partition distribution of a sample of n genes into allelic types when there are no selective differences between types. A conditional univariate distribution of the Ewens distribution is included here because of its relationship to the Stirling distribution of the first kind.

Consider a sample of size n taken from an infinite collection of items that belong to various distinguishable species. The assumption is that the species

Table 7.5 Hypothetical Ewens Sample of Size n

Number of Items per Species, j	Number of Species, $C_j(n)$	Total Number of Items, $j\,C_j(n)$
1	98	98
2	15	30
3	4	12
>3	0	0
	$117 = k$	$140 = n$

have random relative frequencies $P_1, P_2, \ldots$, where

$$\begin{aligned} P_1 &= W_1, \\ P_i &= (1 - W_1)(1 - W_2)\cdots(1 - W_{i-1})W_i, \qquad i = 2, 3, \ldots, \end{aligned} \tag{7.58}$$

where $0 < P_i < 1$, $\sum_{i=1}^{\infty} P_i = 1$, and the $W_1, W_2, \ldots$ are iid random variables with pdf $\theta(1-x)^{\theta-1}$, $0 < x < 1$, $0 < \theta < \infty$. Let $C_j(n)$ be the number of species represented by j items in a sample of size n; then $\sum_{i=1}^{\infty} jC_j(n) = n$. The number of distinct species in the sample is $K_n = \sum_{i=1}^{\infty} C_j(n)$; see Table 7.5.

Ewens (1972) obtained the multivariate distribution of the vector $[C_1(n), C_2(n), \ldots]$. The conditional distribution of K_n is

$$\Pr[K_n = k] = \frac{\theta^k(\theta-1)!\,|s(n,k)|}{(\theta+n-1)!}, \qquad k = 1, 2, \ldots, n. \tag{7.59}$$

This is a weighted Stirling distribution of the first kind. The mean and variance are

$$E[K_n] = \sum_{i=0}^{n-1} \frac{\theta}{\theta+i} \qquad \text{and} \qquad \mathrm{Var}(K_n) = \sum_{i=0}^{n-1} \frac{\theta}{(\theta+i)^2}. \tag{7.60}$$

The unimodality of a number of distributions with pmf's involving Stirling numbers of the first and second kinds was explored by Sibuya (1988).

The extended Stirling family of Nishimura and Sibuya (1997) is formed by replacing the Stirling numbers in the Stirling distributions of the first and second kinds by the Stirling–Carlitz polynomials of the first and second kinds (Charalambides and Singh, 1988; Branson, 2000; Charalambides, 2002). These are $R_1(n, m, t)$ and $R_2(n, m, t)$, where

$$(z+t)_n \equiv \sum_{m=0}^{m} R_1(n, m; t) z^m,$$

$$(z+t)^n \equiv \sum_{m=0}^{m} R_2(n,m;t) z^{(m)},$$

and $(z)_n = (z+n-1)!/(z-1)!$ (Pochhammer's symbol for a rising factorial) and $z^{(m)} = z!/(z-m)!$ (a falling factorial). Nishimura and Sibuya distinguished eight subfamilies:

$$\begin{array}{ll}
\text{Str1F} & \Pr[X=x] = R_1(n,x;\tau)\theta^x/(\theta+\tau)_n, \\
& \qquad x = 0,1,2,\ldots,n \\
\text{Str1W} & \Pr[X=x] = R_1(x-1,k-1;\tau)\theta^k/(\theta+\tau)_x, \\
& \qquad 1 \le k \le x < \infty \\
\text{Str1C} & \Pr[X=x] = R_1(x-1,x-k;\tau)(x-1+\tau)\theta^{x-k}/(\theta+\tau)_x, \\
& \qquad k \le x < \infty \\
\text{Str1|} & \Pr[X=x] = R_1(x,k;\tau)k!(x!)^{-1}(1-\theta)^\tau\theta^x/[-\log(1-\theta)]^k x!, \\
& \qquad 0 < k \le x < \infty
\end{array}$$

and

$$\begin{array}{ll}
\text{Str2F} & \Pr[X=x] = R_2(n,x;\tau)m^{(x)}/(m+\tau)^n, \\
& \qquad x = 0,1,\ldots,\min(n,m) \\
\text{Str2W} & \Pr[X=x] = R_2(x-1,k-1;\tau)m^{(k)}/(m+\tau)^x, \\
& \qquad 1 \le k \le x < \infty \\
\text{Str2C} & \Pr[X=x] = R_2(x-1,x-k;\tau)(x-k+\tau)m^{(x-k)}/(m+\tau)^x, \\
& \qquad k \le x < k+m \\
\text{Str2|} & \Pr[X=x] = R_2(x,k;\tau)k!(x!)^{-1}e^{-\tau\theta}\theta^x/(e^\theta-1)^k, \\
& \qquad 0 < k \le x < \infty
\end{array}$$

The conditions on the parameters are $0 \le \tau < \infty$ and

$$\begin{array}{rl}
\text{Str1F, Str1W, Str1C, Str2|:} & 0 < \theta < \infty \\
\text{Str1|:} & 0 < \theta < 1 \\
\text{Str2F, Str2W, Str2C:} & m \in \mathbb{Z}^+
\end{array}$$

there is also the possibility for Str2W that $m \in \mathbb{R}$ such that $n-1 < m < \infty$ and for Str2C that $m \in \mathbb{R}$ such that $k-1 < m < \infty$. Nishimura and Sibuya gave the pgf's and provided random-walk and other models for these distributions.

A distribution obtained by assigning a lognormal distribution to the parameter of a Poisson distribution is termed a *discrete lognormal distribution* (Anscombe, 1950). This mixed Poisson distribution is mentioned here because a zero-truncated form is an important competitor of the log-series distribution as a model for the distribution of observed abundancies of species and similar phenomena. Bliss (1965) made comparisons of the fidelity with which the two distributions represented five sets of data from moth-trap experiments; see also Preston (1948) and Cassie (1962). He found that in each case the truncated discrete lognormal distribution gave the better fit (as judged by χ^2-probabilities), though both distributions gave acceptable representations (again judged by a χ^2-test). If we suppose that the Poisson parameter λ is distributed lognormally with expected value ξ and standard deviation η, then the mixed distribution has pmf

$$\Pr[X = x] = \int_0^\infty \frac{e^{-\lambda}\lambda^{x-1}}{x!} \cdot \frac{1}{[2\pi \ln(a)]^{1/2}} \exp\left[\frac{[\ln(\lambda) - a]^2}{2\ln(a)}\right] d\lambda. \tag{7.61}$$

The mean and variance are

$$\mu = \exp\left(\xi + \tfrac{1}{2}\eta^2\right), \qquad \sigma^2 = \mu + \mu^2[\exp(\eta^2 - 1)]. \tag{7.62}$$

A summary of its properties appears in Shaban (1988). It is computationally difficult to fit to data using maximum likelihood. Shaban found a way to compute the first derivatives of the likelihood function. Weems and Smith (2004) have devised a method for fitting it by maximum likelihood in a mixed Poisson regression context.

Darwin (1960) obtained an ecological distribution (with finite support) from the beta–binomial distribution by a process analogous to that used by Fisher when obtaining the logarithmic distribution from the negative binomial.

Shenton and Skees (1970) found that the logarithmic distribution, depending as it does on only one parameter, was inadequate as a descriptor of storm duration measured in discrete units of time. They suggested (on an empirical basis) two J-shaped variants:

1:
$$\Pr[X = x] = a(1-p)p^{x-1} + \frac{(1-a)\theta^x}{-x\ln(1-\theta)},$$
$$0 \le p < 1, \qquad 0 \le \theta < 1, \qquad x = 1, 2, \dots. \tag{7.63}$$

2:
$$\Pr[X = 1] = 1 - \frac{a\theta}{b+1} \ge 0,$$
$$\Pr[X = x] = a\theta^{x-1}[(x+b-1)^{-1} - \theta(x+b)^{-1}],$$
$$0 < a, \qquad -1 < b, \qquad 0 < \theta < 1, \qquad x = 2, 3, \dots. \tag{7.64}$$

They noted that the latter distribution involves negative mixing and considered that it fit their data "rather well."

The *logarithmic distribution of order k*, with pgf

$$G(z) = (k\ln p)^{-1} \ln\left(\frac{1 - z + qp^k z^{k+1}}{1 - pz}\right), \tag{7.65}$$

where $0 < p < 1$, $q = 1 - p$, k an integer, was introduced by Hirano et al. (1984); see Section 10.7.4. A second logarithmic distribution of order k with pgf

$$G(z) = \frac{-\ln[1 - \theta(z + z^2 + \cdots + z^k)/(1 + \theta k)]}{\ln(1 + \theta k)} \tag{7.66}$$

was studied by Panaretos and Xekalaki (1986a); see Section 10.7.4 also. For both logarithmic distributions of order k, taking $k = 1$ gives the ordinary logarithmic distribution.

The last decade has seen much research concerning q-series analogs of the classical discrete distributions; see Section 10.8. The Euler distribution (Section 10.8.2) was so named by Benkherouf and Bather (1988) in their solution of an oil exploration problem. It is an infinitely divisible q-analog of the Poisson distribution. Given a Poisson distribution for the number of clusters, the Euler distribution is the outcome when the cluster size distribution has the pgf

$$G(z) = \frac{\sum_{i\geq 1}\{p^i z^i/[i(1-q^i)]\}}{\sum_{i\geq 1}\{p^i/[i(1-q^i)]\}}, \tag{7.67}$$

$0 < p < 1$, $0 < q < 1$, p and q unrelated. This cluster size distribution is a q-series analog of the logarithmic distribution. C. D. Kemp (1997) obtained a different q-analog of the logarithmic distribution (Section 10.8.2) with pgf

$$G(z) = \frac{z\ {}_2\phi_1(q, q; q^2; q, \theta z)}{{}_2\phi_1(q, q; q^2; q, \theta)}. \tag{7.68}$$

A discrete survival distribution with pmf

$$\Pr[X = x] = \frac{A\theta^{x-1}}{x^{\beta}}, \qquad x = 1, 2, \ldots,$$

where A is a normalizing constant, was investigated by Kulasekera and Tonkyn (1992). Kemp (1995) called this a polylogarithmic distribution. It exists when $0 < \theta < 1$, $\beta \leq 1$ and when $0 < \theta \leq 1$, $1 < \beta$ and becomes the logarithmic distribution when $\beta = 1$. It is a member of the Zipf family; see Section 11.2.20.

7.2 LAGRANGIAN DISTRIBUTIONS

Lagrangian expansions for the derivation of expressions for the probabilities of certain discrete distributions have been used for many years. Early examples include Otter's multiplicative process (Section 7.2.1) and Haight and Breuer's (1960) treatment of the Borel and the Tanner–Borel distributions (Section 7.2.2). The Consul and the Geeta distributions are discussed in Sections 7.2.3 and 7.2.4.

The potential of this technique for deriving distributions and their properties has been systematically explored by Consul, Shenton, Janardan, and their co-workers, beginning with a group of key papers in the early 1970s; see Section 7.2.5. Research concerning these distributions continues to be prolific and there is a large literature. Here we have space to mention only the most important of these distributions and the key references.

The *Lagrangian Poisson distribution* (Section 7.2.6) was obtained by Consul and Jain (1973a) as a limiting form of the *Lagrangian negative binomial distribution*. Its very close relationship to a shifted Borel–Tanner distribution has been increasingly appreciated. Consul's (1989) book on the distribution highlights how intensively it has been studied for its many properties and its various modes of genesis.

The *Lagrangian binomial distribution* was obtained by Mohanty (1966) as the distribution of the number of failures x encountered in getting $\beta x + n$ successes given a sequence of independent Bernoulli trials. It was given a queueing process interpretation by Takács (1962) and Mohanty (1966). It is very closely related to the *Lagrangian negative binomial distribution* of Jain and Consul (1971); see Section 7.2.7. The *Lagrangian logarithmic distribution* (Section 7.2.8) was derived from the Lagrangian negative binomial distribution by Jain and Gupta (1973).

The use of Lagrange's second expansion enabled Janardan and Rao (1983) to create *Lagrangian distributions of the second kind* (Section 7.2.9). Recent work by Janardan (1997) and Consul and Famoye (2001) has extended this family.

7.2.1 Otter's Multiplicative Process

Consider the multiplicative process corresponding to the equation

$$\mathcal{G}(z, w) \equiv zf(w) - w = 0, \tag{7.69}$$

where $f(w)$ is the pgf of the number of segments from any vertex in a rooted tree with numbered vertices and $w = \mathcal{P}(z)$ is the pgf for the number of vertices in the rooted tree. This scenario was first studied by Otter (1949). He showed that the number of vertices that are n segments removed from the root can be interpreted as the number of members in the nth generation of a branching process and hence that his process has applications in the study of population growth, in the spread of rumors and epidemics, and in nuclear chain reactions.

Otter took as his first example

$$f(w) = p_0 + p_1 w + p_2 w^2. \tag{7.70}$$

This yields

$$w = \mathcal{P}(z) = \frac{1 - p_1 z - [(1 - p_1 z)^2 - 4p_0 p_2 z^2]^{1/2}}{2p_2 z}. \tag{7.71}$$

The *Haight distribution* (Haight, 1961a) is the outcome when $f(w) = 1 + P - Pw$.

Kemp and Kemp (1969a) showed that when $f(w)$ is the pgf for a binomial distribution with pgf

$$f(w) = (p + qw)^2 \tag{7.72}$$

we have $p_1^2 = 4p_0p_2$, and hence Otter's multiplicative process gives a *lost-games distribution* (Section 11.2.10) with pgf

$$w = \mathcal{P}(z) = \frac{[1 - (1 - 4pqz)^{1/2}]^2}{4q^2z}. \tag{7.73}$$

Other models that are based on Otter's process and lead to particular lost-games distributions were investigated by Kemp and Kemp (1969a).

Neyman and Scott (1964) recognized the relevance of Otter's multiplicative process to their study of stochastic models for the total number of individuals infected during an epidemic started by a single infective individual; see also Neyman (1965). Their epidemiological model also uses (7.72).

For his second example Otter took

$$f(w) = e^{\lambda(w-1)} \tag{7.74}$$

and found that for $\lambda < 1$, $w = \mathcal{P}(z) = \sum_x P_x z^x$, where

$$P_x = \frac{e^{-x\lambda}(x\lambda)^{x-1}}{x!}, \qquad x = 1, 2, \ldots; \tag{7.75}$$

this is the Borel distribution (Section 7.2.2).

Restating (7.69) in terms of u and $g(z)$ [instead of z and $f(w)$] shows that it is the defining equation for basic Lagrangian distributions of the first kind (BLD_1, see Section 2.5).

Otter proved that in his general case where $f(w) = \sum_{j=0}^{\infty} p_j w^j$ the probabilities for the pgf $u = \mathcal{P}(z) = \sum_x P_x z^x$ can be obtained as

$$P_x = \frac{1}{x!}\left[\frac{\partial^{x-1}}{\partial w^{x-1}}[f(w)]^x\right]_{w=0}.$$

In the more modern notation this becomes

$$P_x = \frac{1}{x!}\left[\frac{\partial^{x-1}}{\partial z^{x-1}}[g(z)]^x\right]_{z=0}. \tag{7.76}$$

Examples of BLD_1 include

1. the *Borel distribution* (Section 7.2.2), where $g(z) = e^{a(z-1)}$,
2. the *Consul distribution* (Section 7.2.3), where $g(z) = (1 - \theta + \theta z)^m$, and
3. the *Geeta distribution* (Section 7.2.4),
   where $g(z) = (1 - \theta)^{m-1}/(1 - \theta z)^{m-1}$.

The *geometric distribution*, with $g(z) = 1 - p + pz$, is a special case of the Borel distribution.

7.2.2 Borel Distribution

The *Borel–Tanner distribution* (Tanner–Borel distribution) of Tanner (1953) describes the distribution of the total number of customers served before a queue vanishes given a single queue with random arrival times of customers (at constant rate ℓ) and a constant time (β) occupied in serving each customer. We suppose that the probability of arrival of a customer during the period $(t, t + \Delta t)$ is $\ell \Delta t + o(\Delta t)$ and that the probability of arrival of two or more customers in this period is $o(\Delta t)$. If there are initially n customers in the queue, then the probability that the total number (Y) of customers served before the queue vanishes is equal to y is

$$\Pr[Y = y] = \frac{n}{(y-n)!} y^{y-n-1} (\ell\beta)^{y-n} e^{-\ell\beta y}, \qquad y = n, n+1, \ldots. \tag{7.77}$$

The case $n = 1$ gives the *Borel distribution*; this was obtained by Borel (1942). The parameters ℓ and β appear only in the form of their product $\ell\beta$. It is convenient to use a single symbol for this product and to put $\ell\beta = a$, say, giving

$$\Pr[Y = y] = \frac{n}{(y-n)!} y^{y-n-1} a^{y-n} e^{-ay}, \qquad y = n, n+1, \ldots. \tag{7.78}$$

For (7.78) with $n = 1$ to represent a proper distribution, it is necessary to have $0 < a < 1$. If $a < 0$, the "probabilities" change sign, while if $a > 1$, $\sum_{x=n}^{\infty} \Pr[Y = y] < 1$.

In the classification scheme of Section 2.5, the Borel distribution is a basic Lagrangian distribution of the first kind (i.e., a BLD_1), with $g(z) = e^{a(z-1)}$. The Tanner–Borel distribution is an n-fold convolution of the Borel distribution and so is a delta Lagrangian distribution of the first kind, DLD_1.

Let $a(b)$ be the solution of the equation $b = ae^{-a}$. Using this inverse function, Haight and Breuer (1960) were able to show that the pgf for (7.78) can be expressed as

$$\begin{aligned} H(z) = \left[\frac{a(bz)}{a(b)}\right]^n &= z^n e^{na(bz) - na(b)} \\ &= z^n e^{n\ell\beta\{[H(z)]^{1/n} - 1\}}. \end{aligned} \tag{7.79}$$

Clearly $H(z)/z^n$ is a Poisson–stopped sum (generalized Poisson) distribution. The pgf of the generalizing distribution is $[H(z)]^{1/n}$, that is, the generalizing distribution is a Borel distribution.

Haight and Breuer found from (7.78) that

$$H'(1) = n + aH'(1),$$

$$H''(1) = n(n-1) + naH'(1) + \frac{a[H'(1)]^2}{n} + aH''(1),$$

and so

$$\mu = \frac{n}{1-a} \qquad \text{and} \qquad \mu'_{[2]} = \frac{n(n-1)}{1-a} + \frac{n^2 a}{(1-a)^2} + \frac{na}{(1-a)^3}, \tag{7.80}$$

that is,

$$\mu_2 = \frac{na}{(1-a)^3}. \tag{7.81}$$

Haight and Breuer remarked that the moment properties can also be obtained by successive differentiation of

$$K(t) = \ln G(e^t) = n(t-a) + na[G(e^t)]^{1/n}, \tag{7.82}$$

where $K(t)$ is the cumulant generating function and $G(e^t)$ is the mgf.

The modal value lies approximately between $k-1$ and k, where

$$ae^{1-a} = (k-n)k^{n+1/2}(k-1)^{-n-3/2} \tag{7.83}$$

and $n + 1 \le k \le 2n^2/3 + 3n$.

Tables of the cdf $\Pr[Y \le y]$ to five decimal places were given in Haight and Breuer (1960) and in Owen (1962). One of the problems of tabulating the distribution is that it has a very long tail except when a is small.

The distribution is a generalized power series distribution (Section 2.2), and hence the maximum-likelihood equation for a, assuming n is known, is the first-moment equation, giving

$$\hat{a} = \frac{\bar{x} - n}{\bar{x}}. \tag{7.84}$$

7.2.3 Consul Distribution

The *Consul distribution* is a BLD_1 distribution (Section 2.5) with $g(z) = (1 - \theta + \theta z)^m$, where m is a positive integer [see Consul and Shenton (1975) and Consul (1983)]. The pmf is

$$\Pr[X = x] = \frac{1}{x}\binom{mx}{x-1}\left(\frac{\theta}{1-\theta}\right)^{x-1}(1-\theta)^{mx}, \qquad x = 1, 2, \ldots. \tag{7.85}$$

Consul and Shenton (1975) showed that all basic Lagrangian distributions of the first kind are closed under convolution and that their first four cumulants κ_r, $r = 1, 2, 3, 4$, are

$$\begin{aligned}
\kappa_1 &= \mu = \left[\frac{1}{1-g'(z)}\right]_{z=1}, \\
\kappa_2 &= G_2\mu^3, \\
\kappa_3 &= G_3\mu^4 + 3G_2^2\mu^5, \\
\kappa_4 &= G_4\mu^5 + 10G_3G_2\mu^6 + 15G_3^2\mu^7,
\end{aligned} \tag{7.86}$$

where G_r is the rth cumulant of the distribution with pgf $g(z)$. Consul and Shenton also gave formulas for the higher cumulants and for the uncorrected moments, and they interpreted these distributions as busy-period distributions in queueing theory.

The moments of a BLD_1 can be obtained either from the cumulants or from the factorial moments. The first two factorial moments are

$$\mu = \frac{1}{1-g'}, \qquad \mu'_{[2]} = \frac{2g'}{(1-g')^2} + \frac{g''}{(1-g')^3} \tag{7.87}$$

and hence

$$\mu_2 = \frac{g'}{(1-g')^2} + \frac{g''}{(1-g')^3}, \tag{7.88}$$

where $g' = [dg(z)/dz]_{z=1}$, $g'' = [d^2g(z)/dz^2]_{z=1}$, and so on.

For the Consul distribution (7.87) and (7.88) become

$$\mu = \frac{1}{1-m\theta}, \qquad \mu_2 = \frac{m\theta(1-\theta)}{(1-m\theta)^3}. \tag{7.89}$$

The delta Consul distribution is an n-fold convolution of Consul distributions. The pmf is

$$\Pr[X=x] = \frac{n}{x}\binom{mx}{x-n}\left(\frac{\theta}{1-\theta}\right)^{x-n}(1-\theta)^{mx}, \qquad x = n, n+1, \ldots, \tag{7.90}$$

and the mean and variance are $\mu = n/(1-m\theta)$ and $\mu_2 = nm\theta/(1-m\theta)^3$.

7.2.4 Geeta Distribution

The *Geeta distribution* is both a BLD_1 Lagrangian-type distribution (Section 2.5), with $g(z) = (1-\theta)^{m-1}/(1-\theta z)^{m-1}$, and a modified PSD (Section 2.2.2). It has support $1, 2, 3, \ldots$ and pmf

$$\Pr[X=x] = \frac{1}{\beta x - 1}\binom{\beta x - 1}{x}\theta^{x-1}(1-\theta)^{\beta x - x}, \tag{7.91}$$

where $0 < \theta < 1$ and $1 < \beta < \theta^{-1}$; see Consul (1990b). The mean is infinite if $\beta\theta \geq 1$. As $\beta \to 1$, we find that $\Pr[X=1] \to 1$. For $\beta\theta < 1$ the mean and variance are

$$\mu = \frac{1-\theta}{1-\beta\theta} \qquad \text{and} \qquad \mu_2 = \frac{(\beta-1)\theta(1-\theta)}{(1-\beta\theta)^3}. \tag{7.92}$$

By reparameterizing with $\theta = (1-\mu)/(1-\mu\beta)$, the distribution can be shown to be a location parameter distribution in the sense introduced into the literature by Consul (1990c). The pmf becomes

$$\Pr[X = x] = \frac{(\beta x - 2)!}{x!(\beta x - x - 1)!}\left(\frac{\mu - 1}{\mu\beta - 1}\right)^{x-1}\left(\frac{\mu\beta - \mu}{\mu\beta - 1}\right)^{\beta x - x},$$
$$x = 1, 2, 3, \ldots, \qquad (7.93)$$

where the mean $\mu > 1$ and $\beta > 1$. The variance $\mu_2 = \mu(\mu - 1)(\beta\mu - 1)/(\beta - 1)$; this clearly increases with μ for fixed β and decreases to $\mu^2(\mu - 1)$ as β increases for fixed μ. The general recurrence formulas for the moments of Lagrangian distributions of the first kind simplify and enable the higher moments to be obtained recursively. A numerical approach has indicated that these distributions are reversed J shaped with a mode at $x = 1$ and with the length and weight of the tail dependent on the values of β and θ.

Estimation using (1) moments, (2) sample mean and first frequency, (3) maximum likelihood, and (4) minimum-variance unbiased estimation was studied by Consul (1990a). Two modes of genesis (a two-urn model and a regenerative stochastic process) are given in Consul (1990b).

Consul (1991) put forward a birth-and-death process model and a branching process model for the distribution in his discussion on the evolution of surnames. He fitted real data using the Geeta distribution and compared the fit with those given by the discrete Pareto and Yule distributions.

The delta Geeta distribution is an n-fold convolution of Geeta distributions. The pmf is

$$\Pr[X = x] = \frac{n}{x}\binom{\beta x - n - 1}{x - n}\theta^{x-n}(1 - \theta)^{\beta x - x}, \qquad x = n, n + 1, \ldots, \qquad (7.94)$$

with $\mu = n(1 - \theta)/(1 - \beta\theta)$ and $\mu_2 = n(\beta - 1)\theta(1 - \theta)/(1 - \beta\theta)^3$.

For the basic Lagrangian Katz distribution, Consul and Famoye (1996) took $g(z) = (1 - \beta)^{b/\beta}(1 - \beta z)^{-b/\beta}$, giving the pmf

$$\Pr[X = x] = \frac{1}{x}\binom{bx/\beta + x - 2}{x - 1}\beta^{x-1}(1 - \beta)^{\beta x/b}, \qquad x = 1, \ldots. \qquad (7.95)$$

Taking $\beta = 1/n$ and letting $n \to \infty$ give a Borel distribution. When $0 < \beta = \theta < 1$ and $b = (m - 1)\theta$, it becomes a Consul distribution.

Janardan (1998, 1999) showed that his four-parameter generalized Pólya–Eggenberger (GPED) family of the first kind and his GPED family of the second kind both contain the Lagrangian Katz distribution as a special case. The pmf's, pgf's, and moments of both families were discussed in detail, other special cases were identified, and a method for fitting via maximum-likelihood estimation was described.

7.2.5 General Lagrangian Distributions of the First Kind

The generalization process for *general Lagrangian distributions of the first kind* (GLD_1) was clarified in two important papers by Consul and Shenton (1972, 1973); see also Consul (1983).

Suppose that $f(z)$ is another pgf for which

$$\left[\frac{\partial^{x-1}}{\partial z^{x-1}}\left([g(z)]^x \frac{\partial f(z)}{\partial z}\right)\right]_{z=0} \geq 0 \quad \text{for } x \geq 1. \tag{7.96}$$

[Note that this is not the same as Otter's $f(\cdot)$.] Then the pgf for the general Lagrangian distribution of the first kind is formed from $f(z)$ and $g(z)$, where $z = \ell(u)$ is the smallest root of $z = ug(z)$, as

$$f(z) = f(\ell(u)) = f(0) + \sum_{x>0} \frac{u^x}{x!}\left[\frac{\partial^{x-1}}{\partial z^{x-1}}\left([g(z)]^x \frac{\partial f(z)}{\partial z}\right)\right]_{z=0}. \tag{7.97}$$

The probabilities are

$$\begin{aligned}
&\Pr[X = 0] = f(0),\\
&\Pr[X = x] = \frac{1}{x!}\left[\frac{\partial^{x-1}}{\partial z^{x-1}}\left([g(z)]^x \frac{\partial f(z)}{\partial z}\right)\right]_{z=0}, \qquad x > 0;
\end{aligned} \tag{7.98}$$

see Section 2.5.

Consul and Shenton (1972) presented in table form the outcomes from more than a dozen combinations of specific pgf's for $g(z)$ and $f(z)$. It is convenient to express the following pmf's in terms of binomial coefficients, provided that $c!$ is taken to mean $\Gamma(c+1)$ when c is not an integer.

1. Consider first the "binomial–binomial" distribution. Let $g(z) = (q + pz)^m$, where $q = 1 - p$ and $mp < 1$, and suppose that $z = u \cdot g(z)$ such that $u = 0$ for $z = 0$ and $u = 1$ for $z = 1$. Then $f(z) = (q' + p'z)^n$, where $q' = 1 - p'$, expanded as a power series in u, is the pgf for the binomial–binomial Lagrangian distribution. We have

$$\begin{aligned}
\Pr[X = 0] &= f(0) = (q')^n,\\
\Pr[X = x] &= \frac{1}{x!}\left[\frac{d^{x-1}}{dz^{x-1}}[(q + pz)^{mx} np'(q' + p'z)^{n-1}]\right]_{t=0}\\
&= \frac{n}{x}(q')^n (pq^{m-1})^x \sum_{j=0}^{k} \binom{n-1}{j}\binom{mx}{x-j-1}\left(\frac{p'q}{pq'}\right)^{j-1}\\
&= \frac{n}{mx+1}\binom{mx+1}{x}(q')^n \left(\frac{p'q}{pq'}\right)(pq^{m-1})^x\\
&\quad \times {}_2F_1\left[1-x, 1-n; mx - x + 2; \frac{p'q}{pq'}\right], \qquad x = 1, 2, 3, \ldots,
\end{aligned} \tag{7.99}$$

where $k = \min(x-1, n-1)$. When $g(z) = 1$, the outcome is the usual binomial distribution if $0 < p' < 1$ and n is a positive integer; however, $g(z) = 1$ gives the negative binomial distribution if $q' = 1 + P$, $0 < P$, $n = -k < 0$.

2. The "binomial–Poisson" distribution with $g(z) = (q + pz)^m$, $mp < 1$, $f(z) = e^{M(z-1)}$ has the pmf

$$\Pr[X = x] = e^{-M}\frac{(Mq^m)^x}{x!}\,{}_2F_0\left[1 - x, -mx; \ ; \frac{p}{Mq}\right], \qquad x \geq 0. \tag{7.100}$$

3. The "binomial–negative binomial" distribution with $g(z) = (q + pz)^m$, $mp < 1$, $f(z) = (Q - Pz)^{-k}$ has the pmf (where $x \geq 0$)

$$\Pr[X = x] = \frac{\Gamma(k + x)}{x!\Gamma(k)}Q^{-k}\left(\frac{Pq^m}{Q}\right)^x {}_2F_1\left[1 - x, -mx; 1 - x - k; \frac{-pQ}{qP}\right]. \tag{7.101}$$

4. The "Poisson–binomial" distribution with $g(z) = e^{\theta(z-1)}$, $\theta < 1$, and $f(z) = (q + pz)^n$, has the pmf

$$\Pr[X = 0] = q^n,$$

$$\Pr[X = x] = \frac{(\theta x)^{x-1}}{x!}e^{-\theta x}npq^{n-1}\,{}_2F_0\left[1 - x, 1 - n; \ ; \frac{p}{\theta q x}\right], \qquad x > 0. \tag{7.102}$$

5. The "Poisson–Poisson" distribution with $g(z) = e^{\theta(z-1)}$, $f(z) = e^{M(z-1)}$, $\theta < 1$ has the pmf

$$\Pr[X = x] = \frac{M(M + \theta x)^{x-1}e^{-(M+\theta x)}}{x!}, \qquad x \geq 0. \tag{7.103}$$

6. The "Poisson–negative binomial" distribution with $g(z) = e^{\theta(z-1)}$, $\theta < 1$, $f(z) = (Q - Pz)^{-k}$ has the pmf

$$\Pr[X = 0] = Q^{-k},$$

$$\Pr[X = x] = \frac{(\theta x)^{x-1}}{x!}e^{-\theta x}kPQ^{-k-1}\,{}_2F_0\left[1 - x, 1 + k; \ ; \frac{-P}{\theta Q x}\right], \quad x \geq 1. \tag{7.104}$$

7. The "negative binomial–binomial" distribution with $g(z) = (Q - Pz)^{-k}$, $kP < 1$, $f(z) = (q + pz)^n$ has the pmf

$$\Pr[X = 0] = q^n,$$
$$\Pr[X = x] = npq^{n-1}\frac{\Gamma(kx + x - 1)}{x!\Gamma(kx)}\left(\frac{P}{Q}\right)^{x-1} Q^{-kx} \times {}_2F_1\left[1 - x, 1 - n; 2 - x - kx; \frac{-pQ}{Pq}\right], \qquad x \geq 1. \tag{7.105}$$

8. The "negative binomial–Poisson" distribution with $g(z) = (Q - Pz)^{-k}$, $kP < 1$, $f(z) = e^{M(z-1)}$ has the pmf

$$\Pr[X = x] = \frac{e^{-M}M^x}{x!} Q^{-kx}\, {}_2F_0\left[1 - x, kx; -; \frac{-P}{MQ}\right], \qquad x \geq 0. \tag{7.106}$$

9. The "negative binomial–negative binomial" distribution with $g(z) = (Q - Pz)^{-k}$, $kP < 1$, $f(z) = (Q' - P'z)^{-M}$ has the pmf

$$\Pr[X = 0] = (Q')^{-M},$$
$$\Pr[X = x] = (Q')^{-M}\left(\frac{P'}{Q'Q^k}\right)^x \frac{\Gamma(M + x)}{x!\Gamma(M)} \times {}_2F_1\left[1 - x, kx; 1 - M - x; \frac{PQ'}{P'Q}\right], \qquad x \geq 1. \tag{7.107}$$

Consider now the moment properties of general Lagrangian distributions. The first two factorial moments are

$$\mu_1' = \frac{f_1}{1 - g_1}, \qquad \mu_{[2]}' = \frac{f_2 + 2f_1g_1}{(1 - g_1)^2} + \frac{f_1g_2}{(1 - g_1)^3}, \tag{7.108}$$

whence

$$\mu = \frac{f_1}{1 - g_1}, \qquad \mu_2 = \frac{f_2 + f_1g_1 + f_1 - (f_1)^2}{(1 - g_1)^2} + \frac{f_1g_2}{(1 - g_1)^3}, \tag{7.109}$$

(Consul and Shenton, 1972), where f_r and g_r, $r = 1, 2$, are the rth derivatives of $f(z)$ and $g(z)$ with respect to z evaluated at $z = 1$.

The moments can be obtained straightforwardly via the cumulants. Let F_r be the rth cumulant for the pgf $f(z)$ as a function of z and let D_r be the rth cumulant for the basic Lagrangian distribution obtained from $g(z)$ [as in (7.86)]. Then

$$\begin{aligned}
\kappa_1 &= F_1 D_1, \\
\kappa_2 &= F_1 D_2 + F_2 D_1^2, \\
\kappa_3 &= F_1 D_3 + 3F_2 D_1 D_2 + F_3 D_1^3, \\
\kappa_4 &= F_1 D_4 + 3F_2 D_2^2 + 4F_2 D_1 D_3 + 6F_3 D_1^2 D_2 + F_4 D_1^4
\end{aligned} \tag{7.110}$$

(Consul and Shenton, 1975).

Consul and Shenton (1973) investigated the relationship of these distributions to queueing theory. They also studied their limiting forms. Under one set of limiting conditions all discrete Lagrangian distributions of the first kind tend to normality, and under another set of limiting conditions they tend to inverse-Gaussian distributions.

These distributions should not be confused with the quasi-binomial I and II distributions of Consul and Mittal (1975), Fazal (1976), Mishra and Sinha (1981), and Lingappaiah (1987). The binomial–binomial distribution belongs to the Gould family, whereas the quasi-binomial distributions belong to the Abel family; see Charalambides (1990) and Section 2.6.

The following theorems hold:

1. Distributions with $f(z) = z^n$ are n-fold convolutions of the basic Lagrangian-type distribution.
2. The general distribution with $f(z) = g(z)$ is the same as the corresponding basic Lagrangian-type distribution, except that it is shifted one step to the left.
3. The general Lagrangian distribution can be derived by randomizing the index parameter m in the distribution with pgf z^m according to another distribution with pgf $f(z)$.
4. All the general Lagrangian distributions corresponding to the same basic Lagrangian distribution are closed under convolution.

In Consul (1981) the author considerably widened the scope of Lagrangian distributions by removing the restriction that $g(z)$ and $f(z)$ be pgf's. Instead, $g(z)$ and $f(z)$ are assumed to be two functions that are successively differentiable, with $g(1) = f(1) = 1$, $g(0) \neq 0$, and $0 \leq f(0) < 1$. As an example Consul instanced the use of $g(z) = (1 - p + pz)^{3/2}$ and $f(z) = (1 - p + pz)^{1/2}$; this leads to a valid distribution.

The advantage of Consul's extended definition of a Lagrangian distribution is that it generates a wider class of distributions that encompasses not only Patil's generalized PSDs but also Gupta's modified PSDs. However, models that have been constructed for general Lagrangian distributions will not on the whole be valid for distributions in the extended class.

For weighted Lagrange distributions and their characterizations, see Janardan (1987).

Minimum-variance unbiased estimation for Lagrangian distributions has been examined by Consul and Famoye (1989).

The application of Lagrangian-type distributions in the theory of *random mappings* has been researched by Berg and Mutafchiev (1990) and Berg and Nowicki (1991).

Devroye (1992) has studied the computer generation of Lagrangian-type variables.

7.2.6 Lagrangian Poisson Distribution

The Lagrangian Poisson distribution is the most heavily researched and widely applied member of the GLD_1 family. Consul and his co-workers have studied it in considerable detail; see Consul (1989). It is sometimes called *Consul's generalized Poisson distribution*.

It is the "Poisson–Poisson" distribution (7.103). It can also be obtained by shifting the Tanner–Borel distribution (7.78) so that it has support $0, 1, 2, \ldots$, that is, by transforming to the rv $X = Y - n$. The usual notation has $\theta = an$ and $\lambda = a$. Once the distribution has been shifted to the origin, it is no longer necessary to have $n = \theta/\lambda$ an integer.

The pmf becomes

$$\Pr[X = x] = \frac{\theta(\theta + x\lambda)^{x-1} e^{-\theta - x\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots. \tag{7.111}$$

When the parameter λ is set linearly proportional to the parameter θ, the model is referred to as "restricted" by Consul, and the parameterization $\theta = an$, $\alpha = a/\theta$ (i.e., $\alpha = \lambda/\theta$) is used.

There has been controversy in the literature regarding the parameter space. The distribution with full properties undoubtedly exists for $\theta > 0$, $0 < \lambda < 1$ (i.e., $0 < \alpha < \theta^{-1}$). Nelson (1975) commented that for $\lambda < 0$ (i.e., $\alpha < 0$) the probabilities sooner or later become negative. The response of Consul and Shoukri (1985) to this problem was to recommend the use of truncation when λ is negative, where $\max(-1, -\theta/m) < \lambda < 0$, and the support of the distribution is restricted to $0, 1, \ldots, m$, where m is the largest integer for which $\theta + m\lambda > 0$; that is, $m = [-\theta/\lambda]$, where $[\cdot]$ denotes the integer part. They imposed the arbitrary condition $m \geq 4$ to ensure that there are at least five classes in the resultant distribution. In their 1985 paper they made a detailed analysis of the numerical effect of such truncation; see also Section 9.1.1 of Consul's (1989) book.

When the distribution is obtained by truncation in this manner, the probabilities no longer sum exactly to unity. Many of its properties cease to hold exactly, even when the probabilities are normalized. For instance, the formulas for the moments become close approximations and infinite divisibility no longer holds.

Consider now the properties of the distribution when $0 < \theta$, $0 < \lambda < 1$ (i.e., $0 < \alpha < \theta^{-1}$). Let $t = ze^{\lambda(t-1)}$. Then the pgf has the form

$$\begin{aligned} G(z) = e^{\theta(t-1)} &= \exp\{\theta[ze^{\lambda(t-1)} - 1]\} \\ &= \exp\{\theta z[G(z)]^{\lambda/\theta} - \theta\}; \end{aligned} \tag{7.112}$$

from this the distribution is clearly seen to be a Poisson–stopped sum (generalized Poisson) distribution when $0 < \alpha < \theta^{-1}$ and so is infinitely divisible.

It is straightforward to show that the convolution of two such distributions with parameters (θ_1, λ) and (θ_2, λ) is another such distribution with parameters $(\theta_1 + \theta_2, \lambda)$. Consul (1975, 1989) characterized the distribution by the property that, if the sum of two independent rv's has a Lagrangian Poisson distribution, then each must have a Lagrangian Poisson distribution; Letac (1991), however, put forward a counterexample. A second characterization theorem in Consul (1975, 1989) concerns a Lagrangian Poisson random variate Z that is split into two components X and Y such that the conditional distribution $\Pr[X = x, Y = c - x | Z = c]$ is quasi-binomial with parameters (c, p, θ). The proof of this characterization would seem to depend on the previous characterization.

A useful recurrence relation for the probabilities is

$$\Pr[X = x] = \frac{(\theta + x\lambda)^{x-1} e^{-\lambda}}{(\theta + x\lambda - \lambda)^{x-2} x} \Pr[X = x - 1], \qquad x = 1, 2, \ldots, \tag{7.113}$$

where

$$\Pr[X = 0] = e^{-\theta}, \qquad \Pr[X = 1] = \theta e^{-\lambda-\theta}.$$

The distribution is unimodal. Recurrence-type formulas for the cumulative probabilities are given in Section 1.7 of Consul's (1989) book.

Consul and Jain (1973b) gave the following expressions for the first four moments:

$$\begin{aligned} \mu = \frac{\theta}{1-\lambda}, \qquad \mu_2 = \frac{\theta}{(1-\lambda)^3}, \qquad \mu_3 = \frac{\theta(1+2\lambda)}{(1-\lambda)^5}, \\ \mu_4 = \frac{3\theta^2}{(1-\lambda)^6} + \frac{\theta(1+8\lambda+6\lambda^2)}{(1-\lambda)^7}. \end{aligned} \tag{7.114}$$

The indices of skewness and kurtosis are

$$\begin{aligned} \beta_1 = \alpha_3^2 = \frac{\mu_3^2}{\mu_2^3} = \frac{(1+2\lambda)^2}{\theta(1-\lambda)}, \\ \beta_2 = \alpha_4 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{(1+8\lambda+6\lambda^2)}{\theta(1-\lambda)}. \end{aligned} \tag{7.115}$$

Consul and Jain also devised a method for calculating the higher moments using Stirling numbers. Consul and Shenton (1975) showed that the uncorrected moments satisfy

$$(1-\lambda)\mu'_{r+1} = \theta\mu'_r + \theta\frac{\partial\mu'_r}{\partial\theta} + \lambda\frac{\partial\mu'_r}{\partial\lambda}, \qquad r = 0, 1, 2, \ldots. \tag{7.116}$$

The central moments and the cumulants satisfy

$$\mu_{r+1} = \frac{r\theta}{(1-\lambda)^3}\,\mu_{r-1} + \frac{1}{1-\lambda}\left[\frac{\partial\mu_r(t)}{\partial t}\right]_{t=1}, \qquad r = 1, 2, 3, \ldots, \tag{7.117}$$

$$(1-\lambda)\kappa_{r+1} = \lambda\frac{\partial\kappa_r}{\partial\lambda} + \theta\frac{\partial\kappa_r}{\partial\theta}, \qquad r = 1, 2, 3, \ldots, \tag{7.118}$$

where $\mu_r(t)$ is μ_r with λ and θ replaced by λt and θt, respectively, and $\kappa_1 = \theta(1-\lambda)^{-1}$.

Gupta (1974) and Gupta and Singh (1981) found that

$$\mu_{r+1} = \left(\frac{\theta}{1-\alpha\theta}\right)\frac{\partial\mu_r}{\partial\theta} + r\mu_2\mu_{r-1}, \tag{7.119}$$

$$\mu'_{[r]} = \left(\frac{\theta}{1-\alpha\theta}\right)\frac{\partial\mu'_{[r]}}{\partial\theta} + \mu'_{[r]}(\mu'_{[1]} - r) \tag{7.120}$$

for the central and factorial moments of the restricted model. Further formulas for the uncorrected moments and for the factorial moments appear in Janardan (1984).

Negative moments for the unrestricted model have been studied by Consul and Shoukri (1986) [see also Section 3.8 in Consul (1989)]. Formulas for the incomplete moments are in Consul (1989, Section 3.10).

Limiting forms of the distribution include the normal and the inverse Gaussian (Consul, 1989). Consul (1989) also studied the distribution of the difference $X_1 - X_2$ of two Lagrangian Poisson variables with parameters (θ_1, λ) and (θ_2, λ). The left-truncated distribution was researched by Medhi (1975). Consul (1984) and Gupta and Gupta (1984) investigated the distribution of the order statistics of a sample whose *size* has a Lagrangian Poisson distribution.

Estimation for the Lagrangian Poisson distribution is described in Chapter 4 of Consul (1989). The distribution is a modified PSD, and hence when $\alpha = \lambda/\theta$ is known, the MLE $\hat{\theta}$ and the moment estimator $\tilde{\theta}$ of θ are

$$\hat{\theta} = \tilde{\theta} = \frac{\overline{x}}{1+\alpha\overline{x}}; \tag{7.121}$$

this is known to be negatively biased.

When estimation is for both parameters, the moment estimators are

$$\tilde{\theta} = \sqrt{\frac{m_1^3}{m_2}}, \qquad \tilde{\alpha} = \sqrt{\frac{m_2}{m_1^3}} - \frac{1}{m_1}, \tag{7.122}$$

where m_1 and m_2 are the first two sample moments. Kumar and Consul (1980) have derived expressions for the asymptotic biases, variances, and covariance of $\tilde{\theta}$ and $\tilde{\alpha}$.

The mean-and-zero-frequency estimators are

$$\theta^* = \ln\left(\frac{f_0}{N}\right), \qquad \lambda^* = 1 - \frac{\theta^*}{\bar{x}}, \tag{7.123}$$

where f_0/N is the observed relative frequency of zero.

Maximum-likelihood estimation by direct search of the likelihood surface is not straightforward when the sample variance is less than the sample mean. Certain properties and existence theorems for the maximum-likelihood estimators were obtained by Consul and Shoukri (1984). For generalized minimum chi-square estimation, see Section 4.7 in Consul (1989).

Shoukri and Consul (1987) gave a helpful account of modes of genesis leading to the distribution. These include the distribution as a limit of the "generalized negative binomial" distribution, as a Lagrangian probability distribution, as the distribution of the total progeny in some Galton–Watson-type branching processes, and as the outcome of a birth-and-death process. These authors drew attention to the following uses of the distribution: modeling the spread of ideas, rumors, fashions, and sales; the distribution of salespersons in "pyramid" dealerships; modeling population counts of biological organisms; thermodynamic processes; and the spread of burnt trees in forest fires. Further applications are cited in Chapter 5 of Consul (1989), where more than 50 data sets are fitted; in more than 20 of these the estimated value of λ is negative. It should be noted that many of the data sets are very short tailed, in some cases consisting entirely of counts of 0, 1, or 2.

The Lagrangian Poisson distribution has also been used to fit overdispersed and underdispersed count data in a regression context with one or more covariates (Consul and Famoye, 1992). Both maximum-likelihood estimation and the method of moments were used to estimate the parameters for four observed data sets to which other regression models had earlier been fitted. Approximate tests for the adequacy of the model were applied. In Famoye and Singh (1995) it was used to explore the relationship between chromosome aberrations and radiation dose in human lymphocytes.

Famoye (1993) used the restricted Lagrangian Poisson in a regression model; he discussed estimation of its parameters and approximate tests for its adequacy.

7.2.7 Lagrangian Negative Binomial Distribution

The binomial–binomial Lagrangian distribution of Consul and Shenton (1972) was mentioned in Section 7.2.5, Equation (7.99). The special case with $p = p'$ (and hence with $q = q'$) simplifies to

$$\Pr[X = x] = \frac{n}{n + mx}\binom{n + mx}{x} p^x q^{n+mx-x}, \tag{7.124}$$

where $x = 0, 1, \ldots$. The restrictions on the parameters are $0 < p < 1, q = 1 - p$, $n > 0$, $p < mp < 1$. This is the "generalized negative binomial" distribution of Jain and Consul (1971).

When $P = P'$ (and hence $Q = Q'$), the pmf for the Lagrangian negative binomial–negative binomial distribution (7.107) simplifies to

$$\Pr[X = x] = \frac{M}{M + kx + x}\frac{\Gamma(kx + M + x + 1)}{x!\Gamma(M + kx + 1)}\left(\frac{P}{Q}\right)^x Q^{-(M+kx)}, \qquad x \geq 0. \tag{7.125}$$

Setting $M = n$, $k + 1 = m$, $P/Q = p$ (and hence $1/Q = q$) in (7.125) gives (7.124). This explains the name *Lagrangian negative binomial distribution of the first kind*.

The first four moments are

$$\begin{aligned}
\mu &= np(1 - mp)^{-1}, \\
\mu_2 &= npq(1 - mp)^{-3}, \\
\mu_3 &= npq[q - p + mp(1 + q)](1 - mp)^{-5}, \qquad (7.126)\\
\mu_4 &= 3n^2p^2q^2(1 - mp)^{-6} + npq[1 - 6pq + 2np(4 - 9p + 4p^2) \\
&\qquad + m^2p^2(6 - 6p + p^2)](1 - mp)^{-7}.
\end{aligned}$$

Kumar and Consul (1979) investigated the negative moments.

When $k \to 0$, the Lagrangian negative binomial distribution tends to the negative binomial, and when $n = m$ in (7.124), it becomes the Consul distribution shifted to support $0, 1, \ldots$.

The distribution has received much attention; see, for instance, Gupta (1974) and Kumar and Consul (1979). Normal and inverse-Gaussian limiting forms were studied by Consul and Shenton (1973). Jain and Consul (1971) fitted the distribution to data using estimation by the method of moments, while minimum-variance unbiased estimation was investigated by Kumar and Consul (1980) and Consul and Famoye (1989).

Characterization theorems were obtained by Jain and Consul (1971), Consul (1974), and Consul and Gupta (1980). In particular, if X and Y are independent nonnegative integer-valued rv's and their sum $X + Y$ has a Lagrangian negative binomial distribution, then X and Y each have a Lagrangian negative binomial distribution.

A stochastic urn model for the distribution was devised by Famoye and Consul (1989). Consul (1989) examined the difference of two such Lagrangian negative binomial variables. In Section 2.6, we commented that the distribution belongs to Charalambides's (1986a) family of Gould distributions.

Famoye and Consul (1993) studied estimation for the zero-truncated Lagrangian negative binomial distribution using the method of moments, maximum-likelihood estimation, a method based on the sample mean and the ratio of frequencies, and a method based on the factorial moments and a ratio of frequencies.

Consul and Famoye (1995) reviewed applications of the (untruncated) distribution in queueing theory, branching processes, and in polymerization reaction in chemistry. The paper also discussed methods of estimating the parameters and compared estimates obtained using the truncated and the untruncated distributions. The asymptotic relative efficiencies for various estimation methods were compared in Famoye (1997). Simulation results favored estimation using the first two moments and proportion of zeroes.

Famoye (1995) examined the use of the distribution for predicting a count response variable affected by one or more explanatory variables. Two data sets were fitted using maximum likelihood and the method of moments; approximate tests for the adequacy of the model were applied.

7.2.8 Lagrangian Logarithmic Distribution

The *"generalized" logarithmic distribution* (*Lagrangian logarithmic distribution*) of Jain and Gupta (1973) is a limiting form as $n \to \infty$ of a zero-truncated Lagrangian negative binomial distribution. The pmf is

$$\Pr[X = x] = \frac{-c^x(1-c)^{x(\beta-1)}\Gamma(\beta x)}{\Gamma(x+1)\Gamma(\beta x - x + 1)\ln(1-c)}, \qquad x = 1, 2, \ldots, \qquad (7.127)$$

where $0 < c \le c\beta < 1$. The distribution can also be obtained by taking $f(z) = -\ln[1 + cz/(1-c)]/\ln(1-c)$ and $g(z) = (1 - c + cz)^m$ in (7.97). The ordinary logarithmic distribution is the special case $\beta = 1$.

Jain and Gupta showed that

$$\mu_1' = \frac{-c}{(1-\beta c)\ln(1-c)} \qquad \text{and} \qquad \mu_2' = \frac{-c(1-c)}{(1-\beta c)^3\ln(1-c)}. \qquad (7.128)$$

The incomplete moments were studied by Tripathi, Gupta, and Gupta (1986).

Gupta (1976) examined the distribution of the sum of a number of independent Lagrangian logarithmic distributions; a special case is the Stirling distribution of the first kind. A modified form of Jain and Gupta's "generalized" logarithmic distribution, with added zeroes, was studied by Jani (1986).

Jain and Gupta (1973) used the method of moments to fit (7.127) to Williams's data on numbers of papers by entomologists. Gupta (1977a) and Jani (1977) examined minimum-variance unbiased estimation for the parameters of the distribution. It was applied by Rao (1981) to the study of correlation between two types of children in a family. Mishra and Tiwary (1985) discussed various estimation methods, including maximum likelihood, mean-and-variance, and a noniterative procedure based on the first three moments. Goodness-of-fit test statistics based on the empirical distribution function have been researched by Famoye (2000).

The logconvexity of the probabilities and hence the infinite divisibility of the distribution when shifted to the support 0, 1, 2, ... were demonstrated by Hansen and Willekens (1990); these authors also discussed the use of the distribution in risk theory in a problem related to the total claim size up to time t [see also Hogg and Klugman (1984)].

7.2.9 Lagrangian Distributions of the Second Kind

The use of Lagrange's second expansion enabled Janardan and Rao (1983) to create *Lagrangian distributions of the second kind* (LD_2, see Section 2.5). Let $f(z)$ and $g(z)$ be functions of z such that $g(0) \neq 0$, $f(0) \geq 0$, $f(1) = g(1) = 1$. Then the second Lagrangian expansion gives

$$\frac{f(z)}{1 - zg'(z)/g(z)} = \sum_{x \geq 0} \frac{u^x}{x!} \left[\frac{\partial^x}{\partial z^x} \left\{ [g(z)]^x f(z) \right\} \right]_{z=0}. \tag{7.129}$$

Suppose also that $0 < g'(1) < 1$ and that

$$\left[\frac{\partial^x}{\partial z^x} \{ [g(z)]^x f(z) \} \right]_{z=0} \geq 0$$

when $x = 0, 1, \ldots$. Then the pmf for Lagrangian distributions of the second kind is

$$\begin{aligned} \Pr[X = 0] &= [1 - g'(1)] f(0), \\ \Pr[X = x] &= \left(\frac{1 - g'(1)}{x!} \right) \left[\frac{\partial^x}{\partial z^x} \{ [g(z)]^x f(z) \} \right]_{z=0}, \qquad x > 0. \end{aligned} \tag{7.130}$$

For a *basic Lagrangian distribution of the second kind* $f(z) = z$, and for a *delta Lagrangian distribution of the second kind* $f(z) = z^n$.

In Janardan and Rao (1983) $f(z)$ and $g(z)$ were assumed to be pgf's defined on the nonnegative integers with $g(0) \neq 0$. Janardan (1997) and Consul and Famoye (2001) have extended the family by modifying the conditions on $f(z)$ and $g(z)$; see also Janardan (2001).

A number of properties of LD_2 distributions were derived by Janardan and Rao (1983). Janardan (1997) showed that the class of modified PSDs is a subclass of LD_2. Consul and Famoye (2001) have derived formulas for the LD_2 moments and cumulants.

CHAPTER 8

Mixture Distributions

8.1 BASIC IDEAS

8.1.1 Introduction

The important class of distributions to be discussed in this chapter consists of *mixtures* of discrete distributions.

The revolution since the previous edition of this book in the speed, cost, and memory of computers has made possible complicated statistical techniques that were formerly not feasible.

This has led to an outpouring of research on mixtures.

The notion of *mixing* often has a simple and direct interpretation in terms of the physical situation under investigation.

The random variable concerned may be the result of the actual mixing of a number of different populations, such as the number of car insurance claims per driver, where the expected number of claims varies with category of driver.

Alternatively, it may come from one of a number of different unknown sources; a mixture rv is then the outcome of ascribing a probability distribution to the possible sources. Sometimes, however, "mixing" is just a mechanism for constructing new distributions for which empirical justification must later be sought.

At one time applications of mixture distributions were discussed in terms of modeling data (from biology, geology, sociology, commerce, etc.). Now the focus has widened to other kinds of statistical methodology:

- discriminant analysis and the classification of new observations (statistical pattern recognition and image enhancement),
- outlier robustness studies with an outlier comprising a component in a mixture model,
- cluster analysis by assignment of items in a data set to various subpopulations,

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

- latent structure models and factor analysis,
- Bayes and empirical Bayes estimation,
- kernel-based density estimation,
- variance component models,
- hierarchical generalized linear models in regression studies,
- random variate generation,
- approximation of the distribution of a test statistic.

The term *compounding* has often been used in place of "mixing," as in the first edition of this book. There is, however, an alternative usage of the term "compounding" to mean a random sum (using distribution $\mathcal{F}_A$) of iid random variables from distribution $\mathcal{F}_B$; see Section 8.3.1. To avoid confusion, we use the term "mixture" for the distributions discussed in this chapter.

The two important categories of mixtures of discrete distributions (finite mixtures and countable/continuous mixtures) are described in the next two sections.

8.1.2 Finite Mixtures

A *k-component finite-mixture distribution* is formed from k different component distributions with cumulative distribution functions (cdf's) $F_1(x), F_2(x), \ldots, F_k(x)$ with mixing weights

$$\omega_1, \omega_2, \ldots, \omega_k, \quad \text{where } \omega_j > 0, \qquad \sum_{j=1}^{k} \omega_j = 1,$$

by taking the weighted average

$$F(x) = \sum_{j=1}^{k} \omega_j F_j(x) \tag{8.1}$$

as the cdf of a new (mixture) distribution. It corresponds to the *actual mixing* of a number of different distributions and is sometimes called a *superposition*. In the theory of insurance ω_j, $j = 1, \ldots,$ is called the *risk function*.

It follows from (8.1) that, if the component distributions are defined on the nonnegative integers with

$$P_j(x) = F_j(x) - F_j(x-1),$$

then the mixture distribution is a discrete distribution with pmf

$$\Pr[X = x] = \sum_{j=1}^{k} \omega_j P_j(x). \tag{8.2}$$

The support of the outcome for this type of mixture is the union of the supports for the individual components of the mixture. The logarithmic-with-zeros distribution (Section 8.2.4) is an example of a mixture formed from two components. Here the support for the two components is disjoint, and the support for the outcome is $0, 1, \ldots$.

An important method for the computer generation of pseudorandom variates involves the *decomposition* of a distribution into components. This involves the representation of the target cdf as a finite mixture; see Peterson and Kronmal (1980, 1982) and Devroye (1986, pp. 66–75).

The concept of a finite mixture of discrete distributions has a long history going back to Pearson (1915). Recent books that deal with various aspects of finite mixtures include Lindsay (1995), Grandell (1997), Böhning (1999), and McLachlan and Peel (2000).

See Section 8.2 for further information on discrete distributions with a finite number of components.

8.1.3 Varying Parameters

A mixture distribution also arises when the cdf of a rv depends on the parameters $\theta_1, \theta_2, \ldots, \theta_m$ [i.e., has the form $F(x|\theta_1, \ldots, \theta_m)$] and some (or all) of those parameters *vary*. The new distribution then has the cdf

$$E[F(X|\theta_1, \ldots, \theta_m)],$$

where the expectation is with respect to the joint distribution of the k parameters that vary. This includes situations where the source of a rv is unknowable.

Suppose now that only one parameter varies (this will be the case in most instances in this volume). It is convenient to denote a mixture distribution of this type by the symbolic form

$$\mathcal{F}_A \bigwedge_{\Theta} \mathcal{F}_B,$$

where $\mathcal{F}_A$ represents the original distribution and $\mathcal{F}_B$ the mixing distribution (i.e., the distribution of Θ).

When Θ has a discrete distribution with probabilities p_i, $i = 0, 1, \ldots$, we will call the outcome a *countable mixture*; the cdf is

$$F(x) = \sum_{i \geq 0} p_i F_i(x), \tag{8.3}$$

where the probabilities p_i replace the weights ω_i in (8.1). The probability mass function of the mixture is

$$\Pr[X = x] = \sum_{i \geq 0} p_i P_i(x), \tag{8.4}$$

where $P_j(x) = F_j(x) - F_j(x - 1)$.

As an example consider a mixture of Poisson distributions for which the Poisson parameter $\Theta = \alpha V$, where α is constant and V has a logarithmic distribution with parameter λ. The outcome is represented by the symbolic form

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\alpha} \text{Logarithmic}(\lambda);$$

the corresponding pmf is

$$\Pr[X = x] = \sum_{j=1}^{\infty} e^{-\alpha j} \frac{(\alpha j)^x}{x!} \times \frac{\lambda^j}{j[-\ln(1-\lambda)]}, \qquad x = 0, 1, \ldots. \tag{8.5}$$

Consider, on the other hand, a mixture of logarithmic distributions with parameter $\Phi = \beta Y$, where β is constant and Y now has a Poisson distribution with parameter θ. This gives rise to a mixture distribution of the form

$$\text{Logarithmic}(\Phi) \bigwedge_{\Phi/\beta} \text{Poisson}(\theta)$$

with pmf

$$\Pr[X^* = x] = \sum_{j=0}^{\infty} \frac{\beta^x j^x}{[-\ln(1-\beta j)]x} \times \frac{e^{-\theta}\theta^j}{j!}, \qquad x = 1, 2, \ldots. \tag{8.6}$$

When the points of increase of the mixing distribution are continuous, we will call the outcome a *continuous mixture*. The cdf is obtained by integration over the mixing parameter Θ; if $H(\Theta)$ is the cdf of Θ, then the mixture distribution has cdf

$$F(x) = \int F(x|\theta)\, dH(\theta), \tag{8.7}$$

where integration is over all values taken by Θ. The support of the outcome of mixing processes such as these is the same as the support of the initial distribution.

Thus, for example,

$$\text{Poisson} \bigwedge_{\Theta} \text{Gamma}(\alpha, \beta)$$

means a mixture of Poisson distributions formed by ascribing the gamma distribution with probability density

$$p_\Theta(t) = \frac{t^{\alpha-1}\exp(-t/\beta)}{\beta^\alpha \Gamma(\alpha)}, \qquad 0 \le t, \qquad \alpha,\ \beta > 0,$$

to the parameter Θ of a Poisson distribution. This mixed Poisson distribution is of course a negative binomial distribution (see Section 5.3).

Discrete mixed distributions that have been obtained using countable or continuous mixing distributions are studied in Section 8.3.

The negative binomial distribution was obtained above as a gamma mixture of Poisson distributions. It is also a Poisson–stopped sum distribution. Many other discrete distributions are similarly both Poisson mixtures and Poisson–stopped sum distributions; see Section 9.3 et seq. Three important theorems that relate to this class of distributions are Lévy's theorem, Maceda's theorem, and Gurland's theorem; see Section 8.3.2.

8.1.4 Bayesian Interpretation

There is an interesting interpretation of mixtures via Bayes's theorem. From (8.4)

$$\sum_i \frac{p_i\, P_i(x)}{\Pr[X=x]} = 1.$$

So if p_i is the pmf for a discrete prior distribution, then $p_i\, P_i(x)/\Pr[X=x]$ can be regarded as the pmf for a posterior distribution. Furthermore from (8.7) the pmf for a mixture of discrete distributions formed using a continuous mixing distribution is

$$\Pr[X=x] = \int \Pr[X=x|\theta]h(\theta)\,d\theta,$$

where integration is over all values of θ; the probability density function

$$\frac{h(\theta)\Pr[X=x|\theta]}{\Pr[X=x]}$$

can be looked upon as a posterior density function for the prior density function $h(\theta)$.

8.2 FINITE MIXTURES OF DISCRETE DISTRIBUTIONS

8.2.1 Parameters of Finite Mixtures

Finite-mixture distributions arise in many probabilistic situations. Smith (1985), in a general overview of finite mixtures of continuous and discrete distributions, has listed the following fields of application:

Fisheries research, where the k components (categories) are different ages.

Sedimentology, where the categories are mineral types.

Medicine, where the categories are disease states.

Economics, where the categories are discontinuous forms of behavior.

Many other modeling applications of finite mixtures are referenced in Table 2.1.3 of Titterington, Smith, and Markov (1985, pp. 16–21). Titterington (1990) has commented on their use in speech recognition and in image analysis.

In Section 8.1.1 we commented on indirect applications for mixture distributions. Indirect contexts that involve finite mixtures include contaminated models for outliers, cluster analysis, latent-structure models, empirical Bayes methods, kernel density estimation, and the classification problem of identifying population membership.

Consider a k-component (finite) mixture of discrete distributions with pmf

$$\Pr[X = x] = \sum_{j=1}^{k} \omega_j P_j(x) \tag{8.8}$$

and let $g(X)$ be some function of X. It follows that

$$E[g(X)] = \sum_{j=1}^{k} \omega_j E[g(X|F_j(x))]; \tag{8.9}$$

in particular

$$\mu'_r(X) = E[X^r] = \sum_{j=1}^{k} \omega_j \mu'_r(X|F_j(x)), \tag{8.10}$$

$$\mu'_{[r]}(X) = E\left[\frac{X!}{(X-r)!}\right] = \sum_{j=1}^{k} \omega_j \mu'_{[r]}(X|F_j(x)), \tag{8.11}$$

and

$$\begin{aligned} G(z) = E[z^X] &= \sum_{j=1}^{k} \omega_j E[z^X|F_j(x)] \\ &= \sum_{j=1}^{k} \omega_j G_j(z), \end{aligned} \tag{8.12}$$

where $G_j(z)$ is the pgf of the jth component.

In most applications of finite mixtures the number of components k is quite small—from two to five, say. A special case with $k = 2$ is the formation of distributions with "added zeros" (see Section 8.2.3).

Most of the work on finite mixtures of discrete distributions has concerned mixtures of binomials and mixtures of Poissons; these researches are reviewed in Sections 8.2.5 and 8.2.6.

In Titterington (1990) mixture data are regarded as realizations of X, or alternatively as realizations of (X, Z) but with Z missing, where X is the mixture variable and Z the mixing variable. The author points out that Z can be treated as an indicator vector of dimension k, with $\omega_j = \Pr[Z = c_j]$ and c_j a k-vector with unity in position j and zeros elsewhere. The set of all parameters is then $\psi = (\omega, \theta)$, where ω is the set $\{\omega_j\}$ and θ is the set $\{\theta_j\}$ of all the parameters of the $F_j(x)$, $j = 1, \ldots, k$.

The parameters of a finite mixture fall therefore into three categories:

(i) There is k, the number of components. If k is not known, the problem of estimating it is usually quite difficult. Often the number of components is decided on an ad hoc basis because of the inherent problems of doing otherwise.

(ii) There are the mixing weights ω_j. Note that there are only $k - 1$ weight parameters to be estimated since their sum is necessarily unity.

(iii) There are the parameters of the component distributions.

The components need not all have the same form, though they usually do. Suppose that each has s parameters unrelated to those of any other component. The total number of unknown parameters possessed by the mixture distribution is therefore $k - 1 + ks$, assuming that k itself is known.

8.2.2 Parameter Estimation

It may not be possible to estimate all of these parameters, however. Consider, for instance, a mixture of two binomial distributions with parameters $n_1 = n_2 = 2$. We have

$$\begin{aligned}
\Pr[X = 0] &= \omega(1 - \pi_1)^2 + (1 - \omega)(1 - \pi_2)^2, \\
\Pr[X = 1] &= 2\omega\pi_1(1 - \pi_1) + 2(1 - \omega)\pi_2(1 - \pi_2), \\
\Pr[X = 2] &= \omega\pi_1^2 + (1 - \omega)\pi_2^2.
\end{aligned} \tag{8.13}$$

Because $\sum_i \Pr[X = i] = 1$, there are here only two independent equations in three unknowns, ω, π_1, π_2. Clearly the solution is not unique—the mixture is said to be unidentifiable.

Seminal in-depth studies of the problem of identifiability were Teicher (1960, 1961, 1963), Barndorff-Neilsen (1965), Yakowicz and Spragins (1968), and Chandra (1977). Sapatinas (1995) has concentrated on the identifiability of discrete mixtures.

Early references concerning the general theory of finite mixtures of discrete distributions were Blischke (1965) and Behboodian (1975). The papers by Shaked (1980) and Titterington (1990) are important.

McLachlan and Peel (2000) give a full account of the major issues involved in estimation from finite mixtures. Besides identifiability and number of components, the choice of estimation procedure and the properties of the resultant estimators need careful consideration.

Let us suppose that k is known, that there are $k(s + 1) - 1$ parameters to be estimated, and that the mixture is identifiable. Then, if the first $k(s + 1) - 1$

sample moments (or sample factorial moments) of the mixture distribution have been obtained, moment estimates of the parameters can be obtained by setting the sample moments equal to their expected values and solving the resultant equations. Note that either the central moments or the uncorrected moments can be used (or indeed the factorial moments, since the factorial moments are linear functions of the uncorrected moments). Titterington, Smith, and Makov (1985) drew attention to the problems inherent in moment estimation for finite mixtures, though they pointed out that there is a long history of its use. These authors, like Everitt and Hand (1981), discussed a number of other methods that are not maximum likelihood.

In the 1980s and 1990s much of the research on finite mixtures was concerned with mixtures of members of the linear exponential family. Titterington et al. described in detail the use of the EM (Expectation–Maximization) algorithm for maximum-likelihood estimation in the general case. They also discussed Bayesian methods, minimum-distance estimation based on the cdf, and minimum-distance estimators based on transforms.

Smith (1985) has tried to explain why there should be so many estimation approaches. In his discussion of maximum-likelihood methods he compared the merits of Newton–Raphson iteration, Fisher's scoring method, and the EM algorithm.

Titterington (1990) has stressed that the usual dependence structure associated with the rv's $Y = (X, Z)$ is not the only possibility. He uses the name "hidden multinomial" for the usual dependence structure; here the term hidden multinomial is motivated by regarding the "hidden" $Z_1, \ldots, Z_N$ (where N is the sample size) as a sample from a k-cell multinomial population. Alternatively, the $Z_1, \ldots, Z_N$ may form a stationary Markov chain on the state space $\{c_1, \ldots, c_k\}$; this is Titterington's "hidden Markov chain" dependence structure, with particular relevance in speech recognition, where Z is the articulatory configuration of the vocal tract and X the resultant signal. Another possibility is for the $Z_1, \ldots, Z_N$ to form a Markov random field (MRF) this is Titterington's "MRF" dependence structure, for example, for model-based image analysis. The difficulties of obtaining maximum-likelihood estimates of the parameters appear to increase dramatically with increasing complexity of the dependence structure.

McLachlan and Peel (2000) strongly advocate the use of the EM algorithm. They describe the use of mixture models for handling overdispersion in generalized linear models, also scaling the EM algorithm to allow mixture models to be used for data mining given huge databases, the sparse/incremental EM algorithm for speeding up the standard EM algorithm, the use of hierarchical mixtures-of-experts models for nonlinear regression, and recent research concerning hidden Markov models. Their preferred software for these purposes is their own EMMIX; in an Appendix they give details of this and other suitable software.

There are situations, however, where maximum-likelihood estimates have large biases. This possibility should not be overlooked. Other methods such as those discussed by Titterington et al. (1985) may be needed.

8.2.3 Zero-Modified and Hurdle Distributions

Empirical distributions obtained in the course of experimental investigations often have an excess of zeros compared with a Poisson distribution with the same mean. This has been a major motivating force behind the development of many distributions that have been used as models in applied statistics. The phenomenon can arise as the result of clustering; distributions with clustering interpretations (see Chapter 9) often do indeed exhibit the feature that the proportion of observations in the zero class is greater than $e^{-\overline{x}}$, where $\overline{x}$ is the observed mean.

A very simple alternative to the use of a cluster model is just to add an arbitrary proportion of zeros, decreasing the remaining frequencies in an appropriate manner. Thus a combination of the original distribution with pmf P_j, $j = 0, 1, 2, \ldots,$ together with the degenerate distribution with all probability concentrated at the origin, gives a finite-mixture distribution with

$$\begin{aligned} \Pr[X = 0] &= \omega + (1 - \omega)P_0, \\ \Pr[X = j] &= (1 - \omega)P_j, \qquad j \geq 1. \end{aligned} \tag{8.14}$$

A mixture of this kind is referred to as a *zero-modified distribution* or as a *distribution with added zeros*. Another epithet is *inflated distribution*. Mechanisms producing these distributions have been discussed by Heilbron (1994) in the context of generalized linear models.

It is also possible to take ω less than zero (*decreasing* the proportion of zeros), provided that

$$\omega + (1 - \omega)P_0 \geq 0, \quad \text{i.e.,} \;\; \omega \geq \frac{-P_0}{1 - P_0}. \tag{8.15}$$

The pgf and moments of a zero-modified distribution are easily derived from those of the original distribution. For example, if $G(z)$ is the original probability generation function, then that of the modified distribution is

$$H(z) = \omega + (1 - \omega)G(z). \tag{8.16}$$

Similarly, if the original distribution has uncorrected moments μ'_r, then the rth moment about zero of the modified distribution is $(1 - \omega)\mu'_r$.

We note that when the original distribution $\mathcal{F}_A$ has a parameter θ with distribution $\mathcal{F}_B$, the outcome can be modeled either as

$$(\mathcal{F}_A \text{ with added zeros}) \bigwedge_{\Theta} \mathcal{F}_B$$

or as

$$(\mathcal{F}_A \bigwedge_{\Theta} \mathcal{F}_B) \text{ with added zeros.}$$

In fitting a zero-modified distribution, the estimation of parameters other than ω can be carried out by ignoring the observed frequency in the zero class and then using a technique appropriate to the original distribution truncated by omission of the zero class. After the other parameters have been estimated, the value of ω can then be estimated by equating the observed and expected frequencies in the zero class. (For the modified distribution an arbitrary probability has in effect been assigned to the zero class.)

Hurdle models were introduced by Mullahy (1986) and are popular in econometrics, possibly because of their interpretation as a two-stage decision process. They allow individuals below and above the hurdle to have different statistical behavior. The commonest hurdle count data model sets the hurdle at zero, in which case the outcome is a reparameterized zero-modified distribution.

Suppose that γ and $1-\gamma$ are the probabilities of failing and of crossing the hurdle. Also let the conditional distribution of the nonzero observations be a zero-truncated distribution for which the parent probabilities are $p_0, p_1, \ldots$. Then the probabilities for the outcome distribution are

$$\begin{aligned} &\Pr[X=0]=\gamma, \\ &\Pr[X=j]=\frac{(1-\gamma)p_j}{1-p_0}=\phi\, p_j, \qquad j=1,2,\ldots. \end{aligned} \tag{8.17}$$

If $H(z)=\sum_{j\geq 0} p_j z^j$ is the parent pgf, then the outcome pgf is

$$G(z)=\gamma+(1-\gamma)\frac{H(z)-p_0}{1-p_0} \tag{8.18}$$

and the mean and variance are

$$\mu=\phi\sum_{j\geq 1} jp_j \qquad \text{and} \qquad \mu_2=\phi\sum_{j\geq 1} j^2 p_j-\left[\phi\sum_{j\geq 1} jp_j\right]^2. \tag{8.19}$$

If $H(z)$ is Poisson with parameter λ, then $\phi=(1-\rho)/(1-e^{-\lambda})$ and the outcome mean and variance are

$$\mu=\phi\lambda \qquad \text{and} \qquad \mu_2=\phi\lambda(1+\lambda)-\phi^2\lambda^2.$$

The distribution is over- or underdispersed according as $\phi \lessgtr 1$. Mullahy (1986) assumed that $\phi_{\max}$ is unbounded from above, but Winkelmann (2000) has pointed out that for the variance to be nonnegative it is necessary to have $\phi_{\max} < (1+\lambda)/\lambda$. This is not sufficiently restrictive, however. For $\Pr[X=0]\geq 0$, we require

$$\sum_{j\geq 1} p_j=\phi(1-e^{-\lambda})\leq 1, \quad \text{i.e., } \phi\leq(1-e^{-\lambda})^{-1}<\frac{1+\lambda}{\lambda}.$$

8.2.4 Examples of Zero-Modified Distributions

1. The important *zero-modified Poisson distribution* has already been mentioned in Section 4.10.3; it is defined by

$$\begin{aligned}\Pr[X=0] &= \omega + (1-\omega)e^{-\lambda}\\ \Pr[X=j] &= \frac{(1-\omega)e^{-\lambda}\lambda^j}{j!}, \qquad j=1,2,\ldots;\end{aligned} \tag{8.20}$$

the pgf is therefore

$$G(z) = \omega + (1-\omega)e^{\lambda(z-1)}. \tag{8.21}$$

Setting $\omega = (\gamma - e^{-\lambda})/(1-e^{-\lambda})$ in (8.21) gives the pgf for the hurdle-Poisson distribution discussed at the end of the previous section. This shows that it is a reparameterized zero-modified Poisson distribution.

Since it is a zero-modified distribution, one of the maximum-likelihood equations is

$$\hat{\omega} + (1-\hat{\omega})e^{-\hat{\lambda}} = \frac{f_0}{N} \tag{8.22}$$

where f_0/N is the observed proportion of zeros. It is also a power series distribution (PSD), so the other maximum-likelihood equation is

$$\overline{x} = \hat{\mu} = \hat{\lambda}(1-\hat{\omega}). \tag{8.23}$$

Eliminating $\hat{\omega}$ gives

$$\overline{x}(1-e^{-\hat{\lambda}}) = \hat{\lambda}\left(1-\frac{f_0}{N}\right), \tag{8.24}$$

and hence $\hat{\lambda}$ (and $\hat{\omega}$) can be obtained by iteration.

Singh (1963) obtained the approximate formulas

$$\begin{aligned}\mathrm{Var}(\hat{\lambda}) &\approx (1-\omega)^{-1}\lambda(1-e^{-\lambda})(1-e^{-\lambda}-\lambda e^{-\lambda})^{-1},\\ \mathrm{Var}(\hat{\omega}) &\approx (1-\omega)[\omega(1-\lambda e^{-\lambda}) + (1-\omega)e^{-\lambda}](1-e^{-\lambda}-\lambda e^{-\lambda})^{-1}.\end{aligned} \tag{8.25}$$

Martin and Katti (1965) fitted the distribution to a number of data sets using maximum likelihood with $\lambda^{(0)} = \overline{x}/(1-f_0/N)$ for the initial estimate of λ. A. W. Kemp's (1986) approximation to the maximum-likelihood equation for $\hat{\lambda}$ gave $\lambda_{(0)} = \ln(N\overline{x}/f_1)$ as an initial estimate. She found that usually

$$\lambda_0 < \hat{\lambda} < \lambda^{(0)}. \tag{8.26}$$

Kemp and Kemp (1988) also obtained bounds for $\hat{\lambda}$.

Böhning (1998) studied numerical aspects of the computer-assisted analysis of zero-inflated Poisson data. In Böhning et al. (1999) the distribution was used in the Belo Horizonte study of caries prevention in children. Rodrigues (2003) used a data augmentation algorithm in a Bayesian type of analysis.

An extended modification of the Poisson distribution, with arbitrary probabilities $p_0, p_1, p_2, \ldots, p_K$ and the remaining probabilities proportional to $e^{-\lambda}\lambda^x/x!$, was studied by Yoneda (1962).

2. The *zero-modified binomial* distribution can be formed similarly, giving

$$\begin{aligned} \Pr[X=0] &= \omega + (1-\omega)q^n, \\ \Pr[X=x] &= (1-\omega)\binom{n}{x}p^x(1-p)^{n-x}, \qquad x \geq 1, \end{aligned} \tag{8.27}$$

and the pgf

$$H(z) = \omega + (1-\omega)(1-p+pz)^n. \tag{8.28}$$

The mean and variance for this distribution are

$$\mu = (1-\omega)np \qquad \text{and} \qquad \mu_2 = (1-\omega)np(1-p+\omega np). \tag{8.29}$$

In their study of the estimation of the parameters of this distribution, Kemp and Kemp (1988) showed that the maximum- likelihood equations are

$$\frac{f_0}{N} = \hat{\omega} + (1-\hat{\omega})(\hat{q})^n \qquad \text{and} \qquad \overline{x} = n(1-\hat{\omega})\hat{p}, \tag{8.30}$$

where f_0/N is the observed relative frequency of zero. These equations do not have an explicit solution, so Kemp and Kemp also studied four other simple, explicit estimation methods suitable for a rapid assessment of data or to provide initial estimates for maximum-likelihood estimation; see also Dowling and Nakamura (1997).

They concluded (i) that estimators that give estimates that are close to one another may not, however, give an estimate that is close to the maximum-likelihood estimate, (ii) that a method that is satisfactory for one data set may be useless for another with parameters in another part of the parameter space, and (iii) that inefficient methods may lead to impossible estimates.

The zero-modified binomial is the special case $m = 1$ of Khatri and Patel's (1961) binomial–binomial distribution with pgf

$$[\omega + (1-\omega)(1-p+pz)^n]^m.$$

3. A zero-inflated geometric distribution was studied by Holgate (1964) as a model for the length of residence of animals in a specified habitat. Holgate (1966) subsequently studied a more complicated model for such data. This led to a mixed zero-inflated geometric distribution with pgf involving an Appell function of the first kind; for the theory of Appell functions see Appell and Kampé de Fériet (1926).

4. Khatri (1961) [see also Patil (1964b)] has studied the *logarithmic-with-zeros distribution* (Section 7.1.10). The probabilities are

$$\begin{aligned} \Pr[X=0] &= \omega, \\ \Pr[X=x] &= \frac{(1-\omega)\theta^x}{-x\ln(1-\theta)}, \qquad x \geq 1. \end{aligned} \tag{8.31}$$

The mean and variance are

$$\mu = \frac{(1-\omega)\alpha\theta}{1-\theta}, \qquad \mu_2 = \frac{(1-\omega)\alpha\theta[1-\alpha\theta(1-\omega)]}{(1-\theta)^2}, \tag{8.32}$$

where $\alpha = -1/\ln(1-\theta)$ and the pgf is

$$H(z) = \omega + (1-\omega)\frac{\ln(1-\theta z)}{\ln(1-\theta)}. \tag{8.33}$$

The distribution arises as a mixed binomial distribution with parameters (n, p), where n has a logarithmic distribution with parameter λ [see Khatri (1961)]; the connection between the parameters is

$$\omega = \frac{\ln(1-\lambda+\lambda p)}{\ln(1-\lambda)}, \qquad \theta = \frac{\lambda p}{1-\lambda+\lambda p}. \tag{8.34}$$

Khatri has shown that the MLEs of ω and θ are given by

$$\sum_{j\geq 1}\frac{jf_j}{N} = \frac{\hat{\theta}}{(1-\hat{\theta})[-\ln(1-\hat{\theta})]}, \qquad \hat{\omega} = 1 - \frac{f_0}{N}, \tag{8.35}$$

where f_j/N is the observed relative frequency of an observation equal to j. The asymptotic variances and covariance of the MLEs are

$$\begin{aligned} \mathrm{Var}(\hat{\omega}) &= \frac{\omega(1-\omega)}{N}, \\ \mathrm{Var}(\hat{\theta}) &\doteqdot \frac{\theta(1-\theta)^2[\ln(1-\theta)]^2}{N(1-\omega)[-\ln(1-\theta)-\theta]}, \\ \mathrm{Cov}(\hat{\omega}, \hat{\theta}) &= 0. \end{aligned} \tag{8.36}$$

5. The *log-zero Poisson* (LZP) *distribution* was constructed by Katti and A. V. Rao (1970). It can be obtained by modifying a Poisson $\bigwedge$ Logarithmic distribution by "adding zeros." This gives

$$\begin{aligned}
\Pr[X = 0] &= \omega + (1-\omega)\frac{\ln(1-\lambda e^{-\phi})}{\ln(1-\lambda)}, \\
\Pr[X = x] &= (1-\omega)\frac{\phi^x}{x!}\sum_{j=1}^{\infty}\frac{j^{x-1}(\lambda e^{-\phi})^j}{-\ln(1-\lambda)}, \qquad x \geq 1,
\end{aligned} \tag{8.37}$$

using the expression for the pmf of the Poisson $\bigwedge$ Logarithmic distribution in Section 8.3.3. The pgf is

$$H(z) = \omega + (1-\omega)\frac{\ln(1-\lambda e^{\phi(z-1)})}{\ln(1-\lambda)}. \tag{8.38}$$

Katti and Rao claim that a wide variety of distributions can be reproduced, given an appropriate choice of the parameters ω, λ, and θ. They used maximum-likelihood estimation to fit the distribution to the 35 data sets in Martin and Katti (1965) and compared the fits with Martin and Katti's fits for a number of two-parameter distributions. The LZP (with three parameters) emerged quite well from these comparisons. Katti and Rao provided a model for the distribution, obtained a recurrence formula for the probabilities, made a thorough study of the distribution's properties, and explained in detail their method of maximum-likelihood estimation. In particular they showed that the variance can be greater or less than the mean according to the values taken by the parameters.

One of the drawbacks to the use of the LZP distribution has been the complexity of (8.37). Willmot (1987a) has derived the following finite-sum expression for the pmf

$$\begin{aligned}
\Pr[X = x] = {}& \frac{(1-\omega)\phi^x}{x![-\ln(1-\lambda)]} \\
& \times \sum_{j=1}^{x}\left[\left(\frac{\lambda e^{-\phi}}{1-\lambda e^{-\phi}}\right)^j \sum_{m=1}^{j}\binom{j-1}{m-1}(-1)^{j-m}m^{x-1}\right]
\end{aligned} \tag{8.39}$$

for $x \geq 1$, and he has shown that asymptotically

$$\Pr[X = x] \approx \frac{(1-\omega)\phi^x(\phi - \ln\lambda)^{-x}}{-x\ln(1-\lambda)}.$$

6. Gupta, Gupta, and Tripathi (1995, 1996) investigated *zero-adjusted modified power series distributions*. Special attention was paid to the zero-inflated form of Consul's (1989) generalized Poisson distribution. They fitted several data sets via maximum-likelihood estimation and studied the nature of the relative error incurred by ignoring the inflation.

8.2.5 Finite Poisson Mixtures

A *k-component Poisson mixture* arises when

$$\Pr[X = x] = \sum_{j=1}^{k} \omega_j \frac{e^{-\theta_j}(\theta_j)^x}{x!}. \tag{8.40}$$

The identifiability of mixtures of Poisson distributions with k components, $\omega_j \neq 0$, $\Sigma w_j = 1$, $j = 1, 2, \ldots, k$, has been established by Feller (1943) and Teicher (1960). The number of components in a Poisson mixture has recently been researched by Karlis and Xekalaki (1999).

The factorial moments of a mixed Poisson distribution are equal to the moments about the origin for the mixing distribution. Consider a sample of size N from a mixed Poisson distribution with k components. Then equating the first $2k - 1$ sample factorial moments,

$$T_r = \sum_{i=1}^{N} \frac{x_i(x_i - 1)\cdots(x_i - r + 1)}{N}, \qquad r = 1, 2, \ldots, 2k - 1,$$

to their expectations

$$\mu'_{[r]} = \sum_{j=1}^{k} \omega_j \theta_j^r, \qquad r = 1, 2, \ldots, 2k - 1,$$

enables moment estimates of the $2k - 1$ parameters (ω_j, $j = 1, \ldots, k - 1$, and θ_j, $j = 1, \ldots, k$,) to be calculated (Rider, 1962a). Everitt and Hand (1981) have illustrated the moments method of estimation using Hasselblad's (1969) data on death notices of elderly women (two components) and also computer-generated data with known values of the parameters (three components).

Tiago de Oliviera (1965) has devised a one-sided test of the hypothesis that there are two Poisson components against a null hypothesis of a pure Poisson distribution based on $s^2 - \overline{x}$, where $\overline{x}$ and s^2 are the sample mean and variance. Böhning (1994) has shown that an amended version of the test statistic is equivalent to a normalized Fisher index of dispersion.

Maximum-likelihood estimation of the parameters of (8.40) is straightforward, but it requires the use of iteration. Everitt and Hand (1981) illustrated its use with a four-component Poisson mixture using generated data with known parameters; as initial estimates they used (1) the known values and (2) guesses (the method of moments failed to give a solution with real roots). The discrepancy between their two sets of final values was very noticeable.

Hasselblad (1969) used the EM algorithm to obtain maximum-likelihood estimates for a mixture of two Poissons for the death notice data. Titterington et al. (1985) reexamined maximum-likelihood estimation for these data using both the EM algorithm and a Newton–Raphson technique, with various sets of initial values. The Newton–Raphson algorithm either failed to converge or converged to four decimal places in fewer than a dozen iterations; the EM algorithm, on the other hand, always converged but took between one and two thousand iterations to achieve the same accuracy.

A modified version of the minimum χ^2 method of estimation that had been suggested by Blischke (1964) was applied to Poisson mixtures by Saleh (1981).

For more complicated mixtures involving Poisson distributions, see Simar (1976) and Godambe (1977). Medgyessy (1977, pp. 200–224) has studied the decomposition of finite mixtures of an unknown number of Poisson components with (1) means that are not too similar (i.e., have distinct separated values) and (2) means less than unity.

Karlis and Xekalaki (1998, 2001) have introduced a minimum Helliger distance method of inference which they consider to be efficient when the model is correct and robust when it is not.

8.2.6 Finite Binomial Mixtures

The usual mixture of binomial distributions is one in which the component distributions have a common (known) value of the exponent parameter n but differing values of p. It has probabilities of the form

$$\Pr[X = x] = \binom{n}{x} \sum_{j=1}^{k} \omega_j p_j^x (1 - p_j)^{n-x}, \tag{8.41}$$

where $\omega_j > 0$, $\sum_{j=1}^{k} \omega_j = 1$. A binomial mixture of this kind is identifiable iff $2k - 1 \leq n$. Teicher (1961) established the necessity of this condition, and Blischke (1964) proved its sufficiency.

A binomial mixture of the type

$$\Pr[X = x] = \sum_{j=1}^{k} \omega_j \binom{n_j}{x} p_j^x (1 - p_j)^{n_j - x}, \tag{8.42}$$

where $\omega_j > 0$, $\sum_{j=1}^{k} \omega_j = 1$, $0 < p_j < 1$, and the n_j are different integers, is also identifiable (Teicher, 1963).

If $n_1 = n_2 = \cdots = n_k$, if k is small, and particularly if k is known, then it may be possible to obtain useful estimates of ω_j and p_j, $j = 1, \ldots, k$. Blischke (1964, 1965) has shown that estimation is possible, provided that there are available at least $2k - 1$ observations on the rv with the binomial mixture distribution. Everitt and Hand (1981) have provided a worked example of Blischke's method using data generated from a mixture of four binomials with known parameters, including a common exponent parameter n. They have also discussed Blischke's (1962, 1964) work on the asymptotic properties of the moment estimators.

Rider (1962a) and Blischke (1962) considered the case of a mixture of two binomials in detail. The mixture pmf is

$$\Pr[X = x] = \omega \binom{n_1}{x} p_1^x q_1^{n_1 - x} + (1 - \omega) \binom{n_2}{x} {p_2}^x {q_2}^{n_2 - x}, \tag{8.43}$$

with $0 < \omega < 1$, $0 < q_j = 1 - p_j < 1$, $j = 1, 2$. When $n_1 = n_2 = n$ (known), simple explicit formulas in terms of the first three moments can be obtained for the parameters ω, p_1, and p_2 [see, e.g., Everitt and Hand (1981) for details].

The special cases $k = 2$, $n_1 = n_2 = 12$ with $p_1 = p_2 = p$ and $p_1 \neq p_2$ occur in Gelfand and Soloman's (1975) analysis of jury decisions. They estimated the two models using the method of moments, maximum likelihood, and minimum χ^2, having first grouped the data into five classes. Reasonably good agreement between methods was observed.

Blischke (1962, 1964) also examined maximum-likelihood estimation for a mixture of k binomials; this requires iteration. An advantage of maximum-likelihood estimation is that it is impossible to get estimates that lie outside the admissible limits if the initial estimates are within those limits (Hasselblad, 1969). Everitt and Hand (1981) applied maximum-likelihood estimation to their mixture of four binomials using (1) the population values and (2) the moment estimates as initial estimates. Their two sets of final estimates were very close; this would not necessarily occur with other data sets, especially if the likelihood surface were uneven.

Bondesson (1988) has described the application of a mixed binomial model with seven components to data on the germination of pine seeds.

Other estimation methods that have been used include a single cycle of maximum-likelihood estimation, a scoring method with the moment estimates as initial estimates, and a minimum χ^2 approach in which the χ^2-statistic is expanded as a Taylor series. With present-day computing power such methods are unlikely to receive much attention.

8.2.7 Other Finite Mixtures of Discrete Distributions

A mixture of a Poisson and a binomial distribution was considered by Cohen (1965). The pmf is

$$\Pr[X = x] = \omega \frac{e^{-\theta}\theta^x}{x!} + (1 - \omega)\binom{n}{x} p^x (1 - p)^{n-x}, \qquad x = 0, 1, \ldots, \quad (8.44)$$

where n is assumed to be known and the second component is zero if $x > n$. For this distribution the pgf is

$$H(z) = \omega\, e^{\theta(z-1)} + (1 - \omega)(1 - p + pz)^n \quad (8.45)$$

and the factorial moments are

$$\mu'_{[r]} = \omega\theta^r + (1 - \omega)n(n - 1) \cdots (n - r + 1)p^r. \quad (8.46)$$

Cohen obtained moment estimators for the parameters ω, θ, and p.

Cohen has also described how to fit a mixture of two positive Poissons. This distribution is defined by

$$\Pr[X = x] = \omega\left(\frac{\theta_1^x/x!}{e^{\theta_1} - 1}\right) + (1-\omega)\left(\frac{\theta_2^x/x!}{e^{\theta_2} - 1}\right), \tag{8.47}$$

where $x = 1, 2, \ldots,$ and now

$$H(z) = \omega\left(\frac{e^{\theta_1 z} - 1}{e^{\theta_1} - 1}\right) + (1-\omega)\left(\frac{e^{\theta_2 z} - 1}{e^{\theta_2} - 1}\right) \tag{8.48}$$

and

$$\mu'_{[r]} = \omega\left(\frac{\theta_1^r}{1 - e^{-\theta_1}}\right) + (1-\omega)\left(\frac{\theta_2^r}{1 - e^{-\theta_2}}\right). \tag{8.49}$$

Daniels' (1961) study of the busy-time distribution in a queueing process involved mixtures of geometric distributions (with components corresponding to priority levels). Mixtures of negative binomials have been studied by Rider (1962a) (who used moments to estimate the parameters), by Harris (1983) (who used a gradient method), and by Medgyessy (1977, pp. 203–222).

John (1970) has described the use of the method of moments and a maximum-likelihood method to identify the population of origin of individual observations drawn from a mixture of two distributions. He discussed the cases where both components in the mixture are (1) binomial, (2) Poisson, (3) negative binomial, and (4) hypergeometric.

8.3 CONTINUOUS AND COUNTABLE MIXTURES OF DISCRETE DISTRIBUTIONS

8.3.1 Properties of General Mixed Distributions

The usefulness of continuous and countable mixtures of discrete distributions (particularly in accident proneness theory and actuarial risk theory) has led to a great deal of research into their properties. This section is about properties of mixed discrete distributions in general. The next section deals with properties of mixed Poisson distributions.

The $\bigwedge$ notation for mixtures of distributions was introduced in Section 8.1.3. Given a rv $\mathcal{F}$ with cdf $F(x|\theta)$ where the parameter θ can be treated as a rv $\mathcal{H}$ with cdf $H(\theta)$, then the distribution obtained by summing $\mathcal{F}$ using the mixing parameter θ has the cdf $F(x) = \sum_{\Theta} F(x|\theta)H(\theta)$, where Θ is the parameter space; integration is used when $\mathcal{H}$ has a continuous distribution. The outcome is written as

$$\mathcal{F}(\Theta) \bigwedge_{\Theta} \mathcal{H}.$$

There is also a notation, using $\bigvee$, for distributions with pgf's of the form $G_1(G_2(z))$. These are called *random sums*, alternatively *generalized* or *compounded* distributions. If the rv $\mathcal{F}_1$ has the pgf $G_1(z) = \sum_{i=0}^{\infty} \alpha_i z^i$ and the rv $\mathcal{F}_2$ has the pgf $G_2(z)$, then the distribution with pgf $G_1(G_2(z)) = \sum_{i=0}^{\infty} \alpha_i [G_2(z)]^i$ is called an $\mathcal{F}_1$ *distribution generalized* by the *generalizing* $\mathcal{F}_2$ *distribution*. It is represented by the symbolic form $\mathcal{F}_1 \bigvee \mathcal{F}_2$. Chapter 9 is devoted to this very important class of distributions.

Convolutions of discrete distributions are also important. The *convolution* of two independent discrete distributions with nonnegative support and pgf's $G_1(z) = \sum_{i \geq 0} \alpha_i z^i$ and $G_2(z) = \sum_{j \geq 0} \beta_j z^j$ has the pmf $p_x = \sum_{i \geq 0} \alpha_i \beta_{x-i}$ and pgf $G_X(z) = G_1(z) \cdot G_2(z)$. The relationship between the rv's is written symbolically as $\mathcal{F}_X \sim \mathcal{F}_1 * \mathcal{F}_2$.

When $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are independent discrete rv's, there are many implications, including the following:

1.

$$\left(\mathcal{F}_1 \bigwedge \mathcal{F}_2\right) \bigwedge \mathcal{F}_3 \sim \mathcal{F}_1 \bigwedge \left(\mathcal{F}_2 \bigwedge \mathcal{F}_3\right), \tag{8.50}$$

$$\left(\mathcal{F}_1 \bigvee \mathcal{F}_2\right) \bigvee \mathcal{F}_3 \sim \mathcal{F}_1 \bigvee \left(\mathcal{F}_2 \bigvee \mathcal{F}_3\right), \tag{8.51}$$

$$(\mathcal{F}_1 * \mathcal{F}_2) * \mathcal{F}_3 \sim \mathcal{F}_1 * (\mathcal{F}_2 * \mathcal{F}_3)\,; \tag{8.52}$$

when $\mathcal{F}_2$ does not depend on θ,

$$[\mathcal{F}_1(\Theta) * \mathcal{F}_2] \bigwedge_{\Theta} \mathcal{F}_3 \sim \left[\mathcal{F}_1(\Theta) \bigwedge_{\Theta} \mathcal{F}_3\right] * \mathcal{F}_2. \tag{8.53}$$

2. The following important properties of a mixed discrete distribution imply that it is overdispersed compared with the unmixed distribution:

$$\begin{aligned} E[X] &= E_\theta[E_{x|\theta}[X]], \\ E[h(X)] &= \int_\Theta E_{x|\theta}[h(X)] g(\theta)\, d\theta, \\ \operatorname{Var}(X) &= \operatorname{Var}(E_{x|\theta}[X]) + E[\operatorname{Var}_{x|\theta}(X)]. \end{aligned} \tag{8.54}$$

The identifiability of countable mixtures was studied by Patil and Bildikar (1966) and Tallis (1969). For identifiability in the general case, see Teicher (1961), Blum and Susarla (1977), and Tallis and Chesson (1982). Blischke's (1965) paper on infinite as well as finite mixtures was a very thorough overview; he included applications, implications, identifiability, estimation, and hypothesis testing. Not all of his "areas requiring further investigation" have been resolved.

8.3.2 Properties of Mixed Poisson Distributions

A *mixed Poisson distribution* has the pmf

$$\Pr[X = x] = \int_{\lambda=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!}\, dF(\lambda), \qquad \lambda > 0, \qquad x = 0, 1, 2, \ldots, \tag{8.55}$$

and the pgf

$$G(z) = \int_0^{\infty} e^{\lambda(s-1)}\, dF(\lambda) \tag{8.56}$$

when the mixing distribution is continuous with cdf $F(x)$. Integration is replaced by summation when the mixing distribution is countable.

Mixed Poisson distributions have many extra properties concerning their moments, probabilities, shape and convergence:

1. If $H(e^{it})$ is the characteristic function for the mixing distribution and $G(z)$ is the pgf of the mixed distribution, then

$$G(1 + t) = H(e^t) \qquad \text{and} \qquad \ln G(1 + t) = \ln H(e^t). \tag{8.57}$$

The factorial moments of the mixture are therefore equal to the uncorrected moments of the mixing distribution and the factorial cumulants of the mixture are equal to the cumulants of the mixing distribution. (Ottestad, 1944; Consael, 1952). Let $\kappa_{[r]}$, κ_r, and μ_r denote the factorial cumulants, cumulants, and corrected moments, respectively, for the mixture; let c_r denote the cumulants of the mixing distribution. Then

$$\begin{aligned} \kappa_{[1]} &= \kappa_1 = c_1, \\ \kappa_{[2]} &= \kappa_2 - \kappa_1 = c_2, \\ \kappa_{[3]} &= \kappa_3 - 3\kappa_2 + 2\kappa_1 = c_3, \\ &\vdots \end{aligned} \tag{8.58}$$

and

$$\begin{aligned} \mu &= \kappa_1 = c_1, \\ \mu_2 &= \kappa_2 = c_2 + c_1, \\ \mu_3 &= \kappa_3 = c_3 + 3c_2 + c_1, \\ &\vdots \end{aligned} \tag{8.59}$$

2. The representation (8.56) for the pgf has the form of a Laplace transform. The uniqueness of the Laplace transform implies that for any $G(z)$ there is a corresponding $F(\theta)$; if $F(\theta)$ is a cdf, then $G(z)$ is a mixture of Poisson distributions. The uniqueness of the Laplace transform also provides a proof of the identifiability of infinite mixtures of Poisson distributions; see Teicher (1960), Douglas (1980, pp. 59, 60), and Sapatinas (1995).

3. Teicher (1960) also showed that no mixed Poisson with more than one component can itself be a Poisson distribution and that the convolution of two mixed Poissons is itself mixed Poisson, with a mixing distribution that is the convolution of the mixing distributions for the two component mixed Poissons.
4. If $\mathcal{F}_1$ and $\mathcal{F}_2$ have mixed Poisson distributions formed using the mixing rv's X and Y, respectively, then $\mathcal{F}_1 \rightarrow \mathcal{F}_2$ iff $X \rightarrow Y$ (Grandell, 1997).
5. Suppose that $\Pr[X = x]$, $x = 0, 1, \ldots,$ is the pmf of a mixed Poisson distribution with more than one component and that Y is a Poisson random variable with the *same mean* μ as X. Then (Feller, 1943)

$$\Pr[X = 0] > \Pr[Y = 0] \tag{8.60}$$

and

$$\frac{\Pr[X = 1]}{\Pr[X = 0]} < \frac{\Pr[Y = 1]}{\Pr[Y = 0]}. \tag{8.61}$$

6. Moreover, the *two-crossings theorem* of Shaked (1980) states that the difference $\Pr[X = j] - \Pr[Y = j]$ changes sign exactly twice, with the sign pattern $\{+, -, +\}$. A mixed Poisson distribution therefore has a heavier tail than a Poisson distribution with the same mean μ. The asymptotic tail behavior of mixtures of Poisson distributions was discussed in Willmot (1989a, 1990). Perline (1998) discusses the conditions for a mixed Poisson distribution to have an upper tail asymptotically equivalent to the upper tail of its mixing distribution.
7. A mixed Poisson distribution is unimodal if the mixing distribution is unimodal and absolutely continuous (Holgate, 1970). Unimodality when the mixing distribution is not absolutely continuous is studied in Bertin and Theodorescu (1995).

Some of the proofs of the above properties involve certain long-standing theorems. These assist greatly in our understanding of relationships among distributions related to the Poisson distribution.

8. *Lévy's theorem* (Feller, 1957). If and only if a discrete probability distribution on the nonnegative integers is infinitely divisible, then its pgf can be written in the form

$$G(z) = e^{\lambda[g(z)-1]}, \tag{8.62}$$

where $\lambda > 0$ and $g(z)$ is another pgf. This implies that:
9. An infinitely divisible distribution with nonnegative support can be interpreted as a stopped sum of Poisson distributions, that is, as the sum of Y iid random variables with pgf $g(z)$, where Y has a Poisson distribution.

10. An infinitely divisible distribution with nonnegative support can be interpreted as a convolution (sum) of a distribution of Poisson singlets, doublets, triplets, and so on, where the successive parameters are proportional to the probabilities given by $g(z)$.

11. *Maceda's theorem* (Maceda, 1948; also Godambe and Patil, 1975). Consider a mixture of Poisson distributions where the mixing distribution has nonnegative support. Then the resultant distribution is infinitely divisible iff the mixing distribution is infinitely divisible. The implications are:

12. A mixed Poisson distribution obtained using an infinitely divisible mixing distribution is a Poisson–stopped sum distribution.

13. Mixing a Poisson–stopped sum distribution using an infinitely divisible mixing distribution gives rise to another Poisson–stopped sum distribution.

14. *Gurland's theorem* (Gurland, 1957). Consider two distributions $\mathcal{F}_1$ and $\mathcal{F}_2$ with pgf's

$$G_1(z) = \sum_{k\geq 0} p_k z^k \qquad \text{and} \qquad G_2(z),$$

respectively, where $G_2(z)$ depends on a parameter ϕ in such a way that

$$G_2(z|k\phi) = [G_2(z|\phi)]^k. \tag{8.63}$$

Then the mixed distribution represented by $\mathcal{F}_2(K\phi) \bigwedge_K \mathcal{F}_1$ has the pgf

$$\sum_{k\geq 0} p_k G_2(z|k\phi) = \sum_{k\geq 0} p_k [G_2(z|\phi)]^k = G_1(G_2(z|\phi)). \tag{8.64}$$

Gurland's theorem applies to linear exponential distributions and a fortiori to Poisson–stopped sum distributions. Stated symbolically it is

$$\mathcal{F}_2 \bigwedge_K \mathcal{F}_1 \sim \mathcal{F}_1 \bigvee \mathcal{F}_2 \quad \text{provided that } G_2(z|k\phi) = [G_2(z|\phi)]^k. \tag{8.65}$$

15. If $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are independent, infinitely divisible, discrete random variables that satisfy the condition for Gurland's theorem, then

$$\begin{aligned}\left(\mathcal{F}_1 \bigwedge \mathcal{F}_2\right) \bigvee \mathcal{F}_3 &\sim \mathcal{F}_3 \bigwedge \mathcal{F}_1 \bigwedge \mathcal{F}_2 \\ &\sim \left(\mathcal{F}_1 \bigvee \mathcal{F}_3\right) \bigwedge \mathcal{F}_2 \\ &\sim \mathcal{F}_2 \bigvee \mathcal{F}_1 \bigvee \mathcal{F}_3. \end{aligned} \tag{8.66}$$

Because the Poisson, binomial, and negative binomial distributions all have pgf's of the form (8.63), it follows that discrete mixtures of Poisson, binomial, and negative binomial distributions are also generalized distributions in the above sense.

Consider, for example, a mixture of binomial distributions with parameters nK and p, where $0 < p < 1$ and K is a rv taking nonnegative integer values according to a Poisson distribution with parameter θ, $0 < \theta$. Symbolically this mixture is represented by

$$\text{Binomial}(nK, p) \bigwedge_K \text{Poisson}(\theta). \tag{8.67}$$

The probabilities are

$$\Pr[X = x] = \sum_{k=0}^{\infty} \frac{(nk)!p^x(1-p)^{nk-x}}{x!(nk-x)!} \times \frac{e^{-\theta}\theta^k}{k!} \tag{8.68}$$

and the pgf is

$$\begin{aligned} G(z) &= \sum_{x=0}^{\infty}\sum_{k=0}^{\infty} \frac{(nk)!p^x(1-p)^{nk-x}e^{-\theta}\theta^k z^x}{x!(nk-x)!k!} \\ &= \sum_{k=0}^{\infty}(1-p+pz)^{nk}\frac{e^{-\theta}\theta^k}{k!} \\ &= e^{\theta(1-p+pz)^n-\theta}. \end{aligned} \tag{8.69}$$

The essential step in this argument is the *binomial* property

$$(1-p+pz)^{nk} = [(1-p+pz)^n]^k.$$

Returning to (8.69), we see that this has the form of a Poisson distribution (with parameter θ) generalized using a binomial distribution with parameters n and p, that is, (8.67) has the second symbolic representation

$$\text{Poisson}(\theta) \bigvee \text{Binomial}(n, p). \tag{8.70}$$

Notice that parameter K in (8.67) does not appear in (8.70).

McKay (1996) has investigated the rates of convergence of density estimators of the mixing density when the mixing density is absolutely continuous.

The most important of the mixed Poisson, mixed binomial, and mixed negative binomial distributions will merely be listed in the following sections in this chapter; they will be dealt with in greater depth elsewhere in the book.

8.3.3 Examples of Poisson Mixtures

The Poisson is a single-parameter distribution with variance equal to the mean. A common practical problem when analyzing sets of data thought to be Poissonian is a breakdown in this variance–mean relationship due to overdispersion (rarely underdispersion). The effects of overdispersion are twofold. First, the summary statistics have larger variances than expected and, second, there may be a loss of efficiency if an inadequate model is used. Cox (1983) studied in detail the effect on an analysis if a Poisson model is used in the presence of overdispersion.

Stochastic comparisons (likelihood ordering, simple stochastic ordering, uniform variability ordering, and expectation ordering) have been researched by Misra, Singh, and Harner (2003).

A better approach is to allow for the overdispersion by adopting a specific model. This leads to the theory of mixtures of Poisson distributions. The development of the theory of such mixtures is linked closely to accident-proneness theory and to actuarial risk theory; seminal works are Greenwood and Yule (1920) and Lundberg (1940).

We now list a number of mixtures of Poisson distributions where the mixing distributions are continuous or countable, beginning with the mixtures that are infinitely divisible (Poisson–stopped sum) distributions.

1. A gamma mixture of Poisson distributions is denoted by

$$\text{Poisson}(\Theta) \bigwedge_{\Theta} \text{Gamma}(\alpha, \beta).$$

The gamma distribution is infinitely divisible, and the outcome is the (infinitely divisible) negative binomial distribution; see Section 5.3. A three-parameter gamma distribution used as the mixing distribution gives rise to the Delaporte distribution; see Section 5.12.5. Chukwu and Gupta (1989) have studied a generalized gamma mixture of a generalized Poisson distribution.

2. An inverse-Gaussian mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta} \text{Inverse Gaussian}(m, \sigma^2),$$

was postulated by Holla (1966). An extension, with the symbolic representation Poisson $\bigwedge$ Generalized inverse Gaussian, is known in the literature as Sichel's distribution; it is a long-tailed distribution that is suitable for highly skewed data. These distributions have been the subject of a long series of papers and are discussed more fully in Section 11.1.5.

3. A Poisson mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Poisson}(\lambda),$$

is an important distribution where the mixing distribution is infinitely divisible and also countably infinite (its cdf is a step function). The resultant distribution is known as the *Neyman type A distribution*. By Gurland's theorem,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Poisson}(\lambda) \sim \text{Poisson}(\lambda) \bigvee \text{Poisson}(\phi).$$

The distribution is therefore a Poisson–stopped sum distribution, with pgf

$$G(z) = \exp[\lambda(e^{\phi(z-1)} - 1)];$$

it is examined in detail in Section 9.6.

4. A negative binomial mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Negative binomial}(k, P),$$

is another mixture with an infinitely divisible, countable mixing distribution. Gurland's theorem implies that

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Negative binomial}(k, P)$$
$$\sim \text{Negative binomial}(k, P) \bigvee \text{Poisson}(\phi);$$

this distribution is again a Poisson–stopped sum distribution. The pgf is

$$G(z) = \{1 + P - P\exp[\phi(z-1)]\}^{-k}, \tag{8.71}$$

where $P > 0$, the probabilities are

$$\Pr[X = x] = \frac{\phi^x}{x!} \sum_{j=0}^{\infty} e^{-j\phi} j^x \binom{k+j-1}{k-1} P^j Q^{-k-j}, \qquad k = 0, 1, 2, \ldots, \tag{8.72}$$

where $Q = 1 + P$, and the moments are

$$\mu = kP\phi,$$
$$\mu_2 = kP\phi + kP(1+P)\phi^2,$$
$$\vdots$$

5. A Poisson mixture of negative binomial distributions has the structure

$$\text{Negative binomial}(Y, P) \bigwedge_{Y/k} \text{Poisson}(\lambda),$$

that is,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta} \text{Gamma}(Y, P) \bigwedge_{Y/k} \text{Poisson}(\lambda).$$

This is called the Poisson–Pascal distribution; see Section 9.8. It should not be confused with the previous distribution. By Gurland's theorem, the pgf is

$$G(z) = \exp\{\lambda[(1 + P - Pz)^{-k} - 1]\}. \tag{8.73}$$

The distribution has had a long history of use as a model for biological (especially entomological) data. Taking $k = 1$ gives a Poisson mixture of geometric distributions, known as the *Pólya–Aeppli distribution*. A Poisson mixture of shifted negative binomial (Pascal) distributions is called the *generalized Pólya–Aeppli* *distribution*. For further details concerning these distributions, see Sections 9.7 and 9.8.

6. A negative binomial mixture of negative binomial distributions can be formed similarly.

The mixing distributions for the remaining distributions in this section are not infinitely divisible and so do not lead to Poisson–stopped sum distributions.

7. A (continuous) rectangular mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Rectangular}(a, b),$$

was studied by Feller (1943). Bhattacharya and Holla (1965) studied it as a possible alternative to the negative binomial in the theory of accident proneness. It is a special case of a beta mixture of Poisson distributions.

8. A beta mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Beta}(a, b),$$

was derived by Gurland (1958) by supposing that the number of insect larvae per egg mass has a Poisson distribution with parameter $\Theta = \phi P$, where P (the probability that an egg hatches into a larva) is a rv having a beta distribution. The distribution was subsequently studied by Katti (1966), who called it a type H_1 distribution. Its pgf is

$$G(z) = {}_1F_1[a; a+b; \phi(z-1)]. \tag{8.74}$$

The probabilities are

$$\Pr[X = x] = \frac{a \cdots (a+x-1)\phi^x}{(a+b)\cdots(a+b+x-1)x!}\, {}_1F_1[a+x; a+b+x; -\phi],$$
$$x = 0, 1, \ldots. \tag{8.75}$$

These may be obtained from $\Pr[X = 0]$ and $\Pr[X = 1]$ by using the recurrence relation

$$\begin{aligned}(x+2)(x+1)\Pr[X = x+2] &= (x+a+b+\phi)(x+1)\Pr[X = x+1]\\ &\quad - \phi(x+a)\Pr[X = x];\end{aligned}$$

see Section 2.4.2 for the method for obtaining the recurrence relation. Because this is a generalized hypergeometric factorial moment distribution, the factorial moments are

$$\mu'_{[r]} = \frac{a(a+1)\cdots(a+r-1)\phi^r}{(a+b)(a+b+1)\cdots(a+b+r-1)}, \tag{8.76}$$

whence

$$\mu = \frac{a\phi}{a+b},$$
$$\mu_2 = \frac{a\phi}{a+b} + \frac{ab\phi^2}{(a+b)^2(a+b+1)}, \tag{8.77}$$
$$\vdots$$

9. A truncated-gamma mixture of Poisson distributions,

$$\text{Poisson}(tY) \bigwedge_Y \text{Truncated gamma}(a, b, p),$$

was studied by Kemp (1968c) in the context of limited collective risk theory. The pgf is

$$G(z) = \frac{\int_0^p e^{ty(z-1)} e^{-ay} a^b y^{b-1}\, dy}{\int_0^p e^{-ay} a^b y^{b-1}\, dy}$$
$$= \frac{{}_1F_1[b; b+1; ptz - pt - ap]}{{}_1F_1[b; b+1; -ap]}. \tag{8.78}$$

This is a mixed (compound) Poisson process with time parameter t. Such processes have been studied widely [see, e.g., Lundberg (1940) and Cox and Isham (1980)]. Expansion of (8.78) followed by the use of Kummer's first transformation gives

$$\Pr[X = x] = \frac{b(pt)^x e^{-pt}}{(b+x)x!} \times \frac{{}_1F_1[1; b+x+1; pt+ap]}{{}_1F_1[1; b+1; ap]}$$
$$= \frac{(pt)^x (ap)^b}{(pt+ap)^{x+b} x!} \times \frac{\gamma(b+x; pt+ap)}{\gamma(b; ap)}, \tag{8.79}$$

where $\gamma(c, d)$ is an incomplete gamma function (Section 1.1.5). The factorial moments are

$$\mu = \mu'_{[1]} = \frac{bpt}{b+1} \times \frac{{}_1F_1[b+1; b+2; -ap]}{{}_1F_1[b; b+1; -ap]},$$
$$\mu'_{[2]} = \frac{b(pt)^2}{b+2} \times \frac{{}_1F_1[b+2; b+3; -ap]}{{}_1F_1[b; b+1; -ap]}, \tag{8.80}$$
$$\vdots$$

Haight's (1965) insurance claims process is an unconditional risk process with removals and can also be regarded as a mixture of Poisson distributions. The pgf is

$$G(z) = \frac{\displaystyle\int_0^\infty e^{Ty(z-1)}\Gamma(N, yt)e^{-ay}a^b y^{b-1}\,dy}{\displaystyle\int_0^\infty \Gamma(N, yt)e^{-ay}a^b y^{b-1}\,dy}$$

$$= \frac{{}_2F_1[N+b, b; b+1; (Tz-T-a)/t]}{{}_2F_1[N+b, b; b+1; -a/t]}; \qquad (8.81)$$

see also Kemp (1968a).

10. A truncated Pearson type III mixture of Poisson distributions,

$$\text{Poisson}(\lambda Y) \bigwedge_Y \text{Truncated Pearson type III}(\lambda, \beta),$$

was proposed by Barton (1966); the pgf is

$$G(z) = \int_0^1 e^{\lambda y(z-1)} \times \frac{e^{\lambda y}(1-y)^{\beta-2}\,dy}{{}_1F_1[1; \beta; \lambda]} = \frac{{}_1F_1[1; \beta; \lambda z]}{{}_1F_1[1; \beta; \lambda]}. \qquad (8.82)$$

As Barton pointed out, this is a hyper-Poisson distribution; see Section 4.12.4. Philipson (1960b) has made a systematic study of Poisson mixtures with mixing distributions that are members of the Pearson family of continuous distributions.

11. A lognormal mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_\Theta \text{Lognormal}(\xi, \sigma, a),$$

has been regarded as a competitor to the logarithmic distribution for certain kinds of ecological data; see Sections 7.1 and 11.1.4.

12. A truncated-normal mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_\Theta \text{Truncated normal}(\xi, \sigma),$$

is closely related to the Hermite distribution; see Section 9.4.

13. A Lindley (1958) mixture of Poisson distributions,

$$\text{Poisson}(\Theta) \bigwedge_\Theta \text{Lindley}(\phi),$$

has the pgf

$$G(z) = \int_0^\infty e^{\theta(z-1)} \frac{\phi^2(\theta+1)e^{-\theta\phi}\,d\theta}{\phi+1} = \frac{\phi^2(\phi+2-z)}{(\phi+1)(\phi+1-z)^2}; \qquad (8.83)$$

the pmf is

$$\Pr[X=x] = \frac{\phi^2(\phi+2+x)}{(\phi+1)^{x+3}} = \frac{(\phi+2+x)}{(\phi+1)(\phi+1+x)}\Pr[X=x-1]. \qquad (8.84)$$

The fmgf is

$$G(1+t) = \frac{1 - t/(1+\phi)}{(1 - t/\phi)^2},$$

whence

$$\mu = \frac{\phi + 2}{\phi(\phi + 1)} \qquad \text{and} \qquad \mu_2 = \frac{\phi^3 + 4\phi^2 + 6\phi + 2}{\phi^2(\phi + 1)^2}. \tag{8.85}$$

This distribution was studied by Sankaran (1970), with applications to errors and accidents. Sankaran called it the *Poisson–Lindley distribution*. It is a special case of Bhattacharya's (1966) more complicated mixed Poisson distribution.

14. A binomial mixture of Poisson distributions is represented by

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Binomial}(n, p).$$

For this distribution

$$\begin{aligned}
\Pr[X = x] &= \sum_{j=0}^{n} \binom{n}{j} \frac{p^j q^{n-j} e^{-j\phi}(j\phi)^x}{x!}, \qquad x = 0, 1, 2, \ldots, \qquad (8.86)\\
&= \frac{\phi^x}{x!} \sum_{j=0}^{x} \binom{n}{j} \Delta^j 0^x (pe^{-\phi})^j (q + pe^{-\phi})^{n-j},
\end{aligned}$$

where $q = 1 - p$. From Gurland's theorem the pgf is

$$G(z) = \{q + p\exp[\phi(z - 1)]\}^n. \tag{8.87}$$

The moment properties can be obtained quite readily from those of the binomial distribution, since the fmgf is

$$(q + pe^{\phi t})^n = \sum_{i \geq 0} \frac{a_i \phi^i t^i}{i!},$$

where a_i is the ith uncorrected moment of the binomial distribution. So

$$\mu = np\phi, \qquad \mu'_{[2]} = [np + n(n-1)p^2]\phi^2,$$

that is,

$$\begin{aligned}
\mu_2 &= np\phi + npq\phi^2,\\
\mu_3 &= np\phi + 3npq\phi^2 + npq(q - p)\phi^3, \qquad (8.88)\\
&\vdots
\end{aligned}$$

This distribution should not be confused with the more commonly used Poisson–binomial distribution of Section 9.5.

15. A logarithmic mixture of Poisson distributions is denoted by

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Logarithmic}(\lambda).$$

The pgf is

$$G(z) = \frac{\log(1 - \lambda e^{\phi(t-1)})}{\log(1 - \lambda)}. \tag{8.89}$$

The probabilities are given by

$$\begin{aligned}
\Pr[X = 0] &= [\log(1 - \lambda)]^{-1}[\log(1 - \lambda e^{-\phi})], \\
\Pr[X = 1] &= [-\log(1 - \lambda)]^{-1}\lambda\phi e^{-\phi}(1 - \lambda e^{-\phi})^{-1}, \\
\Pr[X = x] &= [-\log(1 - \lambda)]^{-1}\left(\frac{\phi^x}{x!}\right)\sum_{j=1}^{\infty} j^{x-1}(\lambda e^{-\phi})^j, \qquad x = 0, 1, 2, \ldots;
\end{aligned} \tag{8.90}$$

the restrictions on the parameters are $0 < \theta$ and $0 < \lambda < 1$. The mean and variance are

$$\begin{aligned}
\mu &= [-\log(1 - \lambda)]^{-1}\frac{\lambda\phi}{1 - \lambda}, \\
\mu_2 &= [-\log(1 - \lambda)]^{-1}\frac{\lambda\phi}{1 - \lambda} \\
&\quad + [-\log(1 - \lambda)]^{-1}\frac{\lambda\phi^2}{(1 - \lambda)^2}\{1 - \lambda[-\log(1 - \lambda)]^{-1}\}.
\end{aligned} \tag{8.91}$$

The zero-modified Poisson $\bigwedge$ Logarithmic distribution (Katti and Rao, 1970) is known as the *log-zero Poisson distribution*; see Section 8.2.4.

16. A hypergeometric mixture of Poisson distributions is

$$\text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Hypergeometric}(n, Np, N).$$

For this distribution

$$\Pr[X = x] = \binom{N}{n}^{-1}\left(\frac{\phi^x}{x!}\right)\sum_j e^{-j\phi}(j\phi)^x \binom{Np}{j}\binom{N - Np}{n - j}; \tag{8.92}$$

the summation is taken over the values of j for which $0 \le j \le Np$ and $0 \le n - j \le N - Np$. The pgf is

$$G(z) = \frac{(N - n)!(N - Np)!}{N!(N - n - Np)!}\,{}_2F_1[-n, -Np; N - Np - n + 1; e^{\phi(z-1)}]. \tag{8.93}$$

The case $n = 1$ is of course the Poisson $\bigwedge$ Bernoulli distribution.

17. Hougaard, Lee, and Whitmore (1997) have reexamined Tweedie's family of distributions [see Section 11.1.2]. Their power–variance mixture models, P-G(α, δ, θ), use mixing distributions belonging to a three-parameter exponential family with power–variance function $1 + 1/(1-\alpha)$. The Poisson $\bigwedge$ Inverse Gaussian distribution is a special case; see also Hougaard (1986) and Aalen (1992).

18. Al-Awadhi and Ghitany (2001) have used the Lomax distribution to create a mixed Poisson distribution useful for accident data.

8.3.4 Mixtures of Binomial Distributions

We now consider binomial distributions. The binomial distribution has two parameters, n and p, and either or both of these may be supposed to have a probability distribution. We will not discuss cases in which both n and p vary, though it is easy to construct such examples.

In most cases discussed in the statistical literature, p has a continuous distribution, while n is discrete. The latter restriction is necessary, but the former is not. The reader will recall that for the Poisson parameter θ both continuous and discrete distributions have been used as mixing distributions. However, discrete distributions for p have not been found to be useful and have not attracted much attention from a theoretical point of view.

Mixtures of binomial distributions have finite support and so cannot be infinitely divisible (and cannot be Poisson–stopped sum distributions). Nevertheless, since

$$(1 - p + pz)^{mj} = [(1 - p + pz)^m]^j, \tag{8.94}$$

Gurland's theorem applies when the parameter n is allowed to vary, just as it applied for Poisson distributions with θ varying.

Hald (1968) derived a general approximation for a binomial mixture where the parameter p has a continuous mixing distribution with probability density function $w(p)$. He found that

$$\begin{aligned}\Pr[X = x] &= \binom{n}{x} \int_0^1 p^x (1-p)^{n-x} w(p)\, dp \\ &= \frac{w(x/n)}{n} \left[1 + \frac{b_1(x/n)}{n} + \frac{b_2(x/n)}{n^2} + O(n^{-3}) \right],\end{aligned} \tag{8.95}$$

where

$$\begin{aligned}b_1(h) &= [w(h)]^{-1} \left[-w(h) + (1-2h)w'(h) + \tfrac{1}{2}h(1-h)w''(h) \right], \\ b_2(h) &= [w(h)]^{-1} \left[w(h) - 3(1-2h)w'(h) + (1 - 6h + 6h^2)w''(h) \right. \\ &\qquad \left. + \tfrac{5}{6}h(1-h)(1-2h)w'''(h) \right] + \tfrac{1}{8}h^2(1-h)^2 w^{iv}(h) \big].\end{aligned}$$

Al-Zaid (1989) showed that a mixed binomial distribution is unimodal if the mixing distribution has support (0, 1) and is unimodal.

8.3.5 Examples of Binomial Mixtures

A number of mixtures of binomials are now listed, beginning with p varying. Because p is restricted to the range (0, 1), a gamma mixture of binomials is impossible.

1. A beta mixture of binomial distributions,

$$\text{Binomial}(n, P) \bigwedge_{P} \text{Beta}(\alpha, \beta)$$

has already been discussed in Chapter 6 under the name *beta–binomial*; occasionally it is called *binomial–beta*. It has been used to model variation in the number of defective items per lot in routine sampling inspection. The continuous rectangular–binomial is a special case. Horsnell (1957) proposed the use of a continuous rectangular mixing distribution with range not extending over the entire interval (0, 1). Horsnell also used

$$\text{Binomial}(n, P) \bigwedge_{P} \text{Triangular}$$

in the same connection.

We come now to mixtures of binomials formed by n varying.

2. A Poisson mixture of binomial distributions is denoted by

$$\text{Binomial}(N, p) \bigwedge_{N/n} \text{Poisson}(\lambda).$$

This is known as the *Poisson–binomial distribution*. The *Hermite distribution* is the particular case $n = 2$. These important distributions are discussed in Sections 9.4 and 9.5.

3. A binomial mixture of binomial distributions,

$$\text{Binomial}(N, p) \bigwedge_{N/n} \text{Binomial}(N', p'),$$

has the pgf

$$G(z) = [1 - p' + p'(1 - p + pz)^n]^{N'}, \tag{8.96}$$

and the probabilities are

$$\Pr[X = x] = \sum_{j \geq x/n} \left[\binom{N'}{j} (p')^j (1 - p')^{N'-j} \binom{nj}{x} p^x (1 - p)^{nj-x} \right], \tag{8.97}$$

where $x = 0, 1, 2, \ldots, N'n$. The pgf is $[1 - p' + p'(1 - p) + pp'z]^{N'}$ if $n = 1$, and the distribution is binomial. This result is also evident on realizing that the model corresponds to N' repetitions of a two-stage experiment with probabilities of success p, p' at the two stages independently. The distribution has not been used very much. The fact that there are four parameters to be estimated (three if n is known) makes fitting the distribution a discouraging task.

4. A negative binomial mixture of binomial distributions,

$$\text{Binomial}(N, p) \bigwedge_{N/n} \text{Negative binomial}(k, P'),$$

has similarly found little application in statistical work. The pgf is

$$G(z) = [1 + P' - P'(1 - p + pz)^n]^{-k}. \tag{8.98}$$

5. A logarithmic mixture of binomial distributions,

$$\text{Binomial}(N, p) \bigwedge_{N/n} \text{Logarithmic}(\theta),$$

has the pgf

$$G(z) = \frac{\ln[1 - \theta(1 - p + pz)^n]}{\ln(1 - \theta)} \tag{8.99}$$

and the probabilities are

$$\Pr[X = x] = [-\ln(1 - \theta)]^{-1} \sum_{j \geq x/n} \frac{\theta^j}{j} \binom{nj}{x} p^x (1 - p)^{nj-x}, \tag{8.100}$$

where $x = 0, 1, 2, \ldots$ and $0 < \theta < 1$. The special case $n = 1$ yields the zero-modified logarithmic distribution; see Section 8.2.4.

Another kind of mixture of binomials is as follows.

6. A hypergeometric mixture of binomial distributions has the form

$$\text{Binomial}\left(m, \frac{Y}{n}\right) \bigwedge_{Y} \text{Hypergeometric}(n, Np, N).$$

This model corresponds to sampling without replacement from a population of size N containing Np "defective" items followed by sampling (with sample size m) with replacement from the resultant set of n individuals. The probabilities are

$$\Pr[X = x] = \frac{\binom{m}{x}}{\binom{N}{n}} \sum_{y} \binom{Np}{y} \binom{N - Np}{n - y} \left(\frac{y}{n}\right)^x \left(1 - \frac{y}{n}\right)^{m-x}. \tag{8.101}$$

The range of summation for y is $\max(0, n - N + Np)$ to $\min(Np, n)$.

Further work on mixtures of binomial distributions is that of Bowman, Kastenbaum, and Shenton (1992).

8.3.6 Other Continuous and Countable Mixtures of Discrete Distributions

A number of other discrete mixtures are now catalogued. Mixed logarithmic distributions are not discussed as they do not lend themselves to analysis and are not at present used in statistical modeling.

Gurland's theorem applies to mixtures of negative binomial distributions as well as to mixtures of Poisson and binomial ones, since

$$(1+P-Pz)^{-mj} = [(1+P-Pz)^{-m}]^{j},$$

where $k = mj$. The power parameter of the negative binomial distribution can take all nonnegative values, however, unlike the power parameter of the binomial distribution.

1. The usual beta mixture of negative binomial distributions is

$$\text{Negative binomial}(k, P) \bigwedge_{Q^{-1}} \text{Beta}(\alpha, \beta),$$

where $Q = 1 + P$ [i.e., using the parameterization $(1+P-Pz)^{-k}$ for the negative binomial pgf]. Here $p = Q^{-1}$ has the pdf

$$\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$

and the resultant distribution is the beta–negative binomial distribution of Section 6.2.3.

2. A gamma mixture of negative binomial distributions,

$$\text{Negative binomial}(K, P) \bigwedge_{K} \text{Gamma}(\alpha, \beta),$$

has the pgf

$$\begin{aligned} G(z) &= [\beta^{\alpha}\Gamma(\alpha)]^{-1} \int_0^{\infty} (1+P-Pz)^{-k} k^{\alpha-1} e^{-k/\beta}\, dk \\ &= \int_0^{\infty} \frac{\exp\{-y[1+\beta\ln(1+P-Pz)]\} y^{\alpha-1}\, dy}{\Gamma(\alpha)} \\ &= [1+\beta\ln(1+P-Pz)]^{-\alpha}. \end{aligned} \tag{8.102}$$

This can also be regarded as a negative binomial–stopped sum of logarithmic distributions,

$$\text{Negative binomial}(\alpha, \rho) \bigvee \text{Logarithmic}(\lambda),$$

with pgf

$$G(z) = \left[1 + \rho - \rho\frac{\ln(1-\lambda z)}{\ln(1-\lambda)}\right]^{-\alpha}, \tag{8.103}$$

by taking $\rho = \beta\ln(1+P)$ and $\lambda = P/(1+P)$. The mean and variance are

$$\mu = \alpha\beta P, \qquad \mu_2 = \alpha\beta P(1 + P + \beta P). \tag{8.104}$$

3. A binomial mixture of negative binomial distributions,

$$\text{Negative binomial}(kY, P) \bigwedge_Y \text{Binomial}(n, p),$$

is rarely used. The pgf is

$$G(z) = [1 - p + p(1 + P - Pz)^{-k}]^n. \tag{8.105}$$

4. A logarithmic mixture of negative binomial distributions,

$$\text{Negative binomial}(kY, P) \bigwedge_Y \text{Logarithmic}(\theta),$$

has the pgf

$$G(z) = \frac{\ln[1 - \theta(1 + P - Pz)^{-k}]}{\ln(1-\theta)}. \tag{8.106}$$

When $k = 1$, the pgf can be written as

$$G(z) = \frac{\ln(1 + P - Pz - \theta) - \ln(1 + P - Pz)}{\ln(1-\theta)}.$$

In the following mixtures of hypergeometric distributions the parameter Y is assumed to vary. Hald (1960) commented that, if the rth factorial moment of the distribution of Y is $\xi'_{[r]}$, then the rth factorial moment of the mixture distribution is

$$\mu'_{[r]} = \frac{n!(N-r)!}{(n-r)!N!}\xi'_{[r]}.$$

5. A binomial mixture of hypergeometric distributions,

$$\text{Hypergeometric}(n, Y, N) \bigwedge_Y \text{Binomial}(N, p),$$

gives rise to another binomial distribution with parameters n, p. This is self-evident on realizing that it represents the results of choosing a random subset of a random sample of fixed size.

6. A hypergeometric mixture of hypergeometric distributions is represented by

$$\text{Hypergeometric}(n, Y, N) \bigwedge_{Y} \text{Hypergeometric}(N, N'p', N').$$

By an argument similar to that for a binomial mixture of hypergeometric distributions, it can be seen that this is a hypergeometric distribution with parameters n, $N'p'$, N'. A special case is

$$\text{Hypergeometric}(n, Y, N) \bigwedge_{Y} \text{Discrete rectangular},$$

where the probabilities for the discrete rectangular distribution are

$$\Pr[X = x] = \frac{1}{N+1}, \qquad x = 0, 1, 2, \ldots, N. \tag{8.107}$$

More details concerning mixed hypergeometric distributions are given in Hald (1960). They were examined by Horsnell (1957) in relation to economical acceptance sampling schemes.

8.4 GAMMA AND BETA MIXING DISTRIBUTIONS

A *continuous mixture* of discrete distributions arises when a parameter corresponding to some feature of a model for a discrete distribution can be regarded as a rv taking a continuum of values. If, as is common, the variable parameter can take any nonnegative value, a frequent assumption is that it has a gamma distribution—the unimodality of the gamma distribution makes it a realistic choice in many situations. The resultant mixture has pgf

$$G(z) = \int_0^\infty g(z|u) e^{-u/\beta} u^{\alpha-1} [\beta^\alpha \Gamma(\alpha)]^{-1}\, du, \tag{8.108}$$

where $g(z|u)$ is the pgf of a distribution with parameter u. The relationship between this integral and the Laplace transform leads to mathematical tractability in many instances.

Suppose that the parameter can take only a restricted range of values, from a to b. It is then often scaled to take values from 0 to 1 and assumed to have a beta distribution on (0, 1). The variety of shapes that the beta density can take leads to its popularity for this purpose. The resultant distribution has the pgf

$$G(z) = \int_0^1 g(z|u) \frac{u^{c-1}(1-u)^{d-c-1}\, du}{B(c, d-c)}. \tag{8.109}$$

This integral is related to the Mellin transform, and often the integration is reasonably tractable.

Kemp GHFDs have pgf's of the form

$${}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta(z-1)];$$

see Section 2.4.2. Gamma and beta mixtures of these distributions have the special property that the mixture distribution is also generalized hypergeometric factorial. For gamma mixtures

$$\begin{aligned} G(z) &= \int_0^\infty {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta u(z-1)] \times \frac{e^{-u}u^{c-1}\,du}{\Gamma(c)} \\ &= {}_{p+1}F_q[a_1, \ldots, a_p, c; b_1, \ldots, b_q; \theta(z-1)]. \end{aligned} \tag{8.110}$$

The well-known model for the negative binomial as a gamma mixture of Poisson distributions is an example.

When the mixing distribution is a beta distribution, we have

$$\begin{aligned} G(z) &= \int_0^1 {}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta u(z-1)] \times \frac{u^{c-1}(1-u)^{d-c-1}du}{B(c, d-c)} \\ &= {}_{p+1}F_{q+1}[a_1, \ldots, a_p, c; b_1, \ldots, b_q, d; \theta(z-1)], \end{aligned} \tag{8.111}$$

where $d > c > 0$. An example is the beta–Poisson (i.e., Poisson $\bigwedge$ Beta) distribution of Gurland (1958); see also Holla and Bhattacharya (1965).

Another example is the beta mixed negative binomial distribution of Kemp and Kemp (1956a). This is not the usual beta–negative binomial distribution of Section 6.2.3; instead it is a special case of Katti's (1966) type H_2 distribution (Section 2.4.2).

Here

$$\begin{aligned} G(z) &= \int_0^1 {}_1F_0[k;\ ; u(z-1)] \times \frac{u^{c-1}(1-u)^{d-c-1}\,du}{B(c, d-c)} \\ &= {}_2F_1[k, c; d; (z-1)]. \end{aligned} \tag{8.112}$$

Kemp GHPDs possess analogous properties (Kemp, 1968a). We recall from Section 2.4.1 that their pgf's have the form

$$\frac{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta yz]}{{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta y]}. \tag{8.113}$$

Mixtures of GHPDs using gamma-type mixing distributions with pdf's

$$\frac{e^{-y}y^{c-1}\,{}_pF_q[a_1, \ldots, a_p, c; b_1, \ldots, b_q; \theta y]}{\Gamma(c)\,{}_{p+1}F_q[a_1, \ldots, a_p, c; b_1, \ldots, b_q; \theta]} \tag{8.114}$$

produce mixture distributions that are again generalized hypergeometric factorial. Bhattacharya's (1966) derivation of the negative binomial distribution as a mixed confluent hypergeometric distribution can, for example, be rewritten as

$$\begin{aligned}
G(z) &= \int_0^\infty \frac{{}_1F_1[v; c; uz]}{{}_1F_1[v; c; u]} \times \frac{e^{-(\alpha+1)u}u^{c-1}\alpha^v(\alpha+1)^{c-v}\,{}_1F_1[v; c; u]\,du}{\Gamma(c)} \\
&= \int_0^\infty \frac{{}_1F_1[v; c; yz/(\alpha+1)]}{{}_1F_1[v; c; y/(\alpha+1)]} \times \frac{e^{-y}y^{c-1}\,{}_1F_1[v; c; y/(\alpha+1)]\,dy}{\Gamma(c)\,{}_1F_0[v;\ ; 1/(\alpha+1)]} \\
&= \frac{{}_1F_0[v;\ ; z/(\alpha+1)]}{{}_1F_0[v;\ ; 1/(\alpha+1)]} = \left(\frac{\alpha}{\alpha+1-z}\right)^v.
\end{aligned} \tag{8.115}$$

Mixtures of GHPDs using beta-type mixing distributions with pdf's

$$\frac{y^{c-1}(1-y)^{d-c-1}\,{}_pF_q[a_1, \ldots, a_p; b_1, \ldots, b_q; \theta y]}{B(c, d-c)\,{}_{p+1}F_{q+1}[a_1, \ldots, a_p, c; b_1, \ldots, b_q, d; \theta]} \tag{8.116}$$

are similarly again GHPDs. Barton's (1966) derivation of the hyper-Poisson distribution as a mixture of Poisson distributions can be rewritten as

$$\begin{aligned}
G(z) &= \int_0^1 \frac{{}_0F_0[\ ;\ ; \theta yz]}{{}_0F_0[\ ;\ ; \theta y]} \times \frac{(1-y)^{d-2}\,{}_0F_0[\ ;\ ; \theta y]\,dy}{{}_1F_1[1; d; \theta]} \\
&= \frac{{}_1F_1[1; d; \theta z]}{{}_1F_1[1; d; \theta]}
\end{aligned} \tag{8.117}$$

(where ${}_0F_0[\ ;\ ; \theta y] = e^{\theta y}$).

We note that these mixing processes are only meaningful when the initial and the resultant distributions have pgf's that either converge or terminate.

Only positive numerator and denominator parameters can be added by these procedures.

CHAPTER 9

Stopped-Sum Distributions

Introduction

The distributions considered in this chapter result from the combination of two independent distributions in a particular way. This process was called "generalization" by Feller (1943). The use of the term generalization was reinforced by Gurland (1957), who introduced the symbolic notation that is customarily employed to represent the process. However, "generalized" is a term that is greatly overused in statistics. Some authors, for example Douglas (1971, 1980), chose to use the term "stopped sum" instead for this type of distribution. This is because the principal model for the process can be interpreted as the summation of observations from the distribution $\mathcal{F}_2$, where the number of observations to be summed is determined by an observation from the distribution $\mathcal{F}_1$ (that is, summation of $\mathcal{F}_2$ observations is stopped by the value of the $\mathcal{F}_1$ observation).

A number of distributions that were listed in the previous chapter are dealt with more fully here; this is because they arise from two distinct models, a mixture model and a stopped-sum model. The relationship between the two kinds of models is a consequence of Gurland's theorem. Readers are advised to refer back to the previous chapter when reading the present chapter and to pay particular attention to the three important theorems in Section 8.3.2.

Much research has been devoted to stopped-sum distributions. Within the space available we have tried to give the main results for the more important ones. Certain of the distributions, including the negative binomial, the order-k, and the lost-games distributions, have other, perhaps more important modes of genesis; these distributions are discussed elsewhere in the book.

9.1 GENERALIZED AND GENERALIZING DISTRIBUTIONS

Neyman (1939) constructed a statistical model of the distribution of larvae in a unit area of a field (in a unit of habitat) by assuming that the variation in the

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

number of clusters of eggs per unit area (per unit of habitat) could be represented by a Poisson distribution with parameter λ, while the numbers of larvae developing per cluster of eggs are assumed to have independent Poisson distributions all with the same parameter ϕ. This is a model of heterogeneity. Neyman described the model as "contagious," but before long there arose a substantial body of opinion distinguishing *heterogeneity* from *true contagion*, that is, from situations in which the events under observation depend on the pattern of previous occurrences of the events.

Consider the initial (zero) and first generations of a branching process. Let the probability generating function (pgf) for the size N of the initial (parent) generation be $G_1(z)$, and suppose that each individual i of this initial generation independently gives rise to a random number Y_i of first-generation individuals, where $Y_1, Y_2, \ldots$ have a common distribution, that of Y with pgf $G_2(z)$. The random variable for the total number of first-generation individuals is then

$$S_N = Y_1 + Y_2 + \cdots + Y_N, \tag{9.1}$$

where N and Y_i, $i = 1, 2 \ldots, N$, are all rv's. The pgf of the distribution of S_N is

$$E[z^{S_N}] = E_N[E[z^{S_N}|N]] = E_N[G_2(z)] = G_1(G_2(z)). \tag{9.2}$$

This result seems to have been formulated for the first time by Watson and Galton (1874) and to have been rediscovered repeatedly. Heyde and Seneta (1972), however, have made an earlier attribution to Bienaymé.

Some authors have termed distributions with pgf's of the form $G_1(G_2(z))$ "clustered" or "compound" (the term compound is very ambiguous as it is frequently used for mixtures distributions such as those in Chapter 8). Douglas (1980) discussed terminology for these distributions at some length; see also Section 4.11. In Section 8.3.1, a distribution with pgf of the form (9.2) was called a *generalized* $\mathcal{F}_1$ *distribution*, or more precisely an $\mathcal{F}_1$ *distribution generalized* by the *generalizing* $\mathcal{F}_2$ *distribution*. Another common terminology is to say that S_n has a randomly $\mathcal{F}_1$–stopped sum–$\mathcal{F}_2$ distribution. [Here, $G_1(z)$ is the pgf for $\mathcal{F}_1$; $G_2(z)$ is the pgf for $\mathcal{F}_2$].

One of the authors who consistently used the term "generalized" to denote these distributions is Gurland (1957). Gurland's notation is especially useful. It enables us to refer to an $\mathcal{F}_1$–stopped sum–$\mathcal{F}_2$ distribution, that is, an $\mathcal{F}_1$ distribution generalized by an $\mathcal{F}_2$ distribution, by means of the symbolic representation

$$S_N \sim \mathcal{F}_1 \bigvee \mathcal{F}_2. \tag{9.3}$$

For example, the negative binomial distribution can be represented as

$$\text{Negative binomial} \sim \text{Poisson} \bigvee \text{Logarithmic};$$

see (5.15) in Section 5.3.

Consider now the characteristic function of S_N. Let $\varphi_1(t) = G_1(e^{it})$ and $\varphi_2(t) = G_2(e^{it})$ be the cf's for N and X, respectively. Then the cf of S_N is

$$\varphi(t) = G_1(G_2(e^{it})) = G_1(\varphi_2(t)) = \varphi_1(-i\ln[\varphi_2(t)]). \tag{9.4}$$

Clearly discrete distributions can arise provided that Y has a discrete distribution, even when $\varphi_1(t)$ is the cf of a continuous distribution; see Douglas (1980). However, under these circumstances the model loses its physical meaning.

If the pgf of Y [i.e., $G_2(z)$] depends on a parameter ϕ in such a way that

$$G_2(z|j\phi) = [G_2(z|\phi)]^j, \tag{9.5}$$

then Gurland's theorem holds, and

$$\mathcal{F}_1 \bigvee \mathcal{F}_2 \sim \mathcal{F}_2 \bigwedge_j \mathcal{F}_1; \tag{9.6}$$

in such circumstances the generalized (stopped-sum) distribution has an alternative genesis as a mixed distribution. (It is customary in the case of a mixture distribution to write under the inverted-vee sign the name of the parameter that is summed or integrated out when the mixture is formed.) To clarify this duality in modes of genesis, readers should refer back to Section 8.3 for the general theory and for examples.

There are special relationships between stopped-sum distributions and their components. Let

$$G_1(z) = \sum_{i\ge 0} a_i z^i, \qquad G_2(z) = \sum_{i\ge 0} b_i z^i, \qquad G_1(G_2(z)) = \sum_{i\ge 0} c_i z^i. \tag{9.7}$$

Then by direct expansion

$$\begin{aligned}
\Pr[X=0] &= c_0 = \sum_{j=0}^{\infty} a_j b_0^j, \\
\Pr[X=1] &= c_1 = \sum_{j=0}^{\infty} j a_j b_0^{j-1} b_1, \\
\Pr[X=2] &= c_2 = \sum_{j=0}^{\infty} a_j \left(j b_0^{j-1} b_2 + \frac{j(j-1)}{2} b_0^{j-2} b_1^2 \right), \\
&\;\;\vdots
\end{aligned} \tag{9.8}$$

A general expression for $\Pr[X = x]$, $x \ge 1$, can be obtained by the use of Faà di Bruno's formula (Jordan, 1950):

$$\begin{aligned}
\frac{d^x}{dz^x} G_1(G_2(z)) = \sum_{\pi(x)} \Bigg\{ & \frac{x!}{n_1!n_2!\cdots n_x!} \left[\frac{d^n G_1(u)}{du^n} \right]_{u=G_2(z)} \\
& \times \left(\frac{G_2^{(1)}}{1!} \right)^{n_1} \left(\frac{G_2^{(2)}}{2!} \right)^{n_2} \cdots \left(\frac{G_2^{(x)}}{x!} \right)^{n_x} \Bigg\},
\end{aligned} \tag{9.9}$$

where $G_2^{(r)} = d^r G_2(z)/dz^r$ and summation is over all partitions $\pi(x)$ of x with $n_i \geq 0$, $i = 1, \ldots, x$, such that

$$x = n_1 + 2n_2 + \cdots + xn_x \quad \text{and} \quad n = n_1 + n_2 + \cdots + n_x \tag{9.10}$$

(for $x > 1$ at least one of the n_i will be zero).

An alternative notation for the derivatives of a composite function involves the use of the Bell polynomials [see, e.g., Riordan (1958, pp. 34–37)], but note that Riordan uses k instead of our n, n instead of our x, and t instead of z. Continuing to use n, x, and z as in (9.7), (9.9), and (9.10), let

$$f_r = \left[\frac{d^r G_1(u)}{du^r}\right]_{u=G_2(z)}, \qquad g_r = \frac{d^r G_2(z)}{dz^r}, \qquad A_r = \frac{d^r G_1(G_2(z))}{dz^r}. \tag{9.11}$$

Then

$$\begin{aligned} A_x &= \sum_{\pi(x)} \frac{x!f_n}{n_1!n_2!\cdots n_x!}\left(\frac{g_1}{1!}\right)^{n_1}\left(\frac{g_2}{2!}\right)^{n_2}\cdots\left(\frac{g_x}{x!}\right)^{n_x} \\ &= \sum_{j=1}^{x} f_j A_{x,j}(g_1, g_2, \ldots, g_x), \end{aligned} \tag{9.12}$$

where the A_x are known as Bell's polynomials. For an exposition of their use in the present context, see Charalambides (1977a).

We have

$$\begin{aligned} A_1 &= f_1 g_1, \\ A_2 &= f_1 g_2 + f_2 g_1^2, \\ A_3 &= f_1 g_3 + f_2(3g_2 g_1) + f_3 g_1^3, \\ A_4 &= f_1 g_4 + f_2(4g_3 g_1 + 3g_2^2) + f_3(6g_2 g_1^2) + f_4 g_1^4, \\ &\vdots \end{aligned} \tag{9.13}$$

Expressions for A_5, A_6, A_7, and A_8 are given in Riordan (1958, p. 49). Interested readers may care to compare these polynomials with the expressions for uncorrected moments in terms of cumulants, given for instance in Stuart and Ord (1987, pp. 86–87), of which the first four are

$$\begin{aligned} \mu_1' &= \kappa_1, \\ \mu_2' &= \kappa_2 + \kappa_1^2, \\ \mu_3' &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3, \\ \mu_4' &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4. \end{aligned} \tag{9.14}$$

Let $u = G_2(z)$. Then returning to (9.9), we have

$$\begin{aligned} [u]_{z=0} &= [G_2(z)]_{z=0} = b_0, \\ [G_2^{(x)}(z)]_{z=0} &= x!b_x, \\ \left[\frac{d^n G_1(u)}{du^n}\right]_{z=0} &= \sum_{j\geq 0} \frac{(n+j)!}{j!} a_{n+j} b_0^j, \end{aligned} \tag{9.15}$$

whence

$$\begin{aligned} \Pr[X = x] = c_x &= \frac{1}{x!}\left[\frac{d^x}{dz^x} G_1(G_2(z))\right]_{z=0} \\ &= \sum_{\pi(x)} \left[\left(\sum_{j\geq 0} \frac{(n+j)!a_{n+j}b_0^j}{j!}\right) \frac{b_1^{n_1} b_2^{n_2} \cdots b_x^{n_x}}{n_1!n_2!\cdots n_x!}\right], \quad x \geq 1. \end{aligned} \tag{9.16}$$

If $\sum_i a_i z^i$ is an infinite series and $b_0 \neq 0$, then the inner summation over j leads to an infinite-series expression for $\Pr[X = x]$. However, whenever $b_0 = 0$, the only nonzero term in the summation over j is $n!a_n$ and we obtain the finite-series expansion

$$\Pr[X = x] = \sum_{\pi(x)} n!a_n \left(\frac{b_1^{n_1} b_2^{n_2} \cdots b_x^{n_x}}{n_1!n_2!\cdots n_x!}\right), \qquad x \geq 1. \tag{9.17}$$

Rearrangement of $G_1(G_2(x))$ to give

$$G_1(G_2(x)) = G_1^*(G_2^*(x)), \tag{9.18}$$

where $G_2^*(z)$ is the pgf of a distribution with *positive* support (i.e., where $b_0^* = 0$), is sometimes possible; it is advantageous in that it leads to simpler expressions for the probabilities of the generalized (stopped-sum) distribution. Such rearrangement is always possible for generalized Poisson (Poisson−stopped sum) distributions. (the latter are also called *multiple Poisson distributions*).

Consider now moment properties. These have been studied by Katti (1966), by Douglas (1980), and in a series of papers by Charalambides (e.g., Charalambides, 1977a, 1986a,c). Let the factorial moment generating functions of N and Y be

$$G_1(1+t) = \sum_r \frac{{}_1\mu'_{[r]} t^r}{r!} \qquad \text{and} \qquad G_2(1+t) = \sum_r \frac{{}_2\mu'_{[r]} t^r}{r!}, \tag{9.19}$$

respectively. Then the factorial moments of $X = S_N$ are

$$\mu'_{[r]} = \sum_{n=1}^{r} \left[{}_1\mu'_{[n]} \sum \frac{r!}{n_1!n_2!\cdots n_r!} \left(\frac{{}_2\mu'_{[1]}}{1!}\right)^{n_1} \left(\frac{{}_2\mu'_{[2]}}{2!}\right)^{n_2} \cdots \left(\frac{{}_2\mu'_{[r]}}{r!}\right)^{n_r}\right], \tag{9.20}$$

with the inner summation over all partitions of r with $n_i \geq 0$, $i = 1, \ldots, r$, such that

$$r = n_1 + 2n_2 + \cdots + rn_r \quad \text{and} \quad n = n_1 + n_2 + \cdots + n_r. \tag{9.21}$$

The factorial cumulants of $X = S_N$ are similarly

$$\kappa_{[r]} = \sum_{n=1}^{r} \left[{}_1\kappa_{[n]} \sum \frac{r!}{n_1!n_2!\cdots n_r!} \left(\frac{{}_2\mu'_{[1]}}{1!}\right)^{n_1} \left(\frac{{}_2\mu'_{[2]}}{2!}\right)^{n_2} \cdots \left(\frac{{}_2\mu'_{[r]}}{r!}\right)^{n_r} \right] \tag{9.22}$$

with the same inner summation range as before.

Setting $r = 1, 2$ in these formulas gives

$$\mu_X = \mu_N \mu_Y \qquad \text{and} \qquad \sigma_X^2 = \mu_N \sigma_Y^2 + \mu_Y^2 \sigma_N^2, \tag{9.23}$$

where (μ_X, σ_X^2), $(\mu_N,\ \sigma_N^2)$, and (μ_Y, σ_Y^2) are the means and variances of X, N, and Y, respectively. In terms of the index of dispersion, $I = \sigma^2/\mu$, (9.23) becomes

$$I_X = I_Y + \mu_Y I_N; \tag{9.24}$$

hence I_X is always greater than I_Y.

Katti (1966) has demonstrated how to obtain expressions for the probabilities for a generalized distribution via (9.20) and

$$\Pr[X = x] = \sum_{j \geq x} \frac{(-1)^{j-x} \mu'_{[j]}}{x!(j-x)!} \tag{9.25}$$

(i.e., by inversion of the fmgf).

9.2 DAMAGE PROCESSES

The concept of a damage process is inherent in the work of Catcheside (1948). For a given dosage and length of exposure to radiation, the number N of chromosome breakages in individual cells can be assumed to have a Poisson distribution with pgf $e^{\theta(z-1)}$. If each breakage has a fixed and independent probability p of persisting and probability $q = 1 - p$ of healing, then the number of observed breakages has pgf $e^{\theta[(q+pz)-1]} = e^{\theta p(z-1)}$ (i.e., has a Poisson distribution with parameter θp; see Section 4.8).

More generally [see, e.g., Kendall (1948)], if $G(z)$ is the pgf for the distribution of the number of animals of a particular species per unit of habitat in animal-trapping experiments and all of the animals have an independent and equal probability p of being trapped, then the number of animals trapped per unit of habitat will have pgf $G(q + pz)$. Similarly, if each child in a family has the same independent probability of having a certain genetic defect and this probability is the same for all families under consideration, then the number of children per family who have the defect has the pgf $G(q + pz)$, where $G(z)$ is the pgf for family size.

The pgf of X is

$$\begin{aligned} G_X(z) &= G_Y(1 - p + pz) \\ &= \sum_{y\geq 0} \sum_{j=0}^{y} \Pr[Y = y] \binom{y}{j} q^{y-j} p^j z^j, \end{aligned} \tag{9.26}$$

whence

$$\Pr[X = x] = \sum_{y\geq x} \Pr[Y = y] \binom{y}{x} p^x q^{y-x}. \tag{9.27}$$

(The distribution of $X|(Y = y)$ is therefore binomial with parameters y and p.) The fmgf of X is

$$G_X(1 + t) = G_Y(1 + pt),$$

and so

$${}_X\mu'_{[r]} = p^r\ {}_Y\mu'_{[r]}. \tag{9.28}$$

The first two moments of X are

$$\mu_X = p\mu_Y \qquad \text{and} \qquad \sigma_X^2 = p^2\sigma_Y^2 + p(1 - p)\mu_Y. \tag{9.29}$$

The Rao–Rubin characterization theorem for the Poisson distribution (Rao and Rubin, 1964; Rao, 1965) has initiated much work on characterizations for damage processes. This has concentrated on conditions characterizing the distribution of Y from that of X; see Sections 3.9 and 4.8 and, for example, C. R. Rao et al. (1980), Johnson and Kotz (1982), M. B. Rao and Shanbhag (1982), Panaretos (1982, 1987a), B. R. Rao and Janardan (1985), and Talwalker (1986). Rao and Rubin showed that the Bernoulli survival pattern described above is the only one that preserves a Poisson model.

Consider now the general class of distributions for which the underlying model is preserved under a Bernoulli damage process. Clearly their pgf's must be of the form

$$G(\alpha(z - 1)), \tag{9.30}$$

with maybe other parameters beside α. This class of distributions includes the Poisson, binomial, negative binomial, and other generalized hypergeometric factorial distributions (GHFDs) (Section 2.4.2) as well as the Neyman type A, Hermite, Poisson–binomial, Pólya–Aeppli, Poisson–Pascal, and Thomas distributions (see, e.g., Boswell, Ord, and Patil, 1979). Sprott (1965) showed that this class of distributions has an important maximum-likelihood feature. Since

$$\alpha\frac{\partial G}{\partial \alpha} = (z - 1)\frac{\partial G}{\partial z},$$

that is, since

$$\alpha \frac{\partial \Pr[X = x]}{\partial \alpha} = x \Pr[X = x] - (x+1) \Pr[X = x+1], \tag{9.31}$$

it follows that

$$\frac{\partial}{\partial \alpha} \sum_x f_x \ln \Pr[X = x] = \sum_x f_x \left(x - \frac{(x+1) \Pr[X = x+1]}{\Pr[X = x]} \right), \tag{9.32}$$

where f_x is the frequency of the observation x in a sample of size N. Hence for this class of distributions the maximum-likelihood estimate of α is obtained from the maximum-likelihood equation

$$\overline{x} = \sum_x \frac{(x+1) f_x \hat{p}_{x+1}}{N \hat{p}_x}, \tag{9.33}$$

where $\hat{p}_x$ is the maximum-likelihood estimate of $\Pr[X = x]$; see, for example, Section 9.6.4 concerning maximum-likelihood estimation for the Neyman type A distribution.

9.3 POISSON–STOPPED SUM (MULTIPLE POISSON) DISTRIBUTIONS

A rv is said to be *infinitely divisible* iff it has a cf $\varphi(t)$ that can be represented for every positive integer n as the nth power of some cf $\varphi_n(t)$:

$$\varphi(t) = [\varphi_n(t)]^n. \tag{9.34}$$

Poisson–stopped sum distributions have pgf's of the form

$$G(z) = \sum_{x \geq 0} \Pr[X = x] z^x = e^{\lambda[g(z)-1]}. \tag{9.35}$$

A distribution of this kind is often called a *generalized Poisson distribution*; a more recent name is *multiple Poisson distribution*.

In (9.35), $g(z)$ is the pgf of the generalizing distribution. In terms of branching processes, $g(z)$ is the pgf for the number of subsequent-generation individuals per previous-generation individual. Since $\varphi(t) = G(e^{it})$ and

$$e^{\lambda[g(z)-1]} = \left(e^{\lambda n^{-1}[g(z)-1]} \right)^n, \tag{9.36}$$

it follows that $G(z)$ is infinitely divisible; therefore Poisson–stopped sum distributions belong to the important class of infinitely divisible distributions.

The converse of this result is also true. If $G(z)$ is the pgf of an infinitely divisible nonnegative integer-valued rv, then it has the form

$$G(z) = e^{\lambda[g(z)-1]}, \tag{9.37}$$

where $g(z)$ is a pgf; $G(z)$ is therefore the pgf of a Poisson–stopped sum distribution [see Feller (1957) for an elementary proof]. De Finetti (1931) has proved furthermore [see also Lukacs (1970)] that *all* infinitely divisible distributions are limiting forms of multiple Poisson distributions.

The importance of the property of infinite divisibility in modeling was stressed by Steutel (1983); see also the monograph by Steutel (1970). Infinitely divisible discrete distributions with pgf's of the form

$$G(z) = \frac{(1-q)g(z)}{1-qg(z)} \tag{9.38}$$

were studied by Steutel (1990) under the name *geometrically infinitely divisible distributions*. Conditions for a discrete distribution to be infinitely divisible are discussed in Katti (1967), Warde and Katti (1971), and Chang (1989).

Suppose now that

$$G(z) = \sum_{j\geq 0} \Pr[X = x]z^x$$

is the pgf of a distribution with a finite mean; suppose also that it is a Poisson–stopped sum distribution where the Poisson parameter is λ and the generalizing pgf is $g(z) = \sum_{j\geq 0} b_j z^j$. Then the mean of $G(z)$ is $\lambda \sum_{j\geq 0} jb_j$. This is finite; also $b_j \geq 0$. Consequently $\Pr[X = 0] = \exp(-\sum_{j\geq 0} b_j) > 0$.

Conversely, if $\Pr[X = 0] = 0$ for a distribution with a finite mean, then the distribution cannot be a multiple Poisson distribution. Similarly a distribution with finite support, $x = 0, 1, \ldots, n$, cannot be a multiple Poisson distribution.

It is straightforward to show that the representation of an infinitely divisible distribution on the nonnegative integers as a Poisson–stopped sum distribution is unique only if we treat it as a Poisson–stopped sum of a distribution with *positive* support. This can be achieved by writing

$$G(z) = \exp\left[\lambda(1-b_0)\left(\frac{g(z)-b_0}{1-b_0} - 1\right)\right], \tag{9.39}$$

that is, by use of the representation

$$X \sim \text{Poisson}(\lambda(1-b_0)) \bigvee \mathcal{F}^*, \tag{9.40}$$

where the pgf for $\mathcal{F}^*$ is

$$\frac{g(z)-b_0}{1-b_0} = \frac{b_1 z + b_2 z^2 + \cdots}{b_1 + b_2 + \cdots}. \tag{9.41}$$

Such rearrangement is always possible for an infinitely divisible distribution with nonnegative integer support and a finite mean.

This rearrangement, together with the use of (9.17), enables the probabilities to be expressed as

$$\Pr[X = x] = \sum_{\pi(x)} e^{\lambda(b_0-1)} \frac{\lambda^n (1-b_0)^n (b_1^*)^{n_1} \cdots (b_x^*)^{n_x}}{n_1! \cdots n_x!}, \tag{9.42}$$

where the summation is over all partitions $\pi(x)$ of x with $n_i \geq 0$, $i = 1, \ldots, x$, $x = n_1 + 2n_2 + \cdots + xn_x$ and $n = n_1 + n_2 + \cdots + n_x$, also $b_i^* = b_i/(1-b_0)$, $i \geq 1$.

The recurrence relationship

$$(x+1)\Pr[X = x+1] = \lambda \sum_{j=0}^{x} (x+1-j) b_{x+1-j} \Pr[X = j] \tag{9.43}$$

was derived by Kemp (1967a) by the simple method of differentiating the pgf once to give

$$\begin{aligned}
\frac{dG(z)}{dz} &= \sum_x (x+1)\Pr[X = x+1] z^x \\
&= G(z) \frac{\lambda\, dg(z)}{dz} \\
&= \lambda \left[\sum_{j \geq 0} \Pr[X = j] z^j \right] \left[\sum_{k \geq 0} (k+1) b_{k+1} z^k \right];
\end{aligned}$$

the recurrence relationship is then obtained by equating coefficients of z^x. It can also be obtained by repeated differentiation of $G(z)$; see Khatri and Patel (1961) and Gurland (1965).

We find that

$$\begin{aligned}
\Pr[X = 1] &= \lambda b_1 \Pr[X = 0], \\
\Pr[X = 2] &= \lambda \left(b_2 \Pr[X = 0] + \frac{b_1}{2} \Pr[X = 1] \right) \\
&= \left(\frac{\lambda^2 b_1^2}{2!} + \lambda b_2 \right) \Pr[X = 0], \\
\Pr[X = 3] &= \lambda \left(b_3 \Pr[X = 0] + \frac{2b_2}{3} \Pr[X = 1] + \frac{b_1}{3} \Pr[X = 2] \right) \\
&= \left(\frac{\lambda^3 b_1^3}{3!} + \lambda^2 b_1 b_2 + \lambda b_3 \right) \Pr[X = 0],
\end{aligned} \tag{9.44}$$

and so on.

Consider now the moment properties of X. Let the moment and factorial moment generating functions of the generalizing distribution be

$$g(e^t) = \sum_{r\geq 0} \frac{{}_g\mu'_r t^r}{r!} \qquad \text{and} \qquad g(1+t) = \sum_{r\geq 0} \frac{{}_g\mu'_{[r]} t^r}{r!}, \tag{9.45}$$

where ${}_g\mu'_r$ and ${}_g\mu'_{[r]}$ denote the rth uncorrected moment and the rth factorial moment of the generalizing distribution. Satterthwaite (1942) showed that the first four moments of X are

$$\begin{aligned} \mu'_1 &= \lambda\, {}_g\mu'_1, \\ \mu_2 &= \lambda\, {}_g\mu'_2, \\ \mu_3 &= \lambda\, {}_g\mu'_3, \\ \mu_4 &= \lambda\, {}_g\mu'_4 + 3\lambda^2({}_g\mu'_2)^2. \end{aligned} \tag{9.46}$$

Since $g(z) = \sum_x b_x z^x$, we have

$${}_g\mu'_1 = \sum_x x b_x \leq \sum_x x^2 b_x = {}_g\mu'_2.$$

Consequently $\mu'_1 \leq \mu_2$ (i.e., the mean is less than or equal to the variance) for all multiple Poisson distributions. Equality is achieved only for the Poisson distribution itself.

More generally, from Feller (1943),

$$\ln G(e^t) = \lambda g(e^t) - \lambda, \tag{9.47}$$

and hence the cumulants of X are proportional to the uncorrected moments of $g(z)$:

$$\kappa_r = \lambda\, {}_g\mu'_r. \tag{9.48}$$

Similarly, $\ln[G(1+t)] = \lambda g(1+t) - \lambda$, whence the factorial cumulants are proportional to the factorial moments of $g(z)$:

$$\kappa_{[r]} = \lambda\, {}_g\mu'_{[r]}. \tag{9.49}$$

For relationships between the factorial moments of X and ${}_g\mu'_{[r]}$, see Katti (1966).

Generalized Poisson distributions can of course be interpreted as mixture distributions (Section 8.3.2). We have

$$G(z) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!}[g(z)]^j. \tag{9.50}$$

Let $\mathcal{F}(n)$ denote the distribution with pgf $[g(z)]^n$. Then this interpretation can be stated symbolically as

$$X \sim \text{Poisson}(\lambda) \bigvee \mathcal{F}(1) \sim \mathcal{F}(N) \bigwedge_N \text{Poisson}(\lambda). \tag{9.51}$$

Generalized Poisson distributions can also be regarded as convolutions of distributions of Poisson singlets, Poisson doublets, Poisson triplets, and so on, since

$$G(z) = e^{\lambda[g(x)-1]} = \exp\left(\sum_{i\geq 0} \lambda(b_i z^i - b_i)\right) = \prod_{i\geq 1} e^{\lambda b_i (z^i - 1)}. \tag{9.52}$$

Kemp and Kemp (1965) numerically decomposed several multiple Poisson distributions into component distributions of Poisson singlets, Poisson doublets, and so on, in an investigation into the similarity of fits to data by different multiple Poisson distributions. Maritz (1952) showed that (9.52) can also arise as the pgf for the sum of k correlated Poisson singlet distributions ($b_i = 0$ for $i > k$).

Poisson–stopped sum distributions with pgf's of the form

$$G(z) = e^{\lambda[g(x)-1]} = \exp\left(\sum_{i=0}^{k} \lambda(b_i z^i - b_i)\right) \tag{9.53}$$

(that is, with a finite number of components) have further special properties. From (9.53) their factorial cumulants are

$$\kappa_{[r]} = \sum_{i=1}^{k} \frac{\lambda b_i i!}{(i-r)!} \tag{9.54}$$

[cf. (9.49)], and their moments can be derived therefrom. Also their maximum-likelihood equations have a particularly simple form (Kemp, 1967a). Let $p_j = \Pr[X = j]$ and $a_i = \lambda b_i$. Then differentiation of the pgf with respect to a_i, $i = 1, 2, \ldots, k$, gives

$$\begin{aligned} \sum_{j=0}^{\infty} \frac{\partial p_j z^j}{\partial a_i} = \frac{\partial G(z)}{\partial a_i} &= (z^i - 1)G(z) \\ &= (z^i - 1)\sum_{j=0}^{\infty} p_j z^j, \end{aligned} \tag{9.55}$$

and equating coefficients of z yields

$$\frac{\partial p_j}{\partial a_i} = p_{j-i} - p_j, \tag{9.56}$$

where $p_{j-i} = 0$ for $j < i$. The maximum-likelihood equations are therefore

$$\frac{\partial L}{\partial a_i} \equiv \sum_{j=0}^{\infty} f_j \frac{1}{p_j}\frac{\partial p_j}{\partial a_i} \equiv \sum_{j=0}^{\infty} f_j \left(\frac{p_{j-i}}{p_j} - 1\right) = 0, \tag{9.57}$$

where $L \equiv \sum_j f_j \ln p_j$ is the log-likelihood and f_x is the observed frequency of x. Patel (1976a) has used the method to estimate the parameters of the two distributions for which $k = 3$ and $k = 4$ (he called these the *triple-stuttering Poisson* and the *quadruple-stuttering Poisson* distributions).

All Poisson–stopped sum distributions have the useful property of reproducibility, for if X_i, $i = 1, \ldots, s$, are independent rv's with multiple Poisson distributions, then $\sum_{i=1}^{s} X_i$ has the pgf

$$\prod_{i=1}^{s} \left(e^{\lambda_i g_i(z) - \lambda_i}\right) = \exp\left(\sum_{i=1}^{s} [\lambda_i g_i(z) - \lambda_i]\right) \tag{9.58}$$

and so is also a multiple Poisson distribution.

Consider now a generalization of Fisher's derivation of the logarithmic distribution as a limiting form of the negative binomial distribution (Section 7.1.2). Let $h(z) = \lambda g(z)$, where $g(z)$ is the pgf of the generalizing distribution. Then $G(z) = \exp[h(z) - h(1)]$. Using l'Hôpital's, rule it can be proved that for all multiple Poisson distributions

$$\lim_{\gamma \to 0} \left[\frac{[G(z)]^\gamma - [G(0)]^\gamma}{1 - [G(0)]^\gamma}\right] = \frac{h(z)}{h(1)} = g(z) \tag{9.59}$$

(Kemp, 1978a).

Parzen's axioms for the Poisson distribution were discussed in Section 4.2.1. If Parzen's axiom 3 is modified to allow a random number of events (possibly more than one) to occur at a given instant of time, then the outcome is a multiple Poisson distribution for the number of events in a given interval of time. A stochastic process of this kind is often called a *compound Poisson process*; however, see Sections 4.11 and 8.1 for an alternative usage of the term "compound."

The many models that give rise to Poisson–stopped sum (multiple Poisson) distributions have led to their study under a variety of names in the older literature. Besides "compound" Poisson distributions [Feller (1957)], the following terms have been used: Pollaczek–Geiringer distributions of multiple occurrences of rare events (Lüders, 1934; Haight, 1961b); composed Poisson distributions (Jánossy et al., 1950); multiple Poisson distributions (Feller, 1957); stuttering-Poisson distributions (Galliher et al., 1959; Kemp, 1967a); distributions *par grappes* (Thyrion, 1960); Poisson power series distributions (Khatri and Patel, 1961); and Poisson distributions with events in clusters (Castoldi, 1963). Thompson (1954) has described a spatial mode of genesis (Darwin's model) that can lead to multiple Poisson distributions.

From the definition of a multiple Poisson distribution it follows that none of the b_j, $j = 0, 1, \ldots$, in $g(z) = \sum_j b_j z^j$ can be negative. Nevertheless, it is interesting to ask whether

$$\exp\left(\sum_{j=1}^{\infty} \lambda b_j (z^j - 1)\right), \qquad \lambda > 0,$$

can be a pgf, albeit not for a multiple Poisson distribution, if any of the b_j, $j \geq 1$, is negative. Lévy (1937a), quoted by Lukacs (1970, p. 252), has answered the question by proving that it cannot be a pgf unless a term with a negative coefficient is preceded by one term and followed by at least two terms with positive coefficients.

Most of the distributions in the following sections of this chapter were developed using one or other of the models described in this section and are multiple Poisson distributions.

9.4 HERMITE DISTRIBUTION

The simplest multiple Poisson distribution arises when the generalizing distribution is a Bernoulli distribution with pgf $1 - p + pz$, giving

$$X \sim \text{Poisson}(\lambda) \bigvee \text{Bernoulli}(p), \tag{9.60}$$

$0 < p \leq 1$. This is equivalent to a damage process, and it leads to another Poisson distribution with parameter λp; see Section 9.2. Taking $p = 1$ gives a degenerate mixing distribution with pgf $g(z) = 1$ that leaves the initial Poisson distribution unaltered.

Consider now the generalization of a Poisson distribution with a binomial distribution with parameter $n = 2$:

$$X \sim \text{Poisson}(\lambda) \bigvee \text{Binomial}(2, p). \tag{9.61}$$

The resultant distribution has pgf

$$G(z) = \exp\{\lambda[2pq(z-1) + p^2(z^2-1)]\} \tag{9.62}$$

$$= \exp\left(\alpha\beta z + \tfrac{1}{2}\alpha^2 z^2 - \alpha\beta - \tfrac{1}{2}\alpha^2\right) \tag{9.63}$$

in the notation of Kemp and Kemp (1965). These authors showed that the pgf can be expanded in terms of Hermite polynomials, giving

$$\begin{aligned} \Pr[X=0] &= e^{-\alpha\beta-\alpha^2/2} \\ \Pr[X=x] &= \frac{\alpha^x H_x^*(\beta)}{x!}\Pr[X=0], \qquad x = 1, 2, \ldots, \end{aligned} \tag{9.64}$$

where $H_x^*(\beta)$ is the modified Hermite polynomial of Fisher (1951):

$$H_x^*(\beta) = \sum_{j=0}^{[n/2]} \frac{n!x^{n-2j}}{(n-2j)!j!2^j},$$

where $[n/2]$ denotes the integer part of $n/2$. Hence

$$\begin{aligned}
\Pr[X=1] &= \alpha\beta \Pr[X=0], \\
\Pr[X=2] &= \frac{\alpha^2(\beta^2+1)}{2!}\Pr[X=0], \\
\Pr[X=3] &= \frac{\alpha^3(\beta^3+3\beta)}{3!}\Pr[X=0], \\
\Pr[X=4] &= \frac{\alpha^4(\beta^4+6\beta^2+3)}{4!}\Pr[X=0], \\
\Pr[X=5] &= \frac{\alpha^5(\beta^5+10\beta^3+15\beta)}{5!}\Pr[X=0], \\
&\vdots
\end{aligned} \tag{9.65}$$

The probabilities can also be expressed in terms of confluent hypergeometric functions. Setting $a_1 = \alpha\beta$, $a_2 = \alpha^2/2$ gives

$$\begin{aligned}
\Pr[X=0] &= e^{-a_1-a_2}, \\
\Pr[X=x] &= e^{-a_1-a_2}\sum_{j=0}^{[x/2]} \frac{a_1^{x-2j}a_2^j}{(x-2j)!\,j!},
\end{aligned} \tag{9.66}$$

and hence

$$\begin{aligned}
\Pr[X=2r] &= e^{-a_1-a_2}\left(\frac{a_2^r}{r!}\right) {}_1F_1\left[-r; \frac{1}{2}; -\frac{a_1^2}{4a_2}\right], \\
\Pr[X=2r+1] &= e^{-a_1-a_2}a_1\left(\frac{a_2^r}{r!}\right) {}_1F_1\left[-r; \frac{3}{2}; -\frac{a_1^2}{4a_2}\right].
\end{aligned} \tag{9.67}$$

Kummer's transformation for the ${}_1F_1[\cdot]$ series,

$${}_1F_1[\alpha;\beta;\xi] = e^{\xi}\,{}_1F_1[\beta-\alpha;\beta;-\xi] \tag{9.68}$$

(Section 1.1.7) yields the alternative infinite-series formula

$$\Pr[X=x] = \exp\left(-a_1 - a_2 - \frac{a_1^2}{4a_2}\right)\sum_{j=[(x+1)/2]}^{\infty} \frac{(2j)!(a_1/2)^{2j-x}a_2^{x-j}}{x!(2j-x)!\,j!}. \tag{9.69}$$

The recursion relationship

$$(x+1)\Pr[X=x+1] = \alpha\beta\Pr[X=x] + \alpha^2\Pr[X=x-1],$$

that is,

$$(x+1)\Pr[X=x+1] = a_1\Pr[X=x] + 2a_2\Pr[X=x-1], \tag{9.70}$$

is useful. It holds for $x \geq 0$, with $\Pr[X=-1]=0$.

Patel (1985) has obtained an asymptotic formula for the cumulative probabilities.

The cumulant generating function is

$$\ln G(e^t) = a_1(e^t - 1) + a_2(e^{2t} - 1), \tag{9.71}$$

whence $\kappa_r = a_1 + 2^r a_2$. Thus

$$\begin{aligned}
\mu &= \kappa_1 = a_1 + 2a_2 = \alpha(\alpha + \beta), \\
\mu_2 &= \kappa_2 = a_1 + 4a_2 = \alpha(2\alpha + \beta), \\
\mu_3 &= \kappa_3 = a_1 + 8a_2 = \alpha(4\alpha + \beta), \\
\mu_4 &= \kappa_4 + 3\kappa_2^2 = a_1 + 16a_2 + 3(a_1 + 4a_2)^2, \\
&= \alpha(8\alpha + \beta) + 3\alpha^2(2\alpha + \beta)^2, \\
&\vdots
\end{aligned} \tag{9.72}$$

The factorial moments are given by

$$G(1 + t) = \exp\left[a_1 t + a_2(t^2 + 2t)\right] = \exp\left[(\alpha + \beta)\alpha t + \tfrac{1}{2}\alpha^2 t^2\right], \tag{9.73}$$

and so

$$\begin{aligned}
\mu'_{[r]} &= \alpha^r H_r^*(\alpha + \beta), \\
&= \alpha(\alpha + \beta)\mu'_{[r-1]} + \alpha^2(r - 1)\mu'_{[r-2]}.
\end{aligned} \tag{9.74}$$

The indices of skewness and kurtosis are

$$\begin{aligned}
\beta_1 &= \alpha_3^2 = \frac{\mu_3^2}{\mu_2^3} = \frac{(4\alpha + \beta)^2}{\alpha(2\alpha + \beta)^3}, \\
\beta_2 &= \alpha_4 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{(8\alpha + \beta)}{\alpha(2\alpha + \beta)^2}.
\end{aligned} \tag{9.75}$$

A number of models give rise to the Hermite distribution. Clearly it is the special case $n = 2$ of the Poisson–binomial distribution, the widespread use of which for biometrical data was stimulated by Skellam (1952) and McGuire, Brindley, and Bancroft (1957); see Section 9.5.

By Gurland's theorem, the Hermite distribution can also be regarded as a Poisson mixture of binomial distributions; we have

$$X \sim \text{Binomial}(N, p) \bigwedge_{N/2} \text{Poisson}(\lambda) \tag{9.76}$$

[cf. (9.61)], giving

$$G(z) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!}(1 - p + pz)^{2j}. \tag{9.77}$$

The Hermite distribution appeared in the highly innovative paper by McKendrick (1926), which was quoted extensively by Irwin (1963). McKendrick derived the distribution from the sum of two correlated Poisson rv's. The bivariate Poisson distribution can be defined as $(X_1, X_2) = (U + V, U + W)$, where U, V, and W are independent Poisson variables with parameters a_2, c_1, and c_2, respectively (Ahmed, 1961). The joint pgf of X_1 and X_2 is

$$\begin{aligned} G(z_1, z_2) &= E[z_1^{X_1} z_2^{X_2}] = E[z_1^V z_2^W (z_1 z_2)^U] \\ &= \exp[c_1 z_1 + c_2 z_2 + a_2(z_1 z_2) - c_1 - c_2 - a_2]. \end{aligned} \tag{9.78}$$

The pgf of $X = X_1 + X_2$ is

$$g(z) = E[z^{X_1+X_2}] = \exp[(c_1 + c_2)z + a_2 z^2 - c_1 - c_2 - a_2], \tag{9.79}$$

and therefore the distribution of X is Hermite with parameters $a_1 = c_1 + c_2$ and a_2. McKendrick fitted the distribution to counts of bacteria in leucocytes and obtained a very much better fit than with a Poisson distribution. Like certain other Poisson–stopped sum (multiple Poisson) distributions, the Hermite distribution can have any number of modes. As an example, the fitted distribution for McKendrick's data has $a_1 = 0.0135$, $a_2 = 0.0932$, and the first five calculated probabilities are 0.899, 0.012, 0.084, 0.001, 0.004.

The Hermite distribution can also be derived as the sum $X = Y_1 + Y_2$, where Y_1 is a Poisson rv with parameter a_1 and support $0, 1, 2, \ldots$ and Y_2 is a Poisson rv with parameter a_2 and support $0, 2, 4, \ldots$, that is, as the convolution

$$X \sim \text{Poisson singlet}(a_1) * \text{Poisson doublet}(a_2). \tag{9.80}$$

The good fit to McKendrick's data could also be explained by supposing that some of the bacteria occurred as singletons and the rest occurred in pairs.

Kemp and Kemp also examined the Hermite distribution as a penultimate limiting form for other multiple Poisson distributions. We have

$$\begin{aligned} &\lim_{a_i \to 0, i>2} \exp(a_1 z + a_2 z^2 + a_3 z^3 + \cdots - a_1 - a_2 - a_3 - \cdots) \\ &\qquad = \exp(a_1 z + a_2 z^2 - a_1 - a_2). \end{aligned} \tag{9.81}$$

When a_2 also tends to zero, the limiting form is of course Poissonian. The negative binomial, Neyman type A, and Pólya–Aeppli are multiple Poisson distributions with generalizing distributions that are logarithmic, Poisson, and geometric, respectively. When the generalizing distributions are reverse-J shaped the corresponding multiple Poisson distributions are often approximated well by Hermite distributions.

McKendrick (1926) fitted the distribution using the method of moments. Sprott (1958) derived a maximum-likelihood method for the more general Poisson–binomial distribution (see Section 9.5) and applied it in the particular (Hermite) case $n = 2$. Kemp and Kemp (1965) reexamined and discussed maximum-likelihood estimation for the Hermite distribution using both a "false position" and a Newton–Raphson iterative procedure. Because the distribution is a power series distribution and also has a pgf of the form $G(z) = h(\alpha(z-1))$, the maximum-likelihood equations are

$$\overline{x} = \hat{\alpha}(\hat{\alpha} + \hat{\beta}) \qquad \text{and} \qquad \overline{x} = \sum_j (j+1)\frac{f_j}{N}\frac{\hat{p}_{j+1}}{\hat{p}_j} \tag{9.82}$$

in the notation of Section 9.2. Alternatively, use of the recursion formula (9.70) for the probabilities together with (9.82) gives

$$\overline{x} = \hat{\alpha}(\hat{\alpha} + \hat{\beta}) \qquad \text{and} \qquad 1 = \sum_j \frac{f_j}{N}\frac{\hat{p}_{j-1}}{\hat{p}_j}. \tag{9.83}$$

In both cases the solution requires iteration.

Patel (1971) made a very thorough comparative study of various estimation procedures for the Hermite distribution in his doctoral thesis. These included maximum likelihood, moment estimators, mean-and-zero-frequency estimators, and the method of even points [for the latter see, e.g., Kemp and Kemp (1989)]. Much of this work appeared subsequently in journal papers; see Patel, Shenton, and Bowman (1974) and Patel (1976b, 1977).

Kemp and Kemp (1966) showed that formally the Hermite distribution can be obtained by mixing a Poisson distribution with parameter θ using a normal mixing distribution for θ; see also Greenwood and Yule (1920). This model assumes, however, that θ can take all real values, not merely nonnegative ones, and so lacks physical interpretation. For this model the recursion formula for the probabilities becomes

$$(x+1)\Pr[X = x+1] = (\mu - \sigma^2)\Pr[X = x] + \sigma^2 \Pr[X = x-1], \qquad x \ge 0, \tag{9.84}$$

with $\Pr[X = -1] = 0$ and $\Pr[X = 0] = e^{-\mu+\sigma^2/2}$. Here μ and σ^2 are the mean and variance of the normal distribution on $(-\infty, \infty)$. A necessary restriction on the parameters of the normal mixing distribution is $\mu \ge \sigma^2$.

A closely related distribution, obtained by letting θ have a normal distribution truncated to the left at the origin (the *Poisson truncated normal distribution*), was found by Kemp and Kemp (1967) to be a special case of Fisher's (1931) *modified Poisson distribution*. The Poisson truncated normal distribution was also studied by Berljand, Nazarov, and Pressman (1962) and Patil (1964a). Its pmf is

$$\Pr[X = x] = \exp\left(\frac{\sigma^2}{2} - \mu\right)\sigma^x \frac{I_x(\sigma - \mu/\sigma)}{I_0(-\mu/\sigma)}, \qquad x \ge 0, \tag{9.85}$$

where

$$\begin{aligned} I_0(x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}\,dt, \\ I_r(x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{(t-x)^r}{r!} e^{-t^2/2}\,dt, \end{aligned} \tag{9.86}$$

and μ and σ^2 are again the mean and variance of the complete normal distribution. Kemp and Kemp found that a more convenient method of handling the probabilities is via

$$\begin{aligned}
\Pr[X=0] &= \exp\left(\frac{\sigma^2}{2}-\mu\right)\frac{I_0(\sigma-\mu/\sigma)}{I_0(-\mu/\sigma)},\\
\Pr[X=1] &= \Pr[X=0]\sigma\frac{I_1(\sigma-\mu/\sigma)}{I_0(\sigma-\mu/\sigma)},\\
(x+1)\Pr[X=x+1] &= (\mu-\sigma^2)\Pr[X=x]+\sigma^2\Pr[X=x-1], \qquad x\ge 1
\end{aligned} \tag{9.87}$$

[cf. (9.84)]. Alternatively, the pgf can be expressed as

$$G(z)=\left\{\sum_{j=0}^{\infty}\left[\frac{\mu+\sigma^2(z-1)}{\sigma\sqrt{2}}\right]^j\Big/\left(\frac{j}{2}\right)!\right\}\div\left\{\sum_{j=0}^{\infty}\left[\frac{\mu}{\sigma\sqrt{2}}\right]^j\Big/\left(\frac{j}{2}\right)!\right\}, \tag{9.88}$$

and the probabilities derived therefrom by expansion. The various formulas simplify considerably when $\mu=0$ and when $\mu=\sigma^2$. The closeness of this distribution to the Hermite distribution increases as μ/σ^2 increases.

Gupta and Jain (1974) have investigated an *extended form* of the Hermite distribution with $X=X_1+mX_2$, where X_1 and X_2 are independent Poisson rv's. The pgf is

$$G(z)=\exp(a_1z+a_mz^m-a_1-a_m), \tag{9.89}$$

and the pmf is

$$\begin{aligned}
\Pr[X=0|m] &= e^{-a_1-a_m},\\
\Pr[X=r|m] &= \Pr[X=0|m]\sum_{j=0}^{[x/m]}\frac{(a_m)^j a_1^{x-mj}}{j!(x-mj)!}.
\end{aligned} \tag{9.90}$$

The cumulants are $\kappa_r=a_1+m^ra_m$, and hence

$$\mu=a_1+ma_m \qquad \text{and} \qquad \mu_2=a_1+m^2a_m. \tag{9.91}$$

The *Gegenbauer distribution* was derived by Plunkett and Jain (1975) by mixing a Hermite distribution with pgf

$$G(z)=\exp[\gamma(z-1)+\gamma\rho(z^2-1)]$$

using a gamma distribution for γ; see Section 11.2.7.

Patil and Raghunandanan (1990) allowed all the parameters $a_i, i = 1, \ldots, k$, in

$$G(z) = \exp[a_1(z-1) + a_2(z^2-1) + \cdots + a_k(z^k-1)]$$

to have gamma distributions and thereby obtained a distribution with pgf of the form

$$G(z) = \prod_{i=1}^{k} \left(\frac{1-q_i}{1-q_i z^i} \right)^{n_i}, \tag{9.92}$$

which they called a *stuttering negative binomial distribution*. When $a_i = a$, $i = 1, 2, \ldots, k$, this is the *negative binomial distribution of order k* discussed by Panaretos and Xekalaki (1986a); see Section 10.7.3. Patil and Raghunandanan (1990) have also studied the distribution that arises when $a_1 = a_2 = \cdots = a_k = a$ and a has a reciprocal gamma distribution.

The *Borel–Hermite* distribution was developed by Jain and Plunkett (1977) by replacing the Poisson singlet and doublet distributions in (9.80) with singlet and doublet forms of Consul's "generalized Poisson" distribution; see Section 9.11. The resultant distribution has pmf

$$\Pr[X = x] = \sum_{j=0}^{[x/2]} \left[\frac{m_1(m_1+j)^{j-1} e^{-\lambda_1(m_1+j)} m(m+x-2j)^{x-2j-1} \lambda^{x-2j} e^{-\lambda(m+x-2j)}}{j!(x-2j)!} \right] \tag{9.93}$$

and mean and variance

$$\mu = \frac{m\lambda}{1-\lambda} + \frac{2m_1\lambda_1}{1-\lambda_1} \qquad \text{and} \qquad \mu_2 = \frac{m\lambda}{(1-\lambda)^3} + \frac{4m_1\lambda_1}{(1-\lambda_1)^3}. \tag{9.94}$$

Here there are four parameters. Jain and Plunkett defined four cases of the distribution (A, B, C, and D) by imposing restrictions on the parameters, thus reducing their number to either two or three. When fitting the distribution to McKendrick's leucocyte data, Jain and Plunkett used estimation by the method of moments.

9.5 POISSON–BINOMIAL DISTRIBUTION

The Poisson–binomial distribution has the representation

$$X \sim \text{Poisson}(\lambda) \bigvee \text{Binomial}(n, p) \sim \text{Binomial}(N, p) \bigwedge_{N/n} \text{Poisson}(\lambda), \tag{9.95}$$

where *N/n* takes integer values; see Section 8.3.2. The distribution was discussed by Skellam (1952) and fitted by him to quadrat data on the sedge *Carex flacca*. In McGuire et al. (1957) it was used to represent variation in the numbers of corn-borer larvae in randomly chosen areas of a field. Since then there have been a number of biometric applications. Computational difficulties have been overcome with improved computational facilities and the development of new formulas.

The pgf is

$$G(z) = \exp\{\lambda[(q + pz)^n - 1]\}. \tag{9.96}$$

Expansion as an infinite series gives

$$\begin{aligned} \Pr[X = x] &= e^{-\lambda} \sum_{j \geq x/n} \frac{\lambda^j}{j!} \binom{nj}{x} p^x q^{nj-x} \\ &= \frac{e^{-\lambda}(p/q)^x}{x!} \sum_{j \geq x/n} \frac{(nj)!(\lambda q^n)^j}{(nj - x)!/j!}. \end{aligned} \tag{9.97}$$

An alternative expression for the probabilities can be obtained in terms of $\mu^*_{[x]}$, the xth factorial moment of nY, where Y has a Poisson distribution with expected value λ; see Shumway and Gurland (1960a,b).

The recurrence relation

$$\Pr[X = x + 1] = \frac{np\lambda}{x + 1} \sum_{j=0}^{x} \binom{n - 1}{j} p^j q^{n-1-j} \Pr[X = x - j], \tag{9.98}$$

$x = 0, 1, \ldots,$ can be used to calculate successive probabilities starting from

$$\Pr[X = 0] = \exp[\lambda(q^n - 1)]. \tag{9.99}$$

In particular

$$\begin{aligned} \Pr[X = 1] &= \exp[\lambda(q^n - 1)]np\lambda q^{n-1}, \\ \Pr[X = 2] &= \tfrac{1}{2} \exp[\lambda(q^n - 1)]np^2\lambda q^{n-2}(n - 1 + nq^n\lambda). \end{aligned} \tag{9.100}$$

If $npq^{n-1}\lambda < 1$, then there is a mode at the origin. Note that, if $npq^{n-1}\lambda < 1$ and $(n - 1 + nq^n\lambda) > 2q/p$ as well, then $\Pr[X = 0] > \Pr[X = 1] < \Pr[X = 2]$ and the distribution is at least bimodal. (Multimodality is a typical property of many Poisson–stopped sum distributions.) Douglas (1980) has suggested that, if there are many modes, then, when n is large, they will be roughly at $n, 2n, 3n \ldots$ because the distribution will then tend to a Neyman type A distribution (see below). He found that this will also happen as $p \to 1$, when the limiting form is an n-let Poisson distribution with support $0, n, 2n, 3n, \ldots$.

The cumulants of the Poisson–binomial distribution can be obtained from the uncorrected moments of the binomial distribution. Simpler still are the factorial cumulants. The fcgf is

$$\ln G(1 + t) = \lambda(1 + pt)^n - \lambda, \tag{9.101}$$

giving $\kappa_{[r]} = \lambda p^r n!/(n-r)!$. The first four moments are

$$\begin{aligned}
\mu &= \lambda np, \\
\mu_2 &= \lambda n^2p^2 + \lambda npq, \\
\mu_3 &= \lambda n^3p^3 + 3\lambda n^2p^2q + \lambda npq(1-2p), \\
\mu_4 &= \lambda n^4p^4 + 6\lambda n^3p^3q + \lambda n^2p^2q(7-11p) \\
&\quad + \lambda npq(1-6pq) + 3\lambda^2(n^2p^2+npq)^2.
\end{aligned} \tag{9.102}$$

The factorial cumulant generating function has been used by Douglas (1980) to explore limiting forms of the distribution. If the generalizing binomial distribution tends to a Poisson distribution as $n \to \infty$, then the limiting form is Neyman type A (see Section 9.6).

Estimation for the three parameters n, λ, and p is made difficult by the integer restriction on n. A common estimation practice has been to estimate λ and p for several fixed values of n (say, $n = 2, 3, 4$) and to choose the set of three values for n, λ, and p that appear to give the best fit judged by some criterion such as a χ^2 goodness-of-fit test. See also Section 9.9, where estimation for the distribution as a member of the Poisson–Katz family is discussed.

Consider now estimation for λ and p, given a sample of N observed values $x_1, x_2, \ldots, x_N$ from the distribution (9.96), n being supposed known. The equations satisfied by the MLEs $\hat{\lambda}$ and $\hat{p}$ of λ and p were obtained in the following form by Sprott (1958):

$$\overline{x} = \sum_{j=1}^{N} \frac{x_j}{N} = n\hat{\lambda}\hat{p} \qquad \text{and} \qquad \sum_{j=1}^{N} \frac{(x_j+1)\hat{p}_{x_{j+1}}}{\hat{p}_{x_j}} = Nn\hat{\lambda}\hat{p}, \tag{9.103}$$

where $\hat{p}_{x_j}$ is $\Pr[X = x_j]$ with λ and p replaced by $\hat{\lambda}$ and $\hat{p}$, respectively. Shumway and Gurland (1960b) introduced an alternative formulation of these equations. The maximum-likelihood equations have to be solved iteratively. Martin and Katti (1965) remarked on the practical difficulty of obtaining convergence using Newton–Raphson iteration. It is important therefore to begin with good initial estimates.

There are several ways of obtaining initial estimates. The method of moments involves equating the first and second sample moments, giving

$$\tilde{p} = \frac{s^2 - \overline{x}}{(n-1)\overline{x}}, \qquad \tilde{\lambda} = \frac{\overline{x}}{n\tilde{p}}; \tag{9.104}$$

see Sprott (1958). An alternative method of estimation uses the equations

$$n\lambda^* p^* = \overline{x}, \qquad n\lambda^* p^*(q^*)^{n-1} = \frac{f_1}{f_0}, \tag{9.105}$$

where f_j/N denotes the proportion of j's among the N observations. These equations give

$$p^* = 1 - q^* = 1 - \left(\frac{f_1}{f_0\bar{x}}\right)^{1/(n-1)}, \qquad \lambda^* = \frac{\bar{x}}{np^*}. \tag{9.106}$$

Instead of the ratio f_1/f_0, the proportion f_0/N alone might be used. Distinguishing this case by primes, we have

$$n\lambda' p' = \bar{x}, \qquad \exp\{-\lambda'[1 - (q')^n]\} = \frac{f_0}{N},$$

whence

$$\frac{\bar{x}}{\ln(f_0/N)} = \frac{np'}{(q')^n - 1}, \tag{9.107}$$

which can be solved by iteration. Katti and Gurland (1962b) found that this last method is markedly superior to that using the first and second moments.

Linear minimum χ^2 estimation for the Poisson–binomial distribution was studied by Katti and Gurland (1962b) and Hinz and Gurland (1967). Fitting a truncated Poisson–binomial distribution, with the zero class missing, was studied by Shumway and Gurland (1960b).

9.6 NEYMAN TYPE A DISTRIBUTION

9.6.1 Definition

This is the Poisson–stopped-sum–Poisson distribution that was discussed at the beginning of Section 9.1; its symbolic representation is

$$X \sim \text{Poisson}(\lambda) \bigvee \text{Poisson}(\phi). \tag{9.108}$$

The pgf is

$$G(z) = \exp[\lambda(e^{\phi(z-1)} - 1)]. \tag{9.109}$$

[David and Moore (1954) called ϕ the *index of clumping*.] By Gurland's theorem (Section 8.3.2),

$$X \sim \text{Poisson}(\lambda) \bigvee \text{Poisson}(\phi) \sim \text{Poisson}(\Theta) \bigwedge_{\Theta/\phi} \text{Poisson}(\lambda). \tag{9.110}$$

The distribution is therefore both a Poisson–stopped sum of Poisson distributions and a Poisson mixture of Poisson distributions (Section 8.3.2).

There are two standard expressions for the probabilities, corresponding to two different ways of expanding the pgf. First,

$$G(z) = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j e^{-j\phi} e^{j\phi z}}{j!} = e^{-\lambda} \sum_{j=0}^{\infty} \left[\frac{(\lambda e^{-\phi})^j}{j!} \sum_{x=0}^{\infty} \frac{(j\phi z)^x}{x!} \right], \tag{9.111}$$

whence

$$\Pr[X = x] = \frac{e^{-\lambda}\phi^x}{x!} \sum_{j=0}^{\infty} \frac{(\lambda e^{-\phi})^j j^x}{j!}, \qquad x = 0, 1, \ldots. \tag{9.112}$$

Alternatively,

$$\begin{aligned} G(z) &= e^{-\lambda+\lambda e^{-\phi}} \sum_{j=0}^{\infty} \frac{\lambda^j e^{-j\phi}(e^{\phi z} - 1)^j}{j!} \\ &= e^{-\lambda+\lambda e^{-\phi}} \sum_{j=0}^{\infty} \left[(\lambda e^{-\phi})^j \sum_{x=j}^{\infty} \frac{S(x, j)\phi^x z^x}{x!} \right] \end{aligned} \tag{9.113}$$

using the generating expression for Stirling numbers of the second kind in Section 1.1.3. Hence

$$\Pr[X = x] = \frac{e^{-\lambda+\lambda e^{-\phi}}\phi^x}{x!} \sum_{j=1}^{x} S(x, j)\lambda^j e^{-j\phi}. \tag{9.114}$$

This expression was given by Cernuschi and Castagnetto (1946); it can also be obtained from (9.42). Whereas (9.112) involves an infinite series, (9.114) involves only a polynomial of degree x in $\lambda e^{-\phi}$; (9.114) is therefore more useful for computation of the probabilities for low values of x. In particular

$$\begin{aligned} \Pr[X = 0] &= e^{-\lambda+\lambda e^{-\phi}}, \\ \Pr[X = 1] &= \lambda\phi e^{-\phi} e^{-\lambda+\lambda e^{-\phi}}, \\ \Pr[X = 2] &= \tfrac{1}{2}\lambda\phi^2 e^{-\phi}(1 + \lambda e^{-\phi})e^{-\lambda+\lambda e^{-\phi}}. \end{aligned} \tag{9.115}$$

It will be noticed that $\Pr[X = 0] > \Pr[X = 1]$ if $\lambda\phi e^{-\phi} < 1$; moreover $\Pr[X = 1] < \Pr[X = 2]$ if $1 < \phi(\lambda e^{-\phi} + 1)/2$. Barton (1957) made a systematic study of the modality of the type A distribution and drew an interesting diagram of the regions of the parameter space having a given number of modes; whenever the distribution has more than one mode, one of the modes is at the origin. The degree of multimodality increases as ϕ increases; this is to be expected since for larger ϕ the accretion of each additional "group" makes a larger contribution to the overall distribution. The modal values are *approximately* multiples of ϕ. Shenton and Bowman (1967) calculated modal values of X for several type A distributions, illustrating this point. For example, with $\lambda = 7$ and $\phi = 25$, the modal values are 0, 25, 50, 76, 103, 129, 153, 173.

Beall (1940) derived the recurrence relation

$$\Pr[X = x] = \frac{\lambda\phi e^{-\phi}}{x} \sum_{j=0}^{x-1} \frac{\phi^j}{j!} \Pr[X = x - 1 - j]. \tag{9.116}$$

Douglas (1955) pointed out that

$$\Pr[X = x] = \frac{\phi^x}{x!}\mu'^*_x \Pr[X = 0], \tag{9.117}$$

where μ'^*_x is the xth moment about zero for a Poisson distribution with expected value $\lambda e^{-\phi}$.

9.6.2 Moment Properties

The cumulants for the Neyman type A distribution are proportional to the uncorrected moments of a Poisson distribution with parameter ϕ (the constant of proportionality is λ). Shenton (1949) established the following recurrence relations among the cumulants:

$$\kappa_{r+1} = \phi\left\{\sum_{j=0}^{r-1}\binom{r}{j}\kappa_{r-j} + \lambda\right\}, \tag{9.118}$$

$$\kappa_{r+1} = \phi\left(\kappa_r + \frac{\partial\kappa_r}{\partial\lambda}\right). \tag{9.119}$$

These are similar to the recurrence relations for the uncorrected moments of the Poisson distribution.

The factorial cumulants can be obtained from $\ln G(1+t) = \lambda(e^{\phi t} - 1)$ and hence are

$$\kappa_{[r]} = \lambda\phi^r, \qquad r = 1, 2, \ldots. \tag{9.120}$$

From the factorial cumulants, or otherwise, the rth factorial moment is

$$\mu'_{[r]} = \phi^r \sum_{j=1}^{r} \frac{\Delta^j 0^r}{j!}\lambda^r = \phi^r \sum_{j=1}^{r} S(r, j)\lambda^j, \tag{9.121}$$

where $S(r, j)$ is the Stirling number of the second kind. It follows that

$$\begin{aligned}
\mu &= \lambda\phi,\\
\mu_2 &= \lambda\phi(1+\phi),\\
\mu_3 &= \lambda\phi(1+3\phi+\phi^2),\\
\mu_4 &= \lambda\phi(1+7\phi+6\phi^2+\phi^3) + 3\lambda^2\phi^2(1+\phi)^2.
\end{aligned} \tag{9.122}$$

The moment ratios are

$$\begin{aligned}
\beta_1 = \alpha_3^2 &= \frac{\mu_3^2}{\mu_2^3} = \frac{(1+3\phi+\phi^2)^2}{\lambda\phi(1+\phi)^3},\\
\beta_2 = \alpha_4 &= \frac{\mu_4}{\mu_2^2} = 3 + \frac{1+7\phi+6\phi^2+\phi^3}{\lambda\phi(1+\phi)^2}.
\end{aligned} \tag{9.123}$$

We recall that $\beta_2 - \beta_1 - 3 = 0$ for the Poisson distribution. For the Neyman type A distribution we find that

$$\frac{\beta_2 - 3}{\beta_1} = \frac{(1 + 7\phi + 6\phi^2 + \phi^3)(1 + \phi)}{(1 + 3\phi + \phi^2)^2}. \tag{9.124}$$

This ratio is independent of λ. When $\phi = 0$, it is equal to unity. It rises to a maximum value of approximately 1.215 near $\phi = 0.5$, falling slowly thereafter to 1 as ϕ tends to infinity:

ϕ	0	1	2	3	4	5	7	10	20
$(\beta_2 - 3)/\beta_1$	1	1.20	1.16	1.14	1.12	1.11	1.09	1.07	1.04

The narrow limits of this ratio restrict the field of applicability of the distribution. The Neyman type B and type C and the more general contagious distributions described later in this chapter in Section 9.9 extend the flexibility of this class of distributions.

9.6.3 Tables and Approximations

Values of $\Pr[X = x]$ were tabulated by Grimm (1964). Douglas (1955) provided tables of μ'^*_{k+1}/μ'^*_k to assist in using (9.117) to calculate the probabilities.

Martin and Katti (1962) considered three approximations to the Neyman type A distribution, applicable when the parameters λ and ϕ take "extreme" values. These approximations are in fact limiting forms of the distribution:

Limiting Form I The distribution of the standardized variable

$$Y = (X - \lambda\phi)[\lambda\phi(1 + \phi)]^{-1/2}$$

is approximately unit normal. This approximation is useful when λ is large and the mean ($=\lambda\phi$) does not approach zero.

Limiting Form II If λ *is small*, then X is approximately distributed as a *modified* Poisson variable ("Poisson-with-added-zeros" distribution, in Section 8.2.3), with

$$\Pr[X = 0] \approx (1 - \lambda) + \lambda e^{-\phi},$$

$$\Pr[X = x] \approx \frac{\lambda e^{-\phi}\phi^x}{x!}, \qquad x = 1, 2, \ldots.$$

Limiting Form III If ϕ *is small*, then X is approximately distributed as a Poisson variable with expected value $\lambda\phi$.

Martin and Katti (1962) gave diagrams that provide a general picture of the regions in which these three limiting forms give useful practical approximations.

Douglas (1965) has suggested the approximate formula

$$\Pr[X = x] \approx \left(\frac{e^{-\lambda}}{\sqrt{2\pi}}\right) \frac{\phi^x \exp[x/g(x)]}{[g(x)]^x \{x[1 + g(x)]\}^{1/2}}, \tag{9.125}$$

where

$$g(x) \exp[g(x)] = x(\lambda e^{-\phi})^{-1}.$$

Bowman and Shenton (1967) quoted the following formula due to Philpot (1964):

$$\Pr[X = x] \approx \Pr[X = 0] \frac{\phi^x x_0 \exp[f(x_0) - \lambda e^{-\phi}]}{x!(x_0 + x - 0.5)^{1/2}}, \tag{9.126}$$

where

$$x_0 \left[\ln\left(\frac{x_0 e^{\phi}}{\lambda}\right)\right] = x - 0.5$$

and

$$f(x_0) = x_0 + (x - 0.5)[\ln(\lambda e^{-\phi}) + x_0^{-1}(x - x_0 - 0.5)].$$

These authors also considered approximations of the form

$$\Pr[X = x] \approx \sum_{j=1}^{s} A_{j,s} \frac{e^{-\theta_j} \theta_j^x}{x!},$$

with the A's and θ_j's chosen to give the correct values for the first $2s - 1$ moments.

9.6.4 Estimation

The two parameters of the type A distribution are λ and ϕ. The distribution is a PSD, and also the pgf is a function of $\phi(z - 1)$. Given observations on N independent rv's $X_1, X_2, \ldots, X_N$, each having the same Neyman type A distribution, the maximum-likelihood equations are therefore

$$\begin{aligned} \sum_{j=1}^{N} \frac{x_j}{N} &= \overline{x} = \hat{\lambda}\hat{\phi}, \\ \sum_{j=1}^{N} \frac{(x_j + 1)\Pr[X = x_j + 1 | \hat{\lambda}, \hat{\phi}]}{\Pr[X = x_j | \hat{\lambda}, \hat{\phi}]} &= N\overline{x}, \end{aligned} \tag{9.127}$$

where $\hat{\lambda}$ and $\hat{\phi}$ are the maximum-likelihood estimates of λ and ϕ.

These equations do not have an explicit solution. Shenton (1949) used a Newton–Raphson iterative procedure to solve them. Tables in Douglas (1955) can be used to facilitate the calculations. Computer methods for solving the equations were discussed in Douglas (1980); see also Douglas (1955). Comments on difficulties associated with the MLEs were made by Shenton and Bowman (1967, 1977).

Initial values can be obtained by the method of moments. This gives

$$\tilde{\lambda} = \frac{\overline{x}}{\tilde{\phi}}, \qquad \tilde{\phi} = \frac{s^2 - \overline{x}}{\overline{x}}. \tag{9.128}$$

For $\phi < 0.2$ the asymptotic efficiency of the moment estimators is at least 85% (whatever the value of λ), while for $0.2 < \phi < 1.0$ the efficiency is generally between 75 and 90%. A table of efficiencies is in Shenton (1949). Katti and Gurland (1962a), Bowman and Shenton (1967), and Shenton and Bowman (1977) have explored the efficiency of the moment estimators with respect to the MLEs.

Shenton and Bowman (1967) showed that even in samples of size 100 there are substantial biases in both the maximum-likelihood and moment estimators. For λ the bias is positive; for ϕ it is negative.

A third method of estimation uses the sample mean ($\overline{x}$) and the observed proportion of zeros (f_0/N). The equations for these estimators, λ^*, ϕ^*, are

$$\overline{x} = \lambda^*\phi^*, \qquad \frac{f_0}{N} = \exp[\lambda^*(e^{-\phi^*} - 1)]. \tag{9.129}$$

Elimination of λ^* between these equations gives the following equation for ϕ^*:

$$\frac{\overline{x}}{\ln(f_0/N)} = \frac{\phi^*}{e^{-\phi^*} - 1}. \tag{9.130}$$

A fourth method uses the ratio of the frequencies of 1's and 0's, f_1/f_0, and the sample mean. This leads to the equations

$$\overline{x} = \lambda^{**}\phi^{**}, \qquad \frac{f_1}{f_0} = \lambda^{**}\phi^{**}e^{-\phi^{**}},$$

whence

$$\phi^{**} = \ln\left(\frac{\overline{x}f_0}{f_1}\right), \qquad \lambda^{**} = \frac{\overline{x}}{\phi^{**}}. \tag{9.131}$$

Katti and Gurland (1962a), Bowman and Shenton (1967), and Shenton and Bowman (1977) also studied the efficiencies of the two latter pairs of estimators. The method of moments would seem to be the best of the three if λ is greater than about 5 for most values of ϕ. For $\lambda < 4.5$, the estimators λ^*, ϕ^* appear to be preferable.

Katti (1965) showed how the different methods of estimation can be combined with some gain in efficiency. This idea was taken further by Gurland (1965) and Hinz and Gurland (1967), who developed a generalized minimum χ^2 method of estimation based on the first few moments and a function of the zero frequency;

this method has very high asymptotic efficiency. Its implementation is described in detail in Douglas (1980). Hinz and Gurland (1967) also studied goodness-of-fit tests for the distribution. The modified test of Bhalerao, Gurland, and Tripathi (1980) has high power.

If a series of sample values are available and it is only desired to estimate ϕ (the mean number per cluster), the following method of estimation has found favor among practical workers. This is simply to plot $\ln(f_0/N)$ against $\overline{x}$. Since

$$\frac{\ln \Pr[X = 0|\lambda, \phi]}{E[X]} = \frac{e^{-\phi} - 1}{\phi}, \tag{9.132}$$

a value of ϕ can be estimated from the slope of the graph; see the discussion about this method of estimation by Pielou (1957).

Maximum-likelihood, moment, and minimum χ^2 estimation for a common value of ϕ (or alternatively for a common value of λ), given a series of samples, were investigated at length by Douglas (1980). Hinz and Gurland (1968, 1970) examined tests of linear hypotheses for series of samples from different Neyman type A distributions. Gurland (1965), Hinz and Gurland (1967), and Grimm (1970) investigated graphical methods for assessing the suitability of a Neyman type A model.

9.6.5 Applications

The Neyman type A distribution has been used to describe plant distributions, especially when reproduction of the species produces clusters. This frequently happens when the species is generated by offshoots from parent plants or by seeds falling near the parent plant. Archibald (1948) found, however, that there is not enough evidence to make an induction from the type of fitted distribution to the type of reproduction. Evans (1953) found that while the Neyman type A gave good results for plant distributions, negative binomial distributions (Chapter 5) were better for insect distributions; see also Wadley (1950). Martin and Katti (1965) fitted 35 data sets with a number of standard distributions; they found that both the negative binomial and the Neyman type A have wide applicability.

Pielou (1957) also has investigated the use of Neyman type A distributions in ecology. She found that the distribution is unlikely to be applicable to plant populations unless the clusters of plants are so compact as not to lie across the edge of the quadrat used to select sample areas. Skellam (1958) pointed out that the compactness of the clusters is a hidden assumption in Neyman's original derivation of the distribution. The choice of quadrat size was found to greatly affect the results.

Cresswell and Froggatt (1963) [see also Kemp (1967b)] derived the Neyman type A in the context of bus driver accidents on the basis of the following assumptions:

1. Every driver is liable to have "spells," where the number of spells per driver in a given period of time is Poissonian, with the same parameter λ for all drivers.

2. The performance of a driver during a spell is substandard; he is liable to have a Poissonian number of accidents during a spell with the same parameter ϕ for all drivers.
3. Each driver behaves independently.
4. No accidents can occur outside of a spell.

These assumptions lead to a Neyman type A distribution via the Poisson(λ) $\bigvee$ Poisson(ϕ) model. Cresswell and Froggatt named this the "long distribution," because of its long tail, in contradistinction to their "short distribution" (for which see Section 9.9). Irwin (1964) remarked that a type A distribution can also be derived by assuming that different drivers have differing levels K of proneness, taking values $0, \phi, 2\phi, 3\phi, \ldots$, with probability $\Pr[K = k\phi] = \exp(-\lambda)\lambda^k/k!$, and that a driver with proneness $k\phi$ has X accidents where $\Pr[X = x|k\phi] = \exp(-k\phi)(k\phi)^x/x!$. This is the Poisson($K\phi$) $\bigwedge$ Poisson(λ) model with mixing over the values taken by K.

Rogers (1965, 1969) has applied the Neyman type A distribution to the clustering of retail food stores.

It has been suggested by David and Moore (1954) that complete distributions need not be fitted to data if one only wants to estimate indices of clustering (i.e., contagiousness) or the mean number of entities per cluster.

9.7 PÓLYA–AEPPLI DISTRIBUTION

The *Pólya–Aeppli distribution* arises in a model formed by supposing that objects (which are to be counted) occur in clusters, the number of clusters having a Poisson distribution, while the number of objects per cluster has the geometric distribution with pmf

$$\Pr[Y = y] = qp^{k-1}, \qquad x = 1, 2, \ldots, \qquad q = 1 - p. \tag{9.133}$$

The Pólya–Aeppli distribution can therefore be represented as

$$X \sim \text{Poisson}(\theta) \bigvee \text{Shifted geometric}(p). \tag{9.134}$$

It also arises as a generalization of the Poisson distribution with parameter θ/p using an unshifted geometric distribution (with parameter p) as the generalizing distribution; hence it can be represented as

$$X \sim \text{Poisson}\left(\frac{\theta}{p}\right) \bigvee \text{Geometric}(p). \tag{9.135}$$

The Pólya–Aeppli distribution was described by Pólya (1930); he ascribed the derivation of the distribution to his student Aeppli in a Zurich thesis in 1924. It is a special case of the Poisson–Pascal distribution; the latter is sometimes called the generalized Pólya–Aeppli distribution (see Section 9.8).

The distribution is defined by the pgf

$$G(z) = \exp\left[\theta\left(\frac{(1-p)z}{1-pz} - 1\right)\right] = \exp\left[\theta\left(\frac{z-1}{1-pz}\right)\right] \tag{9.136}$$

$$= \exp\left[\frac{\theta}{p}\left(\frac{1-p}{1-pz} - 1\right)\right]. \tag{9.137}$$

Expression (9.136) corresponds to model (9.134), while expression (9.137) corresponds to model (9.135). In the literature, formulas for the Pólya–Aeppli distribution are given either in terms of θ and p or alternatively in terms of ζ and p, where $\zeta = \theta/p$. This led to a confusion of notations in the first edition of this book.

By direct expansion of the pgf

$$\Pr[X = 0] = e^{-\theta},$$

$$\Pr[X = x] = e^{-\theta} p^x \sum_{j=1}^{x} \binom{x-1}{j-1} \frac{(\theta q/p)^j}{j!} \tag{9.138}$$

$$= e^{-\theta}\left(\frac{\theta q}{p}\right) p^x \, {}_1F_1\left[1-x; 2; -\frac{\theta q}{p}\right], \qquad x = 1, 2, \ldots,$$

where ${}_1F_1[\cdot]$ is the confluent hypergeometric function. Algebraic manipulation that is equivalent to the use of Kummer's transformation,

$${}_1F_1[a; b; y] = e^y \, {}_1F_1[b-a; b; -y],$$

gives

$$\Pr[X = x] = e^{-\theta/p}\left(\frac{\theta q}{p}\right) p^x \, {}_1F_1\left[x+1; 2; \frac{\theta q}{p}\right], \qquad x = 1, 2, \ldots \tag{9.139}$$

(Evans, 1953; Philipson, 1960a); see also Galliher et al. (1959) for a Laguerre polynomial expression.

Evans (1953), using the notation $m = \theta/q$, $a = 2p/q$, derived a recurrence formula for the probabilities; in our notation this becomes

$$(x+1)\Pr[X = x+1] = (\theta q + 2px)\Pr[X = x] - p^2(x-1)\Pr[X = x-1] \tag{9.140}$$

for $x = 0, 1, 2, \ldots$, with $\Pr[X = 0] = e^{-\theta}$ and $\Pr[X = -1] = 0$. An alternative recurrence formula is

$$(x+1)\Pr[X = x+1] = \theta q \sum_{j=0}^{x} (x+1-j) p^{x-j} \Pr[X = j] \tag{9.141}$$

(Kemp, 1967a). Douglas (1986) has reported that use of the three-term recurrence relation (9.140) may run into numerical difficulties because of the differencing of small terms.

Taking $\alpha = \theta q/p$ gives

$$\begin{aligned}
\Pr[X = 1] &= e^{-\theta}\theta q = e^{-\theta}\alpha p, \\
\Pr[X = 2] &= e^{-\theta}\theta q\left(p + \frac{\theta q}{2}\right) = e^{-\theta}\alpha p^2\left(1 + \frac{\alpha}{2}\right), \\
\Pr[X = 3] &= e^{-\theta}\alpha p^3\left(1 + \alpha + \frac{\alpha^2}{6}\right), \\
\Pr[X = 4] &= e^{-\theta}\alpha p^4\left(1 + \frac{3\alpha}{2} + \frac{\alpha^2}{2} + \frac{\alpha^3}{24}\right).
\end{aligned} \tag{9.142}$$

There is a mode at the origin if $\theta q < 1$. When $2 < \theta < 1/q$, there is a local minimum probability at $X = 1$; that is, the distribution has at least two modes. Douglas (1965, 1980) has obtained approximate formulas for the probabilities; these are useful for probabilities in the tail of the distribution. Evans (1953) gave inequalities that can be used to assess the cumulative probability $\Pr[X \le x - 1]$.

The Pólya–Aeppli distribution was called the *geometric Poisson* by Sherbrooke (1968), who gave tables of individual and cumulative probabilities to four decimal places.

The expected value and first three central moments of the distribution are most easily found by using the fcgf

$$\ln G(1 + t) = \theta t(q - pt)^{-1} = \frac{\theta t}{q}\left(1 - \frac{pt}{q}\right)^{-1}; \tag{9.143}$$

the rth factorial cumulant is

$$\kappa_{[r]} = \frac{r!\theta}{q}\left(\frac{p}{q}\right)^{r-1}, \qquad r = 1, 2, \ldots. \tag{9.144}$$

Hence the central moments are

$$\begin{aligned}
\mu &= \kappa_1 = \frac{\theta}{q}, \\
\mu_2 &= \kappa_2 = \frac{\theta(1 + p)}{q^2}, \\
\mu_3 &= \kappa_3 = \frac{\theta(1 + 4p + p^2)}{q^3}, \\
\mu_4 &= \kappa_4 + 3\kappa_2^2 = \frac{\theta(1 + 11p + 11p^2 + p^3)}{q^4} + \frac{3\theta^2(1 + p)^2}{q^4}.
\end{aligned} \tag{9.145}$$

From these equations we find that

$$\alpha_3^2 = \beta_1 = \frac{(1+4p+p^2)^2}{(1+p)^3\theta}, \qquad \alpha_4 = \beta_2 = 3 + \frac{1+11p+11p^2+p^3}{(1+p)^2\theta}. \tag{9.146}$$

The distribution is the limiting form as $\beta \to \infty$, $\phi \to \infty$, $\beta/(\phi+\beta) \to p$ of Beall and Rescia's generalization of the Neyman types A, B, and C distributions (see Section 9.9). The limiting form of the distribution as $p \to 0$ is a Poisson distribution with parameter θ.

We now consider the problem of estimation, given a sample of size N from the distribution with pgf (9.136), with observed values $x_1, x_2, \ldots, x_N$. The maximum-likelihood equations are equivalent to

$$\overline{x} = \frac{\hat{\theta}}{1-\hat{p}}, \qquad \overline{x} = \sum_{j=1}^{N} \frac{f_j}{N} \frac{(j-1)\widehat{P_{j-1}}}{\widehat{P_j}}, \tag{9.147}$$

that is, to

$$\overline{x} = \frac{\hat{\theta}}{1-\hat{p}}, \qquad \overline{x} = \sum_{j=1}^{N} \frac{f_j}{N} \frac{(j+1)\widehat{P_{j+1}}}{\widehat{P_j}} \tag{9.148}$$

because of the three-term recurrence relationship for the probabilities. Here $P_x = \Pr[X = x]$, f_j/N is the observed proportion of the observations that are equal to j, and $\widehat{P_j}$ denotes P_j with θ replaced by $\hat{\theta}$ and p by $\hat{p}$. The solution of these equations requires iteration.

The moment estimators $\tilde{\theta}$ and $\tilde{p}$ are

$$\tilde{\theta} = \frac{2\overline{x}^2}{s^2+\overline{x}}, \qquad \tilde{p} = \frac{s^2-\overline{x}}{s^2+\overline{x}}, \tag{9.149}$$

where s^2 is the sample variance. The biases, variances, and covariances of the maximum-likelihood and moment estimators were tabulated for certain parameter combinations by Shenton and Bowman (1977).

Use of the mean and zero frequency leads to the estimators θ^* and p^* given by

$$\theta^* = -\ln\left(\frac{f_0}{N}\right), \qquad p^* = 1 - \frac{\theta^*}{\overline{x}}, \tag{9.150}$$

where f_j/N is as above (Evans, 1953).

Another simple method of estimation is based on the first two frequencies. From $\Pr[X=0] = e^{-\theta}$ and $\Pr[X=1] = e^{-\theta}\theta q$, we have

$$\theta = -\ln(\Pr[X=0]) \qquad \text{and} \qquad q = \frac{-\Pr[X=1]}{\Pr[X=0]\ln\Pr[X=0]}.$$

This suggests using as estimators

$$\theta^{**} = -\ln\left(\frac{f_0}{N}\right), \qquad q^{**} = -\frac{f_1}{f_0 \ln(f_0/N)}. \tag{9.151}$$

For minimum χ^2 estimation see Hinz and Gurland (1967) and Douglas (1980). Tripathi, Gurland, and Bhalerao (1986) have discussed the use of the method for the Poisson–Katz family of distributions to which the Pólya–Aeppli belongs (see Section 9.9).

9.8 GENERALIZED PÓLYA–AEPPLI (POISSON–NEGATIVE BINOMIAL) DISTRIBUTION

The *Poisson–Pascal distribution* was introduced in the context of the spatial distribution of plants by Skellam (1952), who called it a *generalized Pólya–Aeppli distribution*. Katti and Gurland (1961) studied its properties and estimation and derived it from an entomological model. Consider the number of surviving larvae when the number of egg masses per plot has a Poisson distribution and the number of survivors per egg mass has a negative binomial distribution. The number of survivors per plot will then have a Poisson–stopped sum of negative binomial distributions. This is the Poisson–Pascal distribution:

$$X \sim \text{Poisson}(\theta) \bigvee \text{Negative binomial}(k, P), \tag{9.152}$$

$0 < \theta$, $0 < k$, $0 < P$. The pgf is

$$G(z) = \exp\{\theta[(Q - Pz)^{-k} - 1]\}, \tag{9.153}$$

where $Q = 1 + P$. The special case with $k = 1$, $P = p/(1 - p)$, $\theta = \theta_1/p$ is the Pólya–Aeppli distribution.

Note that the generalizing distribution in (9.153) is negative binomial with support $0, 1, 2, \ldots$ and *not* a negative binomial waiting-time distribution with support $k, k + 1, \ldots$, as the name Poisson–Pascal might suggest. A distribution with pgf $\exp\{\theta[z^k(Q - Pz)^{-k} - 1]\}$ could of course be constructed, but it would lack biological realism in the contexts in which the Poisson–Pascal distribution has been used.

Katti and Gurland pointed out that the distribution can also be obtained as

$$X \sim \text{Negative binomial}(\Psi, P) \bigwedge_{\Psi/k} \text{Poisson}(\theta), \tag{9.154}$$

with pgf

$$G(z) = E[(Q - Pz)^{\Psi}] = \exp\{\theta[(Q - Pz)^{-k} - 1]\}, \tag{9.155}$$

where Ψ/k has a Poisson distribution with parameter θ and k is a positive constant.

The individual probabilities are

$$\begin{aligned}\Pr[X=0] &= e^{\theta(Q^{-k}-1)},\\ \Pr[X=x] &= \frac{e^{-\theta}(P/Q)^x}{x!}\sum_{j=1}^{\infty}\frac{(kj+x-1)!}{(kj-1)!j!}(\theta Q^{-k})^j.\end{aligned} \tag{9.156}$$

Shumway and Gurland (1960a) have provided tables to assist in the direct calculation of $\Pr[X=x]$ from this formula. Katti and Gurland (1961) obtained a useful recurrence relation for the probabilities. Restated in the above notation, it becomes

$$\Pr[X=x+1] = \frac{\theta kP}{(x+1)Q^{k+1}}\sum_{j=0}^{x}\binom{k+x-j}{k}\left(\frac{P}{Q}\right)^{x-j}\Pr[X=j]. \tag{9.157}$$

The fcgf is

$$\ln G(1+t) = \theta[(1-Pt)^{-k}-1],$$

and hence

$$\kappa_{[r]} = \frac{\theta(k+r-1)!P^r}{(k-1)!}. \tag{9.158}$$

The mean and variance are therefore

$$\mu = \kappa_1 = \theta kP \qquad \text{and} \qquad \mu_2 = \kappa_2 = \theta kP(Q+kP); \tag{9.159}$$

see Douglas (1980) for further moment properties. The flexibility of the Poisson–Pascal compared with certain stopped-sum distributions was assessed quantitatively by Katti and Gurland (1961), who evaluated their relative skewness and kurtosis for fixed mean and variance.

As $k\to\infty$, $P\to 0$ such that $kP\to\lambda$, the generalizing negative binomial tends to a Poisson distribution and the Poisson–Pascal tends to a Neyman type A distribution. As $\theta\to\infty$, $k\to 0$ such that $k\theta\to\lambda$, the limiting form is a negative binomial distribution, and as $\theta\to\infty$, $p\to 0$ so that $\theta P\to\lambda$, a Poisson distribution is the outcome; see Katti and Gurland (1961).

Katti and Gurland described three methods of fitting the distribution and gave examples using plant quadrat data. Their first method uses the first three sample moments. Their second method obtains the estimators P^*, k^*, and θ^* via the first two sample moments and the proportion of zeros (f_0/N) in the sample. Katti and Gurland's third method is based on the first two sample moments and the ratio f_1/f_0. Iteration is necessary for all three methods. Katti and Gurland (1961) calculated the asymptotic efficiency (the ratio of the generalized variances) of each method relative to the method of maximum likelihood for a range of values of θ and P. For small values of θ (≤ 1) they found the third method to be generally the best of the three, with an efficiency greater than 90% for $k\leq 1$, $P\leq 0.5$. For larger θ, the second method appeared to be better. In general the first method was worst.

Derivation of the maximum-likelihood estimates via the maximum-likelihood equations is described in detail in Douglas (1980). The procedure parallels that for the Poisson–binomial distribution (Shumway and Gurland, 1960b). Alternatively, the maximum-likelihood estimates may be obtained by direct search of the likelihood surface; with modern computing facilities this method is straightforward. Good initial estimates are advantageous. Minimum χ^2 estimation is examined in Hinz and Gurland (1967).

9.9 GENERALIZATIONS OF NEYMAN TYPE A DISTRIBUTION

Various generalizations of the Neyman type A distribution have been sought by modifying the assumptions described in Section 9.6.1. Suppose, in addition, that while the number of larvae produced have a Neyman type A distribution, the number surviving to be observed is subjected to a Rao damage process; that is, given that m are produced, the number observed has a binomial distribution with parameters m, p. Then (Feller, 1943), the overall distribution is still a Neyman type A distribution but with parameters λ, $p\phi$. We do not get a new distribution this way.

Gurland (1958) made the further assumptions that the parameter p varies from cluster to cluster (egg mass to egg mass) and that this variation can be represented by a beta distribution. The distribution of the number of larvae per cluster now has the form

$$\text{Poisson}(\phi) \bigvee \text{Binomial}(1, P) \bigwedge_{P} \text{Beta}(\alpha, \beta), \tag{9.160}$$

that is, it is a beta–Poisson distribution with the pgf

$$\int_0^1 e^{\phi p(z-1)} \frac{p^{\alpha-1}(1-p)^{\beta-1}\, dp}{B(\alpha, \beta)} = {}_1F_1[\alpha; \alpha+\beta; \phi(z-1)]; \tag{9.161}$$

see Section 8.3.2.

Gurland assumed also that the number of egg masses per plot is Poissonian. The symbolic representation for the resultant distribution of the number of surviving larvae per plot is

$$X \sim \text{Poisson}(\lambda) \bigvee \left[\left\{ \text{Poisson}(\phi) \bigvee \text{Binomial}(1, P) \right\} \bigwedge_{P} \text{Beta}(\alpha, \beta) \right] \tag{9.162}$$

$$\sim \text{Poisson}(\lambda) \bigvee \left[\left\{ \text{Binomial}(M, P) \bigwedge_{M} \text{Poisson}(\phi) \right\} \bigwedge_{P} \text{Beta}(\alpha, \beta) \right], \tag{9.163}$$

and the pgf is

$$G(z) = \exp\left\{\lambda \left[\int_0^1 e^{\phi(1-p+pz)} \frac{p^{\alpha-1}(1-p)^{\beta-1}\,dp}{B(\alpha,\beta)}\right] - \lambda\right\},$$
$$= \exp\{\lambda\, {}_1F_1[\alpha; \alpha+\beta; \phi(z-1)] - \lambda\}. \tag{9.164}$$

This corresponds to a family of distributions with four parameters. There are the original two parameters, λ (expected number of clusters) and ϕ (expected number of larvae per cluster), and α and β defining the distribution of p (the probability of survival or, more generally, the probability of being observed).

Neyman's (1939) own generalizations of the type A distribution are obtained by putting $\alpha = 1$ and $\beta = 1$, also $\alpha = 1$ and $\beta = 2$. He called these type B and type C, respectively. (Type A corresponds to $\alpha = 1$, $\beta = 0$.) Feller (1943) gave an alternative derivation of the type B and type C distributions. For an early paper on applications of the distributions, see Beall (1940).

The subfamily which is obtained when $\alpha = 1$, $\beta \geq 0$ was studied by Beall and Rescia (1953). Members of this subfamily have pgf's that can be rewritten as

$$G(z) = e^{-\lambda} \exp\left[\lambda\Gamma(\beta+1) \sum_{j=0}^{\infty} \frac{\phi^j (z-1)^j}{\Gamma(\beta+j+1)}\right]. \tag{9.165}$$

Formulas for the probabilities and moments of the type B and type C distributions (with two parameters) and the Beall and Rescia distributions (three parameters) are obtainable as special cases of the formulas that Gurland derived for his four-parameter family. His expressions for the probabilities are

$$\Pr[X = 0] = \exp(\lambda\, {}_1F_1[\alpha; \alpha+\beta; -\phi] - \lambda),$$
$$\Pr[X = x+1] = \frac{\lambda}{x+1} \sum_{j=0}^{x} F_j \Pr[X = x-j], \qquad x = 0, 1, 2, \ldots, \tag{9.166}$$

where ${}_1F_1[\cdot]$ is a confluent hypergeometric function and F_j is defined as

$$F_j = \frac{\phi^{j+1}(\alpha+j)!(\alpha+\beta-1)!\, {}_1F_1[\alpha+j+1; \alpha+\beta+j+1; -\phi]}{j!(\alpha-1)!(\alpha+\beta+j)!} \tag{9.167}$$

when $j = 0, 1, \ldots$ and zero otherwise; F_j satisfies

$$F_j = \left(\frac{\phi+\alpha+\beta+j-1}{j}\right) F_{j-1} - \frac{\phi(\alpha+j-1)}{j(j-1)} F_{j-2}, \qquad j \geq 2. \tag{9.168}$$

(Care should be taken not to confuse the notations ${}_1F_1$ and F_j.) Gurland pointed out that the use of the recurrence relation (9.168) may lead to numerical difficulties. He recommended the use of (9.167), with (9.168) kept as a check on the calculations.

Gurland found from the fcgf that

$$\kappa_{[r]} = \frac{(\alpha + r - 1)!(\alpha + \beta - 1)!}{(\alpha - 1)!(\alpha + \beta + r - 1)!}\lambda\phi^r, \tag{9.169}$$

whence

$$\begin{aligned}
\mu &= \lambda\phi\alpha(\alpha + \beta)^{-1}, \\
\mu_2 &= \lambda\phi\alpha(\alpha + \beta)^{-1}[1 + \phi(\alpha + 1)(\alpha + \beta + 1)^{-1}], \\
\mu_3 &= \lambda\phi\alpha(\alpha + \beta)^{-1}[1 + 3\phi(\alpha + 1)(\alpha + \beta + 1)^{-1} \\
&\qquad + \phi^2(\alpha + 1)(\alpha + 2)(\alpha + \beta + 1)^{-1}(\alpha + \beta + 2)^{-1}].
\end{aligned} \tag{9.170}$$

For the estimation of λ and ϕ, *given* α and β, equating the first and second sample and population moments would appear to suffice. Maximum-likelihood estimation by direct search of the likelihood surface would, nevertheless, be straightforward, given modern computing facilities, but little is known about the properties of these estimators.

When Beall and Rescia fitted their subfamily, with $\alpha = 1$ and β, λ, and ϕ all unknown, they admitted only integer values for β, though fractional values are possible. They suggested fitting their three-parameter distribution by a method that essentially consists of first fixing β and then estimating λ and ϕ by equating first and second sample and population moments. The estimated and observed distributions are then compared by means of a χ^2-test. The process is repeated for a succession of integer values of β, and the value of β that gives the best fit to the data is selected as the estimated value of β. Fortunately, the optimal solution does not seem to be very sensitive to the exact value chosen for β.

If all four parameters are to be estimated, a quick estimation method is to equate the first two observed relative frequencies and the first two sample moments to their theoretical values (Gurland, 1958).

Gurland pointed out that, as $\beta \to \infty$ and ϕ varies so that the first two moments remain fixed, the limiting form of his family is a Pólya–Aeppli distribution (see Section 9.7). He also commented that, for a fixed general value of α, the limiting distribution is a generalized Pólya–Aeppli distribution (see Section 9.8).

Kocherlakota Subrahmaniam (1966) and Kathleen Subrahmaniam (1978) have made a special study of the Negative binomial $\bigvee$ Poisson distribution, which they called the Pascal–Poisson distribution (Section 5.12.6). They obtained it as a limiting case of a more general contagious distribution. For their more general model the number of *egg masses per plot* is assumed to be Poissonian, though subject to a Rao damage process with parameter p having a beta distribution; the number of surviving larvae per egg mass is also assumed to be Poissonian. The symbolic representation for this model is

$$X \sim \left[\left\{\text{Poisson}(\lambda) \bigvee \text{Binomial}(1, P)\right\} \bigvee \text{Poisson}(\phi)\right] \bigwedge_P \text{Beta}(a, b) \tag{9.171}$$

$$\sim \left[\left\{\text{Binomial}(N, P) \bigvee \text{Poisson}(\phi)\right\} \bigwedge_N \text{Poisson}(\lambda)\right] \bigwedge_P \text{Beta}(a, b). \tag{9.172}$$

This is *not* the same model as (9.162) or (9.163). The corresponding pgf is

$$
\begin{aligned}
G(z) &= \int_0^1 \exp\{\lambda[(1-p+pe^{\phi(z-1)})-1]\}\frac{p^{a-1}(1-p)^{b-1}\,dp}{B(a,b)} \\
&= {}_1F_1[a; a+b; \lambda(e^{\phi(z-1)}-1)]. \qquad (9.173)
\end{aligned}
$$

Subrahmaniam (1966) showed how to obtain the probabilities recursively and found that

$$
\mu = \frac{\lambda\phi a}{a+b}, \qquad \mu_2 = \frac{\lambda\phi a}{a+b}\left[1+\lambda+\frac{\lambda\phi b}{(a+b)(a+b+1)}\right]. \qquad (9.174)
$$

He commented that $a = 1$, $b = 0$ give a Neyman type A distribution and that the limiting forms of (9.164) are nearly always the same as those of (9.173). There is one exception, however. When $\beta \to \infty$ for fixed α, μ, and μ_2, Gurland's family tends to a generalized Pólya–Aeppli (Poisson $\bigvee$ Pascal) distribution. Contrariwise, when $b \to \infty$ for fixed a, μ, and μ_2, the limiting form for the Subrahmaniam family has the pgf

$$
\begin{aligned}
G(z) &= \lim_{b\to\infty} {}_1F_1\left[a; a+b; \frac{(a+b)\mu}{a\phi}(e^{\phi(z-1)}-1)\right] \\
&= {}_1F_0\left[a; -; \frac{\mu}{a\phi}(e^{\phi(z-1)}-1)\right] = \left(1+\frac{\mu}{a\phi}-\frac{\mu}{a\phi}e^{\phi(z-1)}\right)^{-a}, \qquad (9.175)
\end{aligned}
$$

that is, it has a Negative binomial $\bigvee$ Poisson ($\equiv$ Pascal–Poisson) distribution; further details concerning this distribution are in Section 5.12.6.

A different extension of the type A distribution was made by Cresswell and Froggatt (1963). They supposed, in addition to their assumptions in Section 9.6.5 for the type A distribution, that accidents can also occur outside a spell according to a Poisson distribution with parameter ρ. This yields a convolution of a Neyman type A with a Poisson distribution that can be represented symbolically as

$$
X \sim \left[\text{Poisson}(\lambda) \bigvee \text{Poisson}(\phi)\right] * \text{Poisson}(\rho).
$$

They used the term "Short" distribution to describe this distribution because the tail length is reduced compared with that for the type A component distribution. The pgf is

$$
G(z) = \exp[\lambda(e^{\phi(z-1)}-1)+\rho(z-1)], \qquad (9.176)
$$

with $\lambda, \phi, \rho > 0$. These authors fitted the distribution by the method of moments to many data sets on accidents to public transport bus drivers. Froggatt (1966) has also examined its usefulness for fitting data on industrial absenteeism.

The probabilities can be obtained by convoluting the type A probabilities and the probabilities for a Poisson distribution with parameter (ρ); this gives

$$\Pr[X = 0] = \exp[\lambda(e^{-\phi} - 1) - \rho], \tag{9.177}$$

$$\Pr[X = x] = \frac{\exp[-(\lambda + \phi)]}{x!} \sum_{j=0}^{\infty} (j\phi + \rho)^x \frac{(\lambda e^{-\phi})^j}{j!} \tag{9.178}$$

$$= \exp[\lambda(e^{-\phi} - 1) - \rho] \sum_{j=0}^{x} \sum_{k=0}^{j} S(j, k) \frac{\rho^{x-j}\phi^j}{(x-j)!j!} (\lambda e^{-\phi})^k, \tag{9.179}$$

$x = 1, 2, \ldots$. The polynomial form (9.178) is advantageous for computing the probabilities.

Kemp (1967b) derived the following useful recurrence relationship:

$$\Pr[X = x + 1] = \frac{\rho}{x+1} \Pr[X = x] + \frac{\lambda\phi e^{-\phi}}{x+1} \sum_{j=0}^{x} \frac{\phi^j}{j!} \Pr[X = x - j]. \tag{9.180}$$

Cresswell and Froggatt (1963) derived formulas for the moments from expressions for the cumulants that involve Stirling numbers. Kemp (1967b) showed that they can be obtained more simply from the factorial cumulants, $\kappa_{[1]} = \rho + \lambda\phi$, $\kappa_{[r]} = \lambda\phi^r$, $r \geq 2$. We find

$$\begin{aligned} \mu &= \rho + \lambda\phi, \\ \mu_2 &= \rho + \lambda\phi(1 + \phi), \\ \mu_3 &= \rho + \lambda\phi(1 + 3\phi + \phi^2). \end{aligned} \tag{9.181}$$

Kemp also studied maximum-likelihood estimation of the (three) parameters using Newton–Raphson iteration and showed how to obtain their variances and covariances; he found that the relative efficiency of estimation by the method of moments is generally poor, particularly for $\phi \geq 1$.

The Poisson $\bigvee$ Katz family is yet another generalization of the Neyman type A distribution; see Bhalerao and Gurland (1977) and Tripathi, Gurland, and Bhalerao (1986). Its pgf is

$$G(z) = \exp\left\{\lambda\left[\left(1 - \frac{\beta}{1-\beta}(z-1)\right)^{-\alpha/\beta} - 1\right]\right\}, \tag{9.182}$$

$\lambda > 0$, $\alpha > 0$, $\beta < 1$. It was so named because it can be obtained via a model where the number of clusters has a Poisson distribution and the number of entities within a cluster has a distribution belonging to the Katz family of distributions (see Section 2.3.1 for the Katz family). The Katz family includes the binomial, Poisson, and negative binomial distributions, and hence the Poisson $\bigvee$ Katz family includes the Hermite ($-\alpha/\beta = 2$), Poisson–binomial ($-\alpha/\beta > 0$, an integer), Neyman type A ($\beta \to 0$), Pólya–Aeppli ($\alpha/\beta = 1$), and Poisson–Pascal ($0 < \beta < 1$). The binomial, Poisson, and negative binomial distributions are obtainable as limiting forms.

The probabilities for this family are given by

$$\Pr[X = 0] = \exp\{\lambda[(1-\beta)^{\alpha/\beta} - 1]\},$$

$$\Pr[X = x+1] = \frac{\lambda}{x+1} \sum_{j=0}^{x} \frac{(\alpha/\beta + j)!\beta^{j+1}(1-\beta)^{\alpha/\beta}}{(\alpha/\beta - 1)!j!} \Pr[X = x - j], \tag{9.183}$$

$x = 0, 1, 2, \ldots$. The moments can be obtained from the factorial cumulants,

$$\kappa_{[r]} = \lambda \frac{(\alpha/\beta + r - 1)!}{(\alpha/\beta - 1)!} \left(\frac{\beta}{1-\beta} \right)^r; \tag{9.184}$$

in particular

$$\mu = \frac{\lambda\alpha}{1-\beta} \qquad \text{and} \qquad \mu_2 = \frac{\lambda\alpha(\alpha+\beta)}{(1-\beta)^2}. \tag{9.185}$$

Tripathi, Gurland, and Bhalerao developed a generalized minimum χ^2 method for the estimation of the (three) parameters; they noted, in passing, that maximum-likelihood estimation could also be used.

The rationale underlying the study of the Poisson $\bigvee$ Katz family is the desire to provide a method for selecting and fitting one out of a number of possibly suitable multiple Poisson distributions by applying an overall estimation procedure to the data and then observing the resulting values of the parameter estimates (particularly their signs).

9.10 THOMAS DISTRIBUTION

This is similar to the Neyman type A distribution in Section 9.6, except that the generalizing Poisson distribution is replaced by the distribution of a Poisson variable increased by unity. Symbolically the distribution can be represented by

$$X \sim \text{Poisson}(\lambda) \bigvee \text{Shifted Poisson}(\phi), \tag{9.186}$$

where the shifted Poisson has support $1, 2, \ldots$.

This distribution was used by Thomas (1949) in constructing a model for the distribution of plants of a given species in randomly placed quadrats. It is well suited to situations in which *the parent as well as the offspring* is included in the count for each cluster arising from an initial individual, assuming that the initial distribution is Poissonian with parameter λ and that the number of offspring per initial individual is also Poissonian but with parameter ϕ. The pgf is therefore

$$G(z) = \exp[\lambda(ze^{\phi(z-1)} - 1)]. \tag{9.187}$$

Thomas called this distribution a "double-Poisson" distribution, though Douglas (1980) has pointed out that the term applies more appropriately to a Neyman type A distribution than to a Thomas distribution. [The name double Poisson is also used for the bivariate Poisson distribution and for the double-Poisson distribution belonging to Efron's (1986) double-exponential family.]

It is easy to verify by differentiation of the pgf that

$$\begin{aligned}
\Pr[X = 0] &= e^{-\lambda}, \\
\Pr[X = 1] &= \lambda e^{-(\lambda+\phi)}, \\
\Pr[X = 2] &= \tfrac{1}{2}\lambda e^{-(\lambda+\phi)}(\phi + \lambda e^{-\phi}), \\
&\;\;\vdots
\end{aligned} \tag{9.188}$$

Thomas (1949) obtained the following general expression for the probabilities:

$$\Pr[X = x] = \frac{e^{-\lambda}}{x!} \sum_{j=1}^{x} \binom{x}{j} (\lambda e^{-\phi})^j (j\phi)^{x-j}. \tag{9.189}$$

Like the Neyman type A, the distribution can have more than one mode; this can be seen by considering $\phi > \max(2, \ln \lambda)$, in which case $\Pr[X = 0] > \Pr[X = 1] < \Pr[X = 2]$.

The fcgf is

$$\ln G(1 + t = \lambda(1 + t)e^{\phi t} - \lambda, \tag{9.190}$$

whence

$$\kappa_{[r]} = \lambda \phi^{r-1}(r + \phi) \tag{9.191}$$

(Ord, 1972). The moments can also be obtained directly from the central mgf

$$e^{-\mu} G(e^t) = \exp\{\lambda[\exp(t + \phi e^t - \phi) - (1 + t + \phi t)]\}. \tag{9.192}$$

The distribution is positively skewed and the index of dispersion is

$$1 + \frac{\phi(2 + \phi)}{1 + \phi}.$$

Thomas applied the distribution to observed distributions of plants per quadrat and obtained marked improvements over fits with Poisson distributions. In particular, Thomas distributions can be bimodal, and bimodality was a feature in both of the fitted distributions. The data were fitted by the method of moments;

this was discussed in detail in Gleeson and Douglas (1975). They were also fitted by setting the observed proportions of zero and unit values equal to their expectations. The latter method gives

$$\lambda^* = -\ln\left(\frac{f_0}{N}\right), \qquad \phi^* = \ln\left(\frac{f_0\lambda^*}{f_1}\right). \tag{9.193}$$

The use of only the observed proportions of zero and unit counts is advantageous when an exhaustive count of actual numbers per quadrat for $x \geq 2$ is not required and would be tedious to obtain. Methods of solution of the maximum-likelihood equations for a complete sample of data from a Thomas distribution were investigated by Russell (1978).

9.11 BOREL–TANNER DISTRIBUTION: LAGRANGIAN POISSON DISTRIBUTION

The *Borel–Tanner distribution* (Tanner–Borel distribution) describes the distribution of the total number of customers served before a queue vanishes given a single queue with random arrival times of customers (at constant rate ℓ) and a constant time β occupied in serving each customer. We suppose that the probability of arrival of a customer during the period $(t, t+\Delta t)$ is $\ell\Delta t + o(\Delta t)$ and that the probability of arrival of two or more customers in this period is $o(\Delta t)$. If there are initially n customers in the queue, then the probability that the total number (Y) of customers served before the queue vanishes is equal to y is

$$\Pr[Y = y] = \frac{n}{(y-n)!} y^{y-n-1}(\ell\beta)^{y-n} e^{-\ell\beta y}, \qquad y = n, n+1, \ldots. \tag{9.194}$$

The distribution was obtained by Borel (1942) for the case $n = 1$ and for general values of n by Tanner (1953). The parameters ℓ and β appear only in the form of their product $\ell\beta$. It is convenient to use a single symbol for this product and to put $\ell\beta = a$, say. For (9.194) to represent a proper distribution, it is necessary to have $0 < a < 1$. If $a < 0$, the "probabilities" change sign, while if $a > 1$, $\sum_{x=n}^{\infty} \Pr[Y = y] < 1$.

Let $a(b)$ be the solution of the equation $b = ae^{-a}$. Using this inverse function, Haight and Breuer (1960) were able to show that the pgf can be expressed as

$$H(z) = \left[\frac{a(bz)}{a(b)}\right]^n = z^n e^{na(bz)-na(b)} = z^n e^{na\{[H(z)]^{1/n}-1\}}. \tag{9.195}$$

Clearly $H(z)/z^n$ is a Poisson–stopped sum (multiple Poisson) distribution. The pgf of the generalizing distribution is $[H(z)]^{1/n}$, that is, the generalizing distribution is a Borel distribution.

Haight and Breuer found from (9.195) that

$$H'(1) = n + aH'(1)$$

$$H''(1) = n(n-1) + naH'(1) + \frac{a[H'(1)]^2}{n} + aH''(1),$$

and hence

$$\mu = \frac{n}{1-a}, \qquad \mu_2 = \frac{na}{(1-a)^3}\,. \tag{9.196}$$

They remarked that the moment properties can also be obtained by successive differentiation of

$$K(t) = \ln G(e^t) = n(t-a) + na[G(e^t)]^{1/n}, \tag{9.197}$$

where $K(t)$ is the cgf and $G(e^t)$ is the mgf.

The distribution is a power series distribution and hence the maximum-likelihood equation for a, assuming n is known, is the first-moment equation, giving

$$\hat{a} = \frac{\overline{x} - n}{\overline{x}}. \tag{9.198}$$

Consul's (1989) Lagrangian Poisson distribution is obtained by shifting the Tanner–Borel distribution so that it has support $0, 1, 2, \ldots$, that is, by transforming to the rv $X = Y - n$. The usual notation sets $\theta = an$ and $\lambda = a$. The pmf becomes

$$\Pr[X = x] = \frac{\theta(\theta + x\lambda)^{x-1}}{x!}, \qquad x = 0, 1, \ldots, \tag{9.199}$$

and the pgf has the form

$$G(z) = e^{\theta(t-1)} = \exp[\theta(ze^{\lambda(t-1)} - 1)] \tag{9.200}$$

$$= \exp\{\theta z[G(z)]^{\lambda/\theta} - \theta\}, \tag{9.201}$$

where $t = ze^{\lambda(t-1)}$. From (9.200) the distribution is clearly seen to be a Poisson–stopped sum (multiple Poisson) distribution.

It has been studied in considerable detail by Consul and his co-workers. For the very extensive literature on the distribution, see Consul (1989). It is a member of the Lagrangian family of distributions and is discussed in greater depth in Section 7.2.

The representation (9.201) is recursive. The *Skellam–Haldane distribution* is another Poisson–stopped sum distribution with a pgf that is recursive. Here the pgf is

$$G(z) = e^{\lambda(z-1)}G(e^{c(z-1)}) \tag{9.202}$$

$$= \exp\{\lambda[(z-1) + (e^{c(z-1)} - 1) + (e^{c(e^{c(z-1)}-1)} - 1) + \cdots]\}. \tag{9.203}$$

The distribution is discussed in depth in Section 11.2.16.

9.12 OTHER POISSON–STOPPED SUM (MULTIPLE POISSON) DISTRIBUTIONS

Multiple Poisson distributions are convolutions of distributions of Poisson singlets, Poisson doublets, Poisson triplets, and so on; see (9.5.2). The convolution of two or more multiple Poisson distributions is therefore also a multiple Poisson distribution (Samaniego, 1976).

In their comprehensive survey paper Wimmer and Altmann (1996) list very many distributions of this kind. They include convolutions of distributions with geometric and negative binomial distributions (see Section 5.12.5) such as the Delaporte distribution. This is a negative binomial $*$ Poisson convolution; see Delaporte (1959) and Willmot (1989b). Another example is the noncentral negative binomial distribution of Gurland, Chen, and Hernandez (1983), Ong and Lee (1992), and others. This is a convolution of a negative binomial and a Pólya–Aeppli distribution.

Consider now the convolution of two distributions with pgf's $G_1(z)$ and $G_2(z)$, where $G_1(z)$ is multiple Poisson but $G_2(z)$ is not. Then $G_1(z) = \exp(\sum_{i\geq 1} a_i(z^i - 1))$, where $a_i \geq 0$ for all i. Suppose that $G_2(z)$ can be restated as $G_2(z) = \exp(\sum_{i\geq 1} b_i(z^i - 1))$; because it is not multiple Poisson, some of the b_i are negative. However, $G_1(z) * G_2(z)$ is the pgf of a multiple Poisson distribution provided that $a_i + b_i \geq 0$ for all i. An example is the convolution of a geometric and a Bernoulli distribution (Hunter, 1983a, 1983b). The pgf is

$$G_1(z) * G_2(z) = \frac{(1-a)(1+bz)}{(1-az)(1+b)} \tag{9.204}$$

$$= \left(\frac{1-a}{1+b}\right)\exp\left[\sum_{i\geq 1}\frac{(az)^i}{i} - \sum_{i\geq 1}\frac{(-bz)^i}{i}\right], \quad 0 < a < 1, \quad 0 < b. \tag{9.205}$$

The convolution is therefore a multiple Poisson distribution provided that $b \leq a$.

It is not even necessary that $G_2(z)$ is the pgf of an honest distribution. Suppose that $c = -b$ in (9.204), giving (Rubinovitch, 1985)

$$G_1(z) * G_2(z) = \frac{(1-a)(1-cz)}{(1-az)(1-c)} \tag{9.206}$$

$$= \left(\frac{1-a}{1-c}\right)\exp\left[\sum_{i\geq 1}\frac{(az)^i}{i} - \sum_{i\geq 1}\frac{(cz)^i}{i}\right], \; 0 < a < 1, \; 0 < c < 1, \tag{9.207}$$

which is also multiple Poisson provided that $c \leq a$.

The pgf (9.206) can be restated as

$$H(z) = \frac{(1-a)(1-cz)}{(1-az)(1-c)} = \frac{(1-a)c}{(1-c)a} + \frac{(a-c)}{(1-c)a}\cdot\frac{(1-a)}{(1-az)},$$

showing that a zero-inflated geometric distribution is a multiple Poisson when $0 < c < a < 1$. These distributions are all Galton–Watson distributions; see Section 11.2.7 and Kemp (1979).

Consider again the Delaporte distribution. The pgf is

$$G(z) = \left(\frac{1-a}{1-az}\right)^k e^{\lambda(z-1)} \tag{9.208}$$

$$= (1-a)^k e^{-c} \exp\left[\lambda z + \sum_{i \geq 1} \frac{(az)^i}{i}\right], \qquad 0 < a < 1, \qquad 0 < \lambda. \tag{9.209}$$

This is the pgf of a multiple Poisson distribution. When the parameter space is enlarged to $0 < a < 1$, $0 < k$, $-ak < \lambda < 0$, it remains an honest multiple Poisson distribution and would be suitable for graduating data, although the "Poisson" parameter λ has lost its physical meaning.

A great number of distributions can be constructed as indicated in this section. Nearly all of the many multiple Poisson distributions in Wimmer and Altmann's (1996) list have arisen in modeling situations.

9.13 OTHER FAMILIES OF STOPPED-SUM DISTRIBUTIONS

Sections 9.3–9.11 have concentrated on Poisson–stopped sum distributions. These are the type A distributions of Khatri and Patel (1961). The type B and type C families in Khatri and Patel's somewhat neglected paper have pgf's of the form

$$\text{Type B:} \qquad G_B(z) = [h(z)]^n, \tag{9.210}$$

$$\text{Type C:} \qquad G_C(z) = c\ln[h(z)], \tag{9.211}$$

where $h(z) = a + bg(z)$ and $g(z)$ is a pgf. These types A, B, and C distributions are not the same as the Neyman types A, B, and C distributions.

Let $f^{(r)}(z)$ denote the rth derivative of $f(z)$. Then

$$G(z) = \sum_x \Pr[X = x]z^x = \sum_x \left[\frac{G^{(x)}(z)}{x!}\right]_{z=0} z^x, \tag{9.212}$$

$$g(z) = \sum_x \pi_x z^x = \sum_x \left[\frac{g^{(x)}(z)}{x!}\right]_{z=0} z^x. \tag{9.213}$$

For the type B family $a = 1 - b$; successive differentiation of

$$[a + bg(z)]G_B^{(1)}(z) = nbg^{(1)}(z)G_B(z)$$

yields

$$aG^{(x)}(z) + b\sum_{j=1}^{x}\binom{x-1}{j-1}g(z)^{(j-1)}(z)G_B^{(x-j+1)}(z)$$
$$= nb\sum_{j=1}^{x}\binom{x-1}{j-1}g^{(j)}(z)G_B^{(x-j)}(z), \qquad (9.214)$$

and setting $z = 0$ gives

$$\Pr[X = x] = \sum_{j=1}^{x}\frac{(nj + j - x)\pi_j}{x(\pi_0 + a/b)}\Pr[X = x - j] \qquad (9.215)$$

with

$$\Pr[X = 0] = (a + b\pi_0)^n. \qquad (9.216)$$

For the type C family

$$G_C(z) = c\ln[a + bg(z)], \qquad (9.217)$$

where $c = [\ln(a + b)]^{-1}$. Successive differentiation of

$$[a + bg(z)]G_C^{(1)}(z) = bcg^{(1)}(z)$$

gives

$$[a + bg(z)]G_C^{(x)}(z) = bcg^{(x)}(z) - b\sum_{j=1}^{x-1}\binom{x-1}{j}g^{(j)}(z)G_C^{(x-j)}(z), \qquad (9.218)$$

whence setting $z = 0$ yields

$$\Pr[X = 0] = c\ln(a + b\pi_0),$$

$$\Pr[X = x] = (a + b\pi_0)^{-1}\left(bc\pi_x - \sum_{j=1}^{x-1}\frac{b(x-j)}{x}\pi_j\Pr[X = x - j]\right), \qquad x \geq 1; \qquad (9.219)$$

[Note the slight alterations to the formulas in Khatri and Patel (1961).] When $a = 1$ and $-1 < b < 0$ the type C family comprises logarithmic–stopped sum distributions.

Khatri and Patel developed not only these probability properties but also moment properties. Let

$$\mu'_{[r]} = [h^{(r)}(z)]_{z=1}, \qquad M'_{[r]} = [G^{(r)}(z)]_{z=1}, \qquad K_{[r]} = \left[\frac{d^r}{dz^r}\ln G(z)\right]_{z=1}.$$

Then $\mu'_{[r]}$ is the rth factorial moment for $h(z)$ if $h(z)$ is a pgf. Also $M'_{[r]}$ and $K_{[r]}$ are, respectively, the rth factorial moment and rth factorial cumulant for $G(z)$.

For the type B family $\ln G_B(z) = n \ln h(z)$, so the rth factorial cumulant of $G_B(z)$ is n times the rth factorial cumulant for $h(z)$ if $h(z)$ is a pgf. Also

$$K_{[r]} = n\mu'_{[r]} - \sum_{j=1}^{r-1} \binom{r-1}{j} \mu'_{[j]} K_{[r-j]}. \tag{9.220}$$

For the type C family Khatri and Patel found that

$$M'_{[r]} = \frac{c\mu'_{[r]} - \sum_{j=1}^{r-1} \binom{r-1}{j} \mu'_{[j]} M'_{[r-j]}}{\mu'_0}, \quad \text{where } \mu'_0 = h(1). \tag{9.221}$$

For Khatri and Patel's type B family the pgf must have the form

$$G_B(z) = [1 - p + pg(z)]^n;$$

consequently, when n is a positive integer and $0 < p < 1$, this family consists of generalized binomial distributions. Furthermore, for n a positive integer and $0 < p < [1 - g(0)]^{-1}$, it can be interpreted as an n-fold convolution of zero-modified distributions. When n and p are both negative (n not necessarily an integer), the family consists of generalized negative binomial distributions.

Khatri and Patel (1961) gave the recurrence formulas for the probabilities for binomial and negative binomial distributions with the following generalizing distributions: hypergeometric, binomial, negative binomial, and Poisson. Special attention was paid to the Binomial $\bigvee$ Negative binomial distribution; Khatri and Patel developed a model for it based on the distribution of organisms that occur in colonies; they studied maximum-likelihood and other forms of estimation, and they fitted the distribution to insect data; see also Khatri (1962).

The Negative binomial $\bigvee$ Hypergeometric distribution appears in Gurland's (1958) paper. The Negative binomial $\bigvee$ Poisson has been investigated in depth by Subrahmaniam (1966) and Subrahmaniam (1978); see Sections 5.12.6 and 9.9.

The modeling of insurance claims is studied in Bowers et al. (1986). Verrall (1989) has taken their ideas further by putting forward a binomial–stopped sum model for individual risk claims. He showed that, if the rv X has the pgf

$$G(z) = [1 - q + qg(z)]^n,$$

where the mean and variance for $g(z)$ are μ and σ^2, then the mean and variance of X are

$$E[X] = nq\mu \qquad \text{and} \qquad \text{Var}(X) = n[q\sigma^2 + q(1-q)\mu^2]. \tag{9.222}$$

Given heterogeneous claims, the pgf becomes

$$G(z) = \prod_{k=1}^{K} [1 - q_k + q_k g_k(z)]^{n_k}$$

and the corresponding mean and variance are

$$E[X] = \sum_{k=1}^{K} n_k q_k \mu_k, \qquad \text{Var}(X) = \sum_{k=1}^{K} n_k [q_k \sigma_k^2 + q_k(1 - q_k)\mu_k^2], \tag{9.223}$$

where μ_k and σ_k^2 are the mean and variance of the kth individual claim distribution. Verrall also studied Poisson–stopped sum approximations to his binomial–stopped sum models.

The aim of Katti's (1966) paper was to examine interrelations between generalized distributions with pgf's of the more general form $G(z) = g_1(g_2(z))$ and their component distributions $g_1(z)$ and $g_2(z)$. He showed that

$$\begin{aligned}
\mu'_{[1]} &= {}_1\mu'_{[1]}\,{}_2\mu'_{[1]}, \\
\mu'_{[2]} &= {}_1\mu'_{[2]}({}_2\mu'_{[1]})^2 + {}_1\mu'_{[1]}\,{}_2\mu'_{[2]}, \\
\mu'_{[3]} &= {}_1\mu'_{[3]}({}_2\mu'_{[1]})^3 + 3\,{}_1\mu'_{[2]}\,{}_2\mu'_{[2]}\,{}_2\mu'_{[1]} + {}_1\mu'_{[1]}\,{}_2\mu'_{[3]}, \\
&\ \ \vdots
\end{aligned} \tag{9.224}$$

and commented on the similarity between these formulas and those for moments in terms of cumulants. He also showed how to obtain corresponding formulas for the factorial cumulants for $G(z)$ in terms of the factorial cumulants for $g_1(z)$ and $g_2(z)$. His formulas simplify not only for Poisson–stopped sum distributions but also when $g_2(z) = \exp[\phi(z - 1)]$.

In addition Katti gave a table containing skewness and kurtosis comparisons for some 21 stopped-sum distributions and appended detailed comments. In making these comparisons, he adopted Anscombe's (1950) method of fixing the first two factorial cumulants as kp and kp^2 and then evaluating $\kappa_{[3]}/kp^3$ and $\kappa_{[4]}/kp^4$.

CHAPTER 10

Matching, Occupancy, Runs, and q-Series Distributions

10.1 INTRODUCTION

Problems concerning *coincidences* (i.e., *matching problems*) arise when sequences of characteristics are compared, for example when two sequences of n items are compared a pair at a time. The occurrence of the same characteristic at the jth comparison is called a *match*.

Occupancy distributions relate to the number of occupied categories when r objects are assigned in a random manner to n categories. Many variants of the classical occupancy problem (involving restraints on the method of assignment) have been studied, some for their practical import and others for their theoretical interest.

A *run* is the occurrence of an uninterrupted sequence of a particular attribute in an observed series of attributes. In a series of variate values, an uninterrupted sequence that is monotonically nondecreasing or nonincreasing is termed a run *up* or *down*, respectively. Results concerning runs are important in the theory of distribution free tests (see, e.g., Weiss, 1988).

The theory of occupancy and matching distributions developed in the context of gambling problems. Twentieth-century uses include the following: Maxwell–Boltzmann, Bose–Einstein, and Fermi–Dirac systems in statistical physics (e.g., Desloge, 1966; Feller, 1957); personality assessment in psychology (Vernon, 1936); applications in genetics (Stevens, 1937, 1939), estimation of insect populations (Craig, 1953); computer storage analysis (Meilijson, Newborn, Tenenbein, and Yechieli, 1982); and matches between two DNA sequences (Goldstein, 1990).

Matching, occupancy, and runs problems are of special interest in discrete distribution theory for the impetus that they have given to methodology. The solution of such problems often depends on the inclusion–exclusion principle.

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

Because of its importance in this area of discrete distribution theory, the next section is devoted to the principle and to associated theorems for the probabilities of combined events.

10.2 PROBABILITIES OF COMBINED EVENTS

In the (1718) edition of the *Doctrine of Chances*, De Moivre was able to generalize results that had previously been used by De Montmort in the (1713) edition of his *Essai d'Analyse sur les Jeux de Hasard.*

Consider k events $E_1, E_2, \ldots, E_k$ and suppose that the probabilities of the simultaneous occurrence of any number of them are known. Suppose also, like De Moivre, that the events are exchangeable, that is, that their probabilities satisfy

$$P[E_{i_1}E_{i_2}\cdots E_{i_j}] = P[E_1E_2\cdots E_j], \tag{10.1}$$

where $\{i_1, i_2, \ldots, i_j\}$ is any reordering of $\{1, 2, \ldots, j\}$. Also let

$$S_j = \sum P[E_{i_1}E_{i_2}\cdots E_{i_j}], \tag{10.2}$$

where summation is over $\{i_1, i_2, \ldots, i_j\}$ such that $1 \le i_1 < i_2 < \cdots < i_j \le k$, and $j = 1, 2, \ldots, k$; also let $S_0 = 1$.

Then, in modern notation rather than De Moivre's, the probability that at least m of the k events occur is

$$\begin{aligned} P_m &= S_m - \binom{m}{1}S_{m+1} + \binom{m+1}{2}S_{m+2} - \cdots + (-1)^{k-m}\binom{k-1}{k-m}S_k \\ &= \sum_{i=0}^{k-m}(-1)^i\binom{m+i-1}{i}S_{m+i} \end{aligned} \tag{10.3}$$

(with $P_0 = 1$), and the probability that exactly m of the k events occur is

$$\begin{aligned} P_{[m]} &= S_m - \binom{m+1}{1}S_{m+1} + \binom{m+2}{2}S_{m+2} - \cdots + (-1)^{k-m}\binom{k}{k-m}S_k \\ &= \sum_{i=0}^{k-m}(-1)^i\binom{m+i}{i}S_{m+i}. \end{aligned} \tag{10.4}$$

Clearly

$$P_m = P_{[m]} + P_{[m+1]} + \cdots + P_{[k]} \qquad \text{and} \qquad P_{[m]} = P_m - P_{m+1}. \tag{10.5}$$

The method of proof used by De Moivre involves the repeated use of

$$P[\overline{E}_1 \cap E_2] = P[E_2] - P[E_1 \cap E_2];$$

see Hald's (1990) lucid historical discussion.

Jordan (1867) proved that these formulas are also valid when the events are not exchangeable (note the definition of S_j given above). This more general case can be proved by a method such as De Moivre's. It can also be proved by the use of the inclusion–exclusion principle. Given N objects, suppose that $N(a)$ have property a, $N(b)$ have property $b, \ldots, N(ab)$ have both a and $b, \ldots, N(abc)$ have a, b, and c, and so on. Then the inclusion–exclusion principle states that the number of objects with none of these properties is

$$\begin{aligned} N(\overline{a}\overline{b}\overline{c}\ldots) = N &- N(a) - N(b) - \cdots \\ &+ N(ab) + N(ac) + \cdots \\ &- N(abc) - \cdots \\ &+ \cdots . \end{aligned} \tag{10.6}$$

The principle is fundamental to Boole's (1854) algebra of classes and is used extensively in McMahon's (1915, 1916) *Combinatory Analysis*.

The formulas can also be proved by Loève's (1963) indicator function method. Let X_i be an indicator variable that is equal to 1 when A_i occurs and zero otherwise. Then, for example,

$$1 - (1 - X_1)(1 - X_2) \cdots (1 - X_k)$$

is the indicator variable for the event that at least one of the A_i occur, and the probability for this event is

$$\begin{aligned} P_1 &= E[1 - (1 - X_1)(1 - X_2) \cdots (1 - X_k)] \\ &= E\left[\sum_i X_i - \sum_{i \neq j} X_i X_j + \cdots + (-1)^{k+1} X_1 X_2 \cdots X_k\right] \\ &= S_1 - S_2 + \cdots + (-1)^{k+1} S_k. \end{aligned} \tag{10.7}$$

The probabilities P_m, $P_{[m]}$, $m = 2, \ldots, k$, can be obtained similarly [see, e.g., Parzen (1960) or Moran (1968)]. Besides a very clear account, Moran gives references to further work by Geiringer (1938) and to a long series of papers from 1941 to 1945 by K. L. Chung.

The formulas for P_m and $P_{[m]}$ have been used extensively in life insurance mathematics ever since De Moivre's (1725) work on the subject. In King's (1902) comprehensive text on life insurance mathematics, the author developed the symbolic representation

$$P_m \equiv S^m (1 + S)^{-m}, \qquad P_{[m]} \equiv S^m (1 + S)^{-m-1}, \tag{10.8}$$

where, after expansion, S^j is replaced by S_j and $S_j = 0$ for $j > m$. This is the method adopted by Jordan (1972) and Riordan (1958). Takács (1967) generalized the formulas. Broderick (1937) and Fréchet (1940, 1943) extended them to situations where the events are not independent.

Consider now the probability generating function (pgf)

$$G(z) = \sum_{m=0}^{k} P_{[m]} z^m = \sum_{m=0}^{k} S_m (z-1)^m \tag{10.9}$$

for the rv X taking the values $0, 1, \ldots, k$ with probabilities $P_{[0]}, P_{[1]}, \ldots, P_{[k]}$. The corresponding factorial moment generating function (fmgf) is

$$G(1+t) = \sum_{m=0}^{k} P_{[m]} (1+t)^m = \sum_{m=0}^{k} S_m t^m,$$

and hence

$$S_r = \frac{\mu'_{[r]}}{r!}, \tag{10.10}$$

where $\mu'_{[r]}$ is the rth factorial moment. See Iyer (1949, 1958) for an interpretation of this result.

Reversing the series in the formulas gives

$$S_r = \sum_{j=r}^{k} \binom{j}{r} P_{[j]} = \sum_{j=r}^{k} \binom{j-1}{r-1} P_j. \tag{10.11}$$

A consequence of (10.3) and (10.4) is

$$S_m - (m+1) S_{m+1} \le P_{[m]} \le S_m \tag{10.12}$$

and

$$S_m - m S_{m+1} \le P_m \le S_m. \tag{10.13}$$

These are widely known as Bonferroni's inequalities; see Galambos (1975, 1977) for a review of methods of proof of such inequalities.

Fréchet's (1940, 1943) more general inequalities include

$$\begin{aligned}
P_m &\le S_m - \binom{m}{1} S_{m+1} + \cdots + \binom{m+\ell-1}{\ell} S_{m+\ell}, \\
P_m &\ge S_m - \binom{m}{1} S_{m+1} + \cdots - \binom{m+\ell}{\ell+1} S_{m+\ell+1}, \\
P_{[m]} &\le S_m - \binom{m+1}{1} S_{m+1} + \cdots + \binom{m+\ell}{\ell} S_{m+\ell}, \\
P_{[m]} &\ge S_m - \binom{m+1}{1} S_{m+1} + \cdots - \binom{m+\ell+1}{\ell+1} S_{m+\ell+1},
\end{aligned} \tag{10.14}$$

where ℓ is an even integer. Collectively such inequalities are termed Boole–Bonferroni–Fréchet inequalities, for example by David and Barton (1962). An improved Bonferroni upper bound has been obtained by Worsley (1982); see also Hunter (1976), Seneta (1988), and Hoppe and Seneta (1990).

In Whitworth's (1878, 1948) *Choice and Chance*, the author proves and uses extensively a result sometimes called *Whitworth's theorem*; it is a restatement of (10.4) with $m = 0$, giving

$$P_{[0]} = \sum_{i=0}^{k} (-1)^i S_i. \tag{10.15}$$

Irwin (1967) drew the attention of his contemporaries to Whitworth's theorem; previously, in Irwin (1955), he had used it to derive three interesting distributional results.

The history of the compound probability theorems (10.3) and (10.4) has been discussed by Takács (1967) and by Hald (1990) in his scholarly book on the history of probability and statistics. Fréchet's (1940, 1943) monograph (in French) gives not only his generalizations and applications but also delineates the works of certain earlier writers, some of which are not easy to obtain. Another useful and readily accessible reference is Feller (1957).

10.3 MATCHING DISTRIBUTIONS

The classical *problem of coincidences* (*problème de rencontre*) arises in the following way: Suppose that a set of k entities, numbered $1, 2, \ldots, k$, are arranged in a random order. Let X be the number of entities for which their position in the random order is the same as the number assigned to them. We seek to find the distribution of X.

For this problem $S_j = 1/j!$ [see, e.g., Alt (1985) for an explanation]. Applying this expression to (10.4) gives

$$\begin{aligned} \Pr[X = x] = P_{[x]} &= \frac{1}{x!} - \binom{x+1}{1}\frac{1}{(x+1)!} + \cdots + (-1)^{k-x}\binom{k}{k-x}\frac{1}{k!} \\ &= \frac{1}{x!}\sum_{i=0}^{k-x}\frac{(-1)^i}{i!}, \end{aligned} \tag{10.16}$$

where $x = 0, 1, \ldots, k$. Note that $\Pr[X = k - 1] = 0$.

From (10.16) it can be seen that, except when $k - x$ is small,

$$\Pr[X = x] \approx \frac{e^{-1}}{x!} \tag{10.17}$$

to a close degree of approximation. The successive probabilities can therefore be approximated by the corresponding probabilities for a Poisson distribution with parameter 1. Irwin (1955) showed numerically that for $k = 10$ there is at most a discrepancy of 1 in the fifth decimal place for the two distributions and stressed that for practical purposes the probabilities are the same for $k \geq 10$.

The close connection between the two distributions is brought out by considering their factorial moments (Olds, 1938). For the matching distribution the fmgf is

$$G(1+t) = \sum_{m=0}^{k} S_m t^m = \sum_{m=0}^{k} \frac{t^m}{m!}, \tag{10.18}$$

and so $\mu'_{[r]} = 1, r = 1, 2, \ldots, k$. For a Poisson distribution with parameter $\theta = 1$, we find that $\mu'_{[r]} = 1,\ r = 1, 2, \ldots$. The first k factorial moments, moments, cumulants, and so on, for the two distributions are therefore identical. In particular

$$\mu = \mu_2 = 1. \tag{10.19}$$

The papers by Chapman (1934) and Vernon (1936) dealt with the use of the distribution in psychology, for example, for "coincidences" between character sketches of people and samples of their handwriting. Vernon (1936) included an extensive bibliography. Kendall (1968) has drawn attention to the work of Young (1819) on coincidences in linguistics concerning the numbers of word root forms in common in various languages.

There are many variants of this simple classical model. Suppose, for example, that there are N objects. Let na of these be labeled $1, 2, \ldots, n$, each label occurring a times, and let $N - na$ be unlabeled. Assume also that the labels are assigned at random. We will say that a match has occurred if an object with label j occurs at the jth position (Haag, 1924). Fréchet (1940, 1943) called this the *Laplace–Haag matching problem* and showed that

$$S_r = \binom{n}{r} \frac{(N-r)!a^r}{N!}.$$

It follows that

$$\mu'_{[r]} = \frac{n!(N-r)!a^r}{(n-r)!N!} \tag{10.20}$$

and the pgf is

$$G(z) = {}_1F_1[-n; -N; a(z-1)] \tag{10.21}$$

(Kemp, 1978b). This is a generalized hypergeometric factorial distribution (GHFD) (see Section 2.4.2) with a pgf that strongly resembles the pgf for the Poisson–beta distribution.

The distribution exists for all positive real a, not just when a is a positive integer. Fréchet recognized that a can be less than unity and interpreted the case $0 < a < 1$ via a visibility bias model. The probabilities for this distribution are

$$\Pr[X = x] = \sum_{i=0}^{n-x} \frac{(-1)^i n!(N-x-i)!a^{x+i}}{x!i!(n-x-i)!N!}. \tag{10.22}$$

Kemp showed from the theory of GHFDs that the probabilities may more conveniently be calculated from

$$\begin{aligned}(x+2)(x+1)\Pr[X=x+2] &= (x+1)(x+a-N)\Pr[X=x+1]\\ &\quad + a(n-x)\Pr[X=x],\end{aligned} \tag{10.23}$$

where

$$\begin{aligned}\Pr[X=n] &= \frac{a^n(N-n)!}{N!},\\ \Pr[X=n-1] &= \frac{(1+N-n-a)na^{n-1}(N-n)!}{N!}.\end{aligned} \tag{10.24}$$

Kemp also showed that as both N and a become large the pgf tends to

$$G(z) = {}_1F_0\left[-n;\ ;(1-z)\left(\frac{a}{N}\right)\right], \tag{10.25}$$

that is, it becomes binomial. When both n and N become large, the limiting distribution is Poisson with parameter *an/N*. However, Barton (1958) has argued that the Poisson limit may not be a very good approximation for even quite large N.

Fréchet (1940, 1943) considered several special cases of this distribution. For the Laplace matching distribution $N = an$. For the Gumbel distribution $N = n$, $a \neq 1$ and for the classical matching problem $N = n$, $a = 1$. Fréchet provided appropriate references.

Many authors, in particular McMahon (1894, 1898, 1902, 1915), have examined the problem of coincidences in the context of matching packs of cards. McMahon's method of specifying the composition of a pack is as follows: Suppose that there are k sorts of cards in a pack and that there are a_i cards of the ith sort, $i = 1, 2, \ldots, k$. Then the composition of the pack is said to be $(a_1, a_2, \ldots, a_k)$; for brevity a set of am cards of m kinds each replicated a times is represented by (a^m). The composition of a standard pack of playing cards with one joker is thereby represented as $(1, 13^4)$. It is usual to speak of one pack as the target pack and the other as the matching pack. For example, for the Laplace–Haag problem above, the target pack is $(N-n, 1^n)$ and the matching pack is $(N-an, a^n)$; the $N-n$ component in the target pack and the $N-an$ component in the matching pack both consist of blank cards.

The equivalence of problems of coincidences and two-pack matching problems occurs because the order of the cards in the target pack is immaterial. What is important is the randomness of the matching pack.

Suppose, for example, that we have two identical packs of cards and that each pack of cards has n suits of c (distinct) cards per suit (so that each pack contains $N = cn$ cards). The two packs are then dealt out simultaneously, in random order, and at the same rate. Let a match be said to occur if the cards dealt from the two packs at the same time have the same face value. (The case $n = 1$ is a special case of the problem of coincidences.) For general values of n

$$\Pr[X = x] = \sum_{i=x}^{N} (-1)^{i-x} \binom{i}{x} \frac{(N-i)!H_i}{N!}, \tag{10.26}$$

where H_i is the coefficient of u^i in the expansion of

$$\left[\sum_{j=0}^{c} \frac{(c!)^2 u^j}{j!(c-j)!(c-j)!}\right]^n.$$

Tables of $\Pr[X \geq x]$ to five decimal places for

$$\begin{array}{lll} c = 1, 2 \text{ and } n = 2(1)11 & \text{also} & n = 2 \text{ and } c = 6(1)12 \\ c = 3 \quad \text{ and } n = 2(1)8 & & n = 3 \text{ and } c = 6, 7 \\ c = 4, 5 \text{ and } n = 2(1)5 & & \end{array}$$

are given in Gilbert (1956) and are also included in Owen (1962). Silva (1941) has shown that there are the following approximations for $\Pr[X = x]$ when a is large:

$$\Pr[X = x] \approx \frac{n^x e^{-n}}{x!} \left[1 - \left(\frac{n-1}{2na}\right)\left(\frac{(n-x)^2 - x}{c}\right)\right]$$

and

$$\Pr[X = x] \approx \frac{n^x}{x!} \left(1 - \frac{1}{a}\right)^{na} \left[1 + \frac{1}{2a} + \frac{(n-1)(2n+1-x)x}{2n^2 a}\right].$$

Greville (1941) gave an appropriate expression for the more general problem in which there are $n_{11}, \ldots, n_{1a}$ cards per face value in the first pack and $n_{21}, \ldots, n_{2a}$ cards per face value in the second pack. Joseph and Bizley (1960) gave formulas from which values of $\Pr[X = x]$ can be calculated for the case $n_{1i} = n_{2i}$ for all i.

McMahon (1915) solved the problem of two packs with completely general composition; however, his formulas are not very tractable. A somewhat simpler operator technique was devised by Kaplansky and Riordan (1945). Barton (1958) discussed their work and also that of Stevens (1939); a particular strength of Barton's paper is the derivation of Poisson limiting forms.

Yet more matching problems arise when there is simultaneous matching among K packs of cards. Barton defined the following types of matches:

1. All K cards in the same position have the same label.
2. The card in a given position in the target pack has the same label as at least one of the $K-1$ corresponding cards in the $K-1$ other packs.
3. At least two of the K cards in a given position have the same label.

McMahon (1915) managed to give some highly abstract formulas in certain of these more complex situations, but otherwise there has been little progress in this area.

A different kind of matching problem has arisen in genetics; see Levene (1949). It can be formulated as the random splitting of a pack with composition (2^N) into two equally sized packs. Barton (1958) showed that the factorial moments for the number of matches are given by

$$\mu'_{[r]} = \frac{N!\left(N-\frac{1}{2}-r\right)!}{(N-r)!\left(N-\frac{1}{2}\right)!2^n}, \tag{10.27}$$

where $\xi!$ is taken as $\Gamma(\xi+1)$ when ξ is positive but not an integer (Section 1.1.2).

Barton (1958) next investigated the generalization to a pack of composition (K^N) randomly split into K equally sized packs; the factorial moments for the number of matches of type 1 are

$$\mu'_{[r]} = \frac{N!N!(K!)^r(KN-Kr)!}{(N-r)!(N-r)!(KN)!}. \tag{10.28}$$

Anderson's (1943) problem concerned a pack of composition (K^N) split into K packs each with composition (1^N). He obtained the following expression for the probabilities of the number of matches of type 1:

$$\Pr[X=x] = \sum_{j=0}^{N-x} \frac{(-1)^j}{x!j!}\left[\frac{(N-x-j)!}{N!}\right]^{K-2}. \tag{10.29}$$

Kemp (1978b) reexamined these three matching schemes. She showed that the pgf corresponding to (10.27) is

$$G(z) = {}_1F_1\left[-N; \tfrac{1}{2}-N; \tfrac{1}{2}(z-1)\right]; \tag{10.30}$$

as N becomes large, this tends to $\exp[N(z-1)/(2N-1)]$, and hence as $N\to\infty$, it tends to a Poisson pgf with parameter $\theta=\frac{1}{2}$. She showed furthermore that the pgf's corresponding to (10.28) and (10.29) are

$$G(z) = {}_1F_{K-1}\left[-N; \frac{1}{K}-N, \frac{2}{K}-N, \ldots, \frac{K-1}{K}-N; \frac{K!(z-1)}{(-K)^K}\right] \tag{10.31}$$

and

$$G(z) = {}_1F_{K-1}[-N; -N, -N, \ldots, -N; (-1)^K(z-1)], \tag{10.32}$$

respectively. In both cases, as N becomes large, the pgf's tend to those of Poisson distributions [with parameters $N^2/\binom{NK}{K}$ and N^{2-K}, respectively]; when N becomes infinite the distributions become the degenerate distribution with pgf $G(z)=1$.

Historical aspects of matching problems have been reviewed by Takács (1980); see also Barton (1958) and David and Barton (1962).

10.4 OCCUPANCY DISTRIBUTIONS

10.4.1 Classical Occupancy and Coupon Collecting

Consider the placement of b balls into c cells (categories, classes). The number of ways in which this can be achieved is clearly c^b, since each of the b balls can be put into any one of the c cells. In the classical occupancy situation each of these ways is assumed to be equiprobable.

There has been considerable interest in the distribution of the number of empty cells. Because this distribution arises in many different contexts, there is no standard notation. We note that previous writers have used the following symbols:

Author	Number of Balls (b)	Number of Cells (c)	Number of Empty Cells (x)
Tukey, 1949	m	N	$N - b'$
Feller, 1957; Harkness, 1970	r	n	m or x
Riordan, 1958	n	m	x
Barton and David, 1959a	n	N	k
Parzen, 1960	n	M	m
Moran, 1968	N	n	m
Johnson and Kotz, 1969	N	k	$k - j$
Kemp, 1978b	N	k	x
Johnson and Kotz, 1977	n	m	k
Fang, 1985	n	m	x
Hald, 1990	n	f	$f - i$

Suppose that the cells are numbered $1, 2, \ldots, c$. Then the probability that there are no balls in any of a specified set of j cells is $(1 - j/c)^b$. Using the notation and methodology of Section 10.2, we have

$$S_j = \binom{c}{j}\left(\frac{c-j}{c}\right)^b, \tag{10.33}$$

and the probability that *at least* x cells are empty is

$$\Pr[X \geq x] = P_x = \sum_{j=x}^{c} \frac{(-1)^{j-x} c!x}{x!(j-x)!(c-j)!j}\left(\frac{c-j}{c}\right)^b. \tag{10.34}$$

Moran (1968) used this method but not the same notation.

The probability that *exactly* x are empty can be written in a number of (equivalent) ways:

$$\begin{aligned}
\Pr[X = x] = P_{[x]} &= \sum_{i=0}^{c-x} (-1)^i \binom{x+i}{i} \binom{c}{x+i} \left(\frac{c-x-i}{c}\right)^b \\
&= \sum_{j=x}^{c} (-1)^{j-x} \frac{c!}{x!(j-x)!(c-j)!} \left(\frac{c-j}{c}\right)^b \\
&= \binom{c}{x} \sum_{\nu=0}^{c-x} (-1)^\nu \binom{c-x}{\nu} \left(1 - \frac{x+\nu}{c}\right)^b \\
&= \binom{c}{x} \frac{\Delta^{c-x} 0^b}{c^b} = \frac{c! S(b, c-x)}{x! c^b},
\end{aligned} \tag{10.35}$$

where $S(b, c-x)$ is a Stirling number of the second kind (Section 1.1.3).

A familiar textbook example of this occupancy distribution relates to birthdays. Consider a community of b people whose birthdays occur randomly and independently and suppose that a year contains $c = 365$ days. Let X be the number of days of the year when nobody has a birthday. Then

$$\Pr[X = x] = \binom{365}{x} \sum_{\nu=0}^{365-x} (-1)^\nu \binom{365-x}{\nu} \left(\frac{365-x-\nu}{365}\right)^b. \tag{10.36}$$

Moran has pointed out that (10.35) still holds for $b < c$ and that alternative derivations are possible; see, for example, Domb's (1952) application of the result to a cosmic ray problem.

The factorial moments for the *distribution of the number of empty cells* are

$$\mu'_{[r]} = \frac{c!}{(c-r)!} \left(\frac{c-r}{c}\right)^b; \tag{10.37}$$

in particular

$$\begin{aligned}
\mu &= (c-1)^b c^{1-b}, \\
\mu_2 &= (c-1)(c-2)^b c^{1-b} + (c-1)^b c^{1-b} - (c-1)^{2b} c^{2-2b}.
\end{aligned} \tag{10.38}$$

The pgf is

$$\begin{aligned}
G(z) &= \sum_{r=0}^{c} \binom{c}{r} \left(\frac{c-r}{c}\right)^b (z-1)^r \\
&= {}_bF_{b-1}[1-c, \ldots, 1-c; -c, \ldots, -c; 1-z],
\end{aligned} \tag{10.39}$$

showing that the distribution is generalized hypergeometric factorial (Kemp, 1978b); see Section 2.4.2.

Von Mises (1939) seems to have been the first to prove that the distribution tends to a Poisson distribution with parameter θ when b and c both tend to infinity such that $c\exp(-b/c) = \theta$ remains bounded. Feller (1957, pp. 93–95) gave an alternative proof and a short illustrative table. David (1950) constructed tables of the probabilities for $b = c$ when $c = 3(1)20$; see Nicholson (1961) for certain probability levels in the tails of the distribution, also see Owen (1962).

Many generalizations of the classical occupancy distribution have been studied. Parzen (1960) and David and Barton (1962) have examined, for example, the problem of specified occupancy. Suppose that precisely ℓ of the c cells are specified ($\ell \le c$). Then the distribution of the number of empty cells among the ℓ specified cells has the factorial moments

$$\mu'_{[r]} = \frac{\ell!}{(\ell - r)!}\left(\frac{c-r}{c}\right)^b \tag{10.40}$$

and the probability mass function (pmf)

$$\Pr[X = x] = \sum_{j=x}^{\ell} \frac{(-1)^{j-x}\ell!}{(\ell - j)!x!(j-x)!}\left(\frac{c-j}{c}\right)^b. \tag{10.41}$$

This is also a GHFD; its pgf can be restated (Kemp, 1978b) as

$$G(z) = {}_{b+1}F_b[-\ell, 1-c, \ldots, 1-c; -c, \ldots, -c; 1-z]. \tag{10.42}$$

The moments of the numbers of cells with given numbers of balls in them have been investigated by von Mises (1939), Tukey (1949), and Johnson and Kotz (1977) inter alia; see also Fang (1985).

An interesting problem of a more general kind was formulated by Tukey (1949) in the contexts of breakages of chains in the oxidation of rubber, and irradiation and mutation of bacteria. Suppose that a total of b balls are distributed at random among an *unknown* number of cells including ℓ specified cells. The total number of balls falling in the ℓ cells is b if the total number of cells is ℓ; otherwise it has a binomial distribution. Let the probability of a specified ball falling into any given one of the ℓ specified cells be p. Then the distribution of the number X_u of (specified) cells containing u balls each has the expected value

$$E[X_u] = \ell\binom{b}{u}(1-p)^b\left(\frac{p}{1-p}\right)^u, \tag{10.43}$$

variance

$$\text{Var}(X_u) = E[X_u]\{1 - [1 - \omega(u,u)]E[X_u]\}, \tag{10.44}$$

and covariance

$$\text{Cov}(X_u, X_v) = [\omega(u,v) - 1]E[X_u]E[X_v], \tag{10.45}$$

where

$$\omega(u,v) = \left(\frac{\ell - 1}{\ell}\right)\frac{(b-u)!(b-v)!}{b!(b-u-v)!}\left[1 - \left(\frac{p}{1-p}\right)^2\right]^b\left(\frac{1-p}{1-2p}\right)^{u+v}.$$

Harkness (1970) [see also Sprott (1957) and Johnson and Kotz (1977)] has examined the problem of "leaking urns"; this is a randomized occupancy problem. Consider, like Sprott, the number of targets hit when b shots are fired randomly at c targets and the probability that any particular shot actually hits its target is p. This is the classical occupancy problem with a probability p that a ball randomly placed into a cell stays in the cell and a probability $1-p$ that it immediately disappears. The probability that x cells are unoccupied is now

$$\Pr[X=x] = \sum_{i=0}^{c-x} \frac{(-1)^i c!}{x!i!(c-x-i)!} \left(\frac{c-(x+i)p}{c} \right)^b . \tag{10.46}$$

Harkness has pointed out that these probabilities satisfy

$$\begin{aligned} \Pr[X=x|b+1,c,p] &= \left(1-\frac{px}{c}\right)\Pr[X=x|b,c,p] \\ &\quad + p(x+1)c^{-1}\Pr[X=x+1|b,c,p] \end{aligned} \tag{10.47}$$

(the same recursion relationship holds with $p=1$ for the classical occupancy distribution).

It can be shown that this leaking-urn distribution tends to a Poisson distribution with parameter $\lambda = c(1-p/c)^b$ as b and c become large. Harkness obtained certain moment properties, discussed applications and a Markov chain derivation, and suggested a binomial approximation to the distribution. His tables of probabilities demonstrate the superiority of his binomial approximation compared with a Poisson approximation.

For work on restricted occupancy models, see Fang (1985) and references therein. Yet other types of occupancy models, such as contagious occupancy models, have been studied by Barton and David (1959a) and David and Barton (1962).

Johnson and Kotz (1977) have written a very full account of occupancy distributions; besides the classical problem and the problem of leaking urns, they have discussed sequential occupancy and committee problems. A typical committee problem is as follows (Mantel and Pasternack, 1968): A group consists of c individuals any w of whom may be selected to serve on a committee. If b committees, each of size w, are chosen randomly, what is the probability that exactly x individuals will be on committees?

Occupancy problems can be reinterpreted as sampling problems. Parzen (1960) gives a very clear account. The classical occupancy problem can, for instance, be viewed as the withdrawal with replacement of a sample of size b from a collection containing c distinguishable items. The number $c-x$ of *occupied cells* now corresponds to the number of different items that have been sampled. The distribution of $Y = c - X$ is known by the following names:

Stevens–Craig	Stevens, 1937; Craig, 1953
Arfwedson	Arfwedson, 1951
Coupon collecting	David and Barton, 1962
Dixie cup	Johnson and Kotz, 1977

The number of different items in the sample has the pmf

$$\Pr[Y = y] = \binom{c}{y}\frac{\Delta^y 0^b}{c^b} = \frac{c!S(b, y)}{(c - y)!c^b}. \tag{10.48}$$

The mean and variance are

$$\begin{aligned} \mu &= c\left[1 - \left(\frac{c-1}{c}\right)^b\right], \\ \mu_2 &= c(c-1)(c-2)^b c^{-b} + c(c-1)^b c^{-b} - c^2(c-1)^{2b}c^{-2b}. \end{aligned} \tag{10.49}$$

The joint distribution of the numbers of the c different types of item can be regarded as multinomial. A modified form of the Stevens–Craig distribution is obtained if it is supposed that there is an upper limit s to the value of each of the multinomial variables. This corresponds to a situation in which the c cells are each capable of containing up to s balls and the b balls are placed at random in the available cells, with the proviso that, if any cells are full, then a choice must be made among the remaining cells. The probability that there are exactly x cells each containing at least one ball is

$$\begin{aligned} \Pr[X = x] = \frac{c!b!(cs - b)!}{x!(c - x)!(cs)!}\Bigg[\binom{sx}{b} - \binom{x}{1}\binom{s(x-1)}{b} + \binom{x}{2}\binom{s(x-2)}{b} \\ - \cdots + (-1)^{\nu}\binom{x}{\nu}\binom{s(x-\nu)}{b}\Bigg], \end{aligned} \tag{10.50}$$

where ν is the greatest integer such that $s(x - \nu) \geq b$. The mean and variance of the distribution of X are

$$\mu = c(1 - q_1), \qquad \mu_2 = c[q_1 - q_2 + c(q_2 - q_1^2)], \tag{10.51}$$

where

$$q_1 = \binom{s(c-1)}{b}\Big/\binom{sc}{b} \qquad \text{and} \qquad q_2 = \binom{s(c-2)}{b}\Big/\binom{sc}{b}.$$

A more general problem of this kind was solved by Richards (1968) using an exponential-type generating function; the use of exponential-type generating functions is explained in Riordan (1958). Barton and David (1959a) obtained limiting distributions for this and a number of similar distributions.

Feller (1957) listed no fewer than 16 applications of the classical occupancy distribution. These include the birthday example mentioned above, irradiation in biology, cosmic ray counts, gene distributions, polymer reactions, and the theory of photographic emulsions. Harkness (1970) has put on record further instances, including Mertz and Davies (1968) predator–prey interpretation, Peto's (1953) dose–response model for invasion by microorganisms, and the epidemic spread model of Weiss (1965).

In Barton and David (1959b) the problem of placement of one-colored balls in cells is generalized to many colors and is interpreted in terms of dispersion of species. An application to computer storage is implicit in Denning and Schwartz (1972). In the usual birthday problem the question is the probability of at least one birthday match among a group of n people. Meilijson, Newborn, Tenenbein, and Yechieli (1982) examined the distribution of the number of matches as well as the distribution of the number of matched people; their work has implications regarding computer storage overflow.

A unified approach to a number of occupancy problems has been achieved by Holst (1986) by embedding in a Poisson process; this shows that many of these problems are closely related to the properties of order statistics from the gamma distribution. Holst gives a number of relevant references. Limit theorems for occupancy problems are studied in Johnson and Kotz (1977, Chapter 6), Kolchin, Sevast'yanov, and Chistyakov (1978), and Barton and David (1959a).

10.4.2 Maxwell–Boltzmann, Bose–Einstein, and Fermi–Dirac Statistics

Consider a physical system comprising a very large number b of "particles" of some kind, for example, electrons, protons, and photons. Suppose that there are c states (energy levels) in which each particle can be. The overall state of the system is then $(b_1, b_2, \ldots, b_c)$, and equilibrium is defined as the overall state with the highest probability of occurrence.

If all c^b arrangements are equally likely, the system is said to behave according to Maxwell–Boltzmann statistics (the term "statistics" is used here in a sense meaningful to physicists). The probability that there are x particles ($x \leq b$) in a particular state is

$$\Pr[X = x] = \binom{b}{x}\left(\frac{1}{c}\right)^x\left(1 - \frac{1}{c}\right)^{b-x}, \tag{10.52}$$

and the probability that exactly k out of ℓ specified states ($\ell \leq c$) are unoccupied is given by the classical occupancy problem documented in the previous section. The classical theory of gases (at low densities and not very low temperatures) was based on Maxwell–Boltzmann statistics.

Modern theory and experimentation (particularly at very low temperatures) have, however, yielded two far more plausible hypotheses concerning the behavior of physical systems, namely Bose–Einstein and Fermi–Dirac statistics. In both cases particles are assumed to be indistinguishable (rather than distinguishable as for Maxwell–Boltzmann statistics). The difference between the two cases is the assumption of whether or not the Pauli exclusion principle holds. This postulates that there cannot be more than one particle in any particular state.

For Bose–Einstein statistics particles are assumed to be indistinguishable but not to obey the Pauli exclusion principle. The total number of arrangements of b particles in the c states can be derived by choosing b particles one at a time, with replacement, from the set of c states and hence is equal to

$$\binom{b + c - 1}{b}.$$

Muthu (1982) reports that particles having integer spin (intrinsic momentum equal to $0, h/2\pi, 2h/2\pi, \ldots$), such as photons and pions, obey Bose–Einstein statistics and are called *bosons*. The probability that there are x particles ($x \leq b$) in a particular state is now

$$\Pr[X = x] = \binom{b+c-x-2}{b-x} \Bigg/ \binom{b+c-1}{b}, \tag{10.53}$$

and the probability that exactly k out of ℓ specified states ($\ell \leq c$) are unoccupied is

$$\Pr[K = k] = \binom{\ell}{\ell-k}\binom{b+c-1-\ell}{b+k-\ell} \Bigg/ \binom{b+c-1}{b}; \tag{10.54}$$

this is of course a hypergeometric distribution (Section 6.2.1).

For Fermi–Dirac statistics particles are assumed not only to be indistinguishable but also to satisfy the Pauli exclusion principle. There can now be no more than one particle per state. The total number of arrangements of the b particles in the c states is now

$$\binom{c}{b}.$$

The probability that a specified state has x particles in it is

$$\Pr[X = x] = \binom{c-1}{b-x} \Bigg/ \binom{c}{b}, \qquad x = 0,\ 1, \tag{10.55}$$

and the probability that exactly k out of ℓ specified states ($\ell \leq c$) are empty is

$$\Pr[K = k] = \binom{\ell}{\ell-k}\binom{c-\ell}{b+k-\ell} \Bigg/ \binom{c}{b}. \tag{10.56}$$

Again we have a hypergeometric distribution.

All known elementary particles have either integer spin or half-integer spin (this determines whether their waveform is symmetric or antisymmetric). Particles with half-integer spin, such as electrons and protons, are thought to obey Fermi–Dirac statistics; they have been called *fermions*.

Johnson and Kotz (1977) gave a full account of the properties of Maxwell–Boltzmann, Bose–Einstein, and Fermi–Dirac occupancy schemes. They commented that the expected number of particles X in a specified state is b/c under each of the three systems. The variance of X is not the same for the three schemes, however. For Maxwell–Boltzmann statistics $\text{Var}(X) = b(c-1)/c^2$, for Bose–Einstein statistics $\text{Var}(X) = b(c-1)(b+c)/[c^2(c+1)]$, and for Fermi–Dirac statistics $\text{Var}(X) = b(c-b)/c^2$.

Loève (1963) studied these three occupancy models as particular cases of a more general model.

10.4.3 Specified Occupancy and Grassia–Binomial Distributions

Reparameterization of the specified occupancy pgf (10.41), using $b \mapsto a$, $c \mapsto 1/b$, $\ell \mapsto n$, and $j \mapsto r$, gives the pmf

$$\Pr[X = x] = \binom{n}{x} \sum_{r=x}^{n} \binom{n-x}{r-x} (-1)^{r-x} (1 - br)^a, \qquad x = 0, 1, \ldots, n, \tag{10.57}$$

where $0 < a$, $0 < b < 1/n$ and the factorial moment generating function

$$G_o(1 + t) = \sum_{r=0}^{n} \frac{n!}{(n-r)!} (1 - br)^a \frac{t^r}{r!} \tag{10.58}$$

(the distribution has finite support and only a finite number of the factorial moments are meaningful).

Reparameterization of the leaking-urns pgf (10.46), using $b \mapsto a, c \mapsto n, p \mapsto bn, i \mapsto r - x$, gives a distribution that is mathematically the same. The pgf has the form

$$G_o(z) = \sum_{x=0}^{n} \sum_{r=x}^{n} \frac{n!(1 - br)^a (-1)^{r-x} z^x}{(n-r)!(r-x)!x!} \tag{10.59}$$

with $0 < a$, $0 < n$, $0 < bn \le 1$ and

$$\lim_{b \to 0} G_o(z) = z^n. \tag{10.60}$$

Kemp and Kemp (2004) showed that (10.60) enables the distribution to be characterized by the factorial moment relationship

$$\frac{d \log \mu'_{[r]}}{db} = \frac{-ar}{1 - br}, \qquad r = 0, 1, \ldots, n. \tag{10.61}$$

Harkness (1970) recognized the marked similarity of the leaking-urns distribution (and hence the specified occupancy distribution) to the distribution of the number of survivors at the end of a carrier-borne epidemic with n susceptibles and a carriers. Here the pmf is

$$\Pr[X = x] = \binom{n}{x} \sum_{r=x}^{n} \binom{n-x}{r-x} (-1)^{r-x} (1 + br)^{-a}, \qquad x = 0, 1, \ldots, n; \tag{10.62}$$

see Weiss (1965), Dietz (1966), and Downton (1967). Variants of the basic model that yield the same distribution are described by Daley and Gani (1999). The pgf is

$$G(z) = \sum_{x=0}^{n} \sum_{r=x}^{n} \frac{n!(1 + br)^{-a} (-1)^{r-x} z^x}{(n-r)!(r-x)!x!}, \qquad 0 < a, \qquad 0 < b, \tag{10.63}$$

and the fmgf is

$$G(1+t) = \sum_{r=0}^{n} \frac{n!}{(n-r)!}(1+br)^{-a}\frac{t^r}{r!}. \tag{10.64}$$

Again

$$\lim_{b\to 0} G(z) = z^n.$$

Kemp and Kemp (2004) found that the distribution is characterized by the factorial moment relationship

$$\frac{d\log \mu'_{[r]}}{db} = \frac{-ar}{1+br}, \qquad r = 0, 1, \ldots, n. \tag{10.65}$$

The carrier-borne epidemic distribution has surfaced in the literature several times under different guises. Grassia (1977) obtained mixed binomial distributions by (i) transforming the binomial parameter p to $e^{-\lambda}$ and (ii) transforming $q(= 1-p)$ to $e^{-\lambda}$ and letting q have a gamma distribution. These are called the Grassia I–binomial and the Grassia II–binomial distributions, respectively.

For the Grassia II–binomial, the pmf of the outcome is

$$\begin{aligned}
\Pr[X=x] &= \int_0^\infty \binom{n}{x}(1-e^{-\lambda})^x(e^{-\lambda})^{n-x}\cdot\frac{e^{-\lambda/b}\lambda^{a-1}\,d\lambda}{\Gamma(a)b^a}\\
&= \binom{n}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^{x-j}\int_0^\infty (e^{-\lambda})^{n-j}\cdot\frac{e^{-\lambda/b}\lambda^{a-1}\,d\lambda}{\Gamma(a)b^a}\\
&= \binom{n}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^{x-j}[1+b(n-j)]^{-a}, \qquad 0<a, \qquad 0<b,
\end{aligned} \tag{10.66}$$

as in Alanko and Duffy (1996). Equivalently it is the mixed binomial distribution where the parameter q has the log-gamma distribution of Consul and Jain (1971). Chatfield and Goodhardt (1970) had previously suggested this mixed binomial model as an alternative to the beta–binomial model for consumer purchasing behavior. Alanko and Duffy (1996) explored the properties and estimation for this distribution in depth in their study of daily consumption of alcohol per week ($n = 7$).

Because

$$\int_0^1 \binom{n}{x} p^x(1-p)^{n-x} f(p)\,dp = \int_0^1 \binom{n}{y}(1-q)^{n-y}q^y f(q)\,dq,$$

where $Y = n - X$ and $q = 1 - p$, the Grassia I–binomial is the reversed Grassia II–binomial distribution. It therefore has the pgf (10.63) and is mathematically the same as the carrier-borne epidemic distribution.

The interest of Kemp and Kemp (2004) in the factorial moment characterization of the Grassia I–binomial distribution arose from a need to discriminate between it and the binomial distribution. They showed that the binomial distribution is characterized by the factorial moment relationship

$$\frac{d \log \mu'_{[r]}}{db} = \frac{-r}{1 - br}, \qquad r = 0, 1, \ldots, n; \tag{10.67}$$

thus $\mu'_{[2]}/\mu^2 = (n-1)/n$, as is well known. The Grassia I–binomial distribution is characterized by (10.65). Hence

$$\frac{\mu'_{[2]}}{\mu^2} = \frac{n-1}{n}\left(\frac{1 + 2b + b^2}{1 + 2b}\right) > \frac{n-1}{n}.$$

Kemp and Kemp used this property to reject a binomial model for Chatfield's (1970) data for brand D; they obtained a satisfactory fit to the data.

10.5 RECORD VALUE DISTRIBUTIONS

Properties and distributional results for record values from continuous distributions have received a great deal of attention over the years. There has been much less research on record values from discrete distributions—the possibility of ties makes their study far less straightforward. Chandler (1952) studied the distribution of the serial numbers of record values in an infinite time series comprising independent sample values $y_1, y_2, y_3, \ldots$ from a fixed continuous distribution. He was concerned with those members of the series that are smaller than all preceding members; he called these lower record values. For example, if the observed sequence of values is 2.1, 3.6, 1.9, 1.7, 1.4, 2.5, 4.3, ..., then the first lower record value is the first observation, 2.1, the second record value is the third observation, 1.9, the third record value is the fourth observation, 1.7, and so on. Similarly, the first, second, third, ..., higher record values are the first, second, seventh, ..., observations. (By convention the first observation is regarded as both a low and a high record.)

Let X_r be the rth lower record value and let U_r be its serial number in the sequence of observations. Chandler obtained the distribution of U_r and also the distribution of $U_r - U_{r-1}$ (the number of observations between the $(r-1)$th and the rth lower record). Clearly the same distributions hold if high records are being considered.

Chandler showed that

$$\Pr[U_r = j] = j^{-1} K_{r-2}(j-1), \qquad j = r, r+1, \ldots, r \geq 2, \tag{10.68}$$

where $K_0(y) = y^{-1}$ when $y \geq 1$ and $K_{r+1}(y) = y^{-1} \sum_{i=r+1}^{y-1} K_r(i)$ when $y \geq r + 2$. He also showed that

$$\Pr[U_r - U_{r-1} \geq t] = M_{r-2}(t), \qquad t = 1, 2, 3, \ldots, r \geq 2, \tag{10.69}$$

where $M_0(y) = y^{-1}$ when $y \geq 1$ and $M_{r+1}(y) = y^{-1} \sum_{i=1}^{y} M_r(i)$ when $y \geq 1$.

Both distributions have infinite means; also under very general conditions they are independent of the distribution of the sample observations. Chandler (1952) provided tables from which the probabilities (10.68) and (10.69) can be obtained.

Foster and Stuart (1954) showed that for the *total* number of records in a sequence of length N the mean and variance are

$$\mu = 2\sum_{j=2}^{N} j^{-1} \qquad \text{and} \qquad \mu_2 = 2\sum_{j=2}^{N} j^{-1} - 4\sum_{j=2}^{N} j^{-2}. \tag{10.70}$$

The pgf is $(N!)^{-1}\prod_{j=1}^{N-2}(j+2z)$. The distribution is asymptotically normal as N increases.

A related distribution is that of the number of *local* records. These are values that are the largest in *some* sequence of k successive values. Thus x_i is a local record if it is the greatest value among $\{x_{i-k+1}, x_{i-k+2}, \ldots, x_i\}$, or among $\{x_{i-k+2}, \ldots, x_{i+1}\}$, or among $\{x_{i-k+3}, \ldots, x_{i+2}\}, \ldots,$ or among $\{x_i, x_{i+1}, \ldots, x_{i+k-1}\}$. The distribution of the number R of such local maxima in a sequence of n values $x_1, x_2, \ldots, x_n$ is sometimes called a *Morse distribution* (Freimer et al., 1959) (this name has also been applied to the distribution of $S = n - R$). The name arose from relevance to a method of machine-decoding hand-keyed Morse code, based upon the identification of the largest and the smallest of each successive sequence of six spaces.

Other relevant early references to distributions of this kind were Austin et al. (1957) and David and Barton (1962). Shorrock (1972a,b) and Resnick (1973) studied record value times. A profound paper on the independence of record processes was that by Goldie and Rogers (1984). Important work on the position of records concentrated on the random walks performed by the elements in an ordered sequence as new elements are inserted; see Blom and Holst (1986) and Blom, Thorburn, and Vessey (1990). A good elementary review paper concerning record value distributions was Glick (1978). Vervaat (1973) remains an important reference. More recent surveys are those by Nevzorov (1987) and Nagaraja (1988a).

The lack-of-memory property of the geometric distribution gives it a role comparable to that of the exponential distribution. There are a number of characterizations of the geometric distribution based on record values. If X_j is observed at time j, then the record time sequence $\{T_n, n \geq 0\}$ is defined as $T_0 = 1$ with probability 1 and $T_n = \min\{j : X_j > X_{t_{n-1}}\}$ for $n \geq 1$. The record value sequence $\{R_n\}$ is defined as $R_n = X_{T_n}$, $n = 0, 1, 2, \ldots$. Suppose that the X_j's are iid geometric variables with pmf

$$p_x = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots. \tag{10.71}$$

Then $R_n = X_{T_n} = \sum_{j=0}^{n} X_j$ is distributed as the sum of $n+1$ iid geometric variables, that is, as a shifted negative binomial rv with pgf

$$G_{R_n}(z) = \left(\frac{pz}{1-qz}\right)^{n+1}, \tag{10.72}$$

where $q = 1 - p$. Each of the following properties characterizes the geometric distribution:

(i) *Independence*: The rv's $R_0, R_1 - R_0, \ldots, R_{n+1} - R_n, \ldots$ are independent.

(ii) *Same Distribution*: $R_{n+1} - R_n$ has the same distribution as R_0.

(iii) *Constant Regression*: $E[R_{n+1} - R_n | R_n]$ is constant.

For these and other characterizations see Srivastava (1979, 1981) and Rao and Shanbhag (1994, 1998). Arnold, Balakrishnan, and Nagaraja (1998) cite further references.

Geometric-tail distributions have also been studied. A rv is said to have a geometric-tail j distribution if

$$\Pr[X = x] = p \Pr[X \geq x] \quad \text{for } x \geq j,$$

that is, $\Pr[X = x] = c(1 - p)^x$ for all $x \geq j$, where c is a constant. Arnold, Balakrishnan, and Nagaraja (1998) give good accounts, with references, of characterizations for these distributions in their book.

These authors also discuss results for weak records. An observation is called a weak record if it is as least as large as any previous record (but maybe no larger). The key reference is Stepanov (1992).

Random record models arise when the sequence of observations has a random length, not an infinite length as assumed above. When the observations occur at time points determined by a point process, the model is called a point process record model (PPRM). Arnold, Balakrishnan, and Nagaraja (1998) devote a chapter to these models, including ones for the Poisson, renewal, and various birth processes.

10.6 RUNS DISTRIBUTIONS

10.6.1 Runs of Like Elements

A considerable variety of forms of distributions have been discovered from the study of *runs* in sequences of observations. Among the simplest of such "runs" are sequences of identical values where the random variables giving rise to the observed values are independent and take the values 0 and 1 with probabilities p and $q = 1 - p$, respectively. A concise account of the early history of attempts to find the distribution of the total number of runs in such binomial (Bernoulli) sequences has been given by Mood (1940). Bortkiewicz (1917) obtained the mean and variance of this distribution. An asymptotic result by Wishart and Hirschfeld (1936) showed that the asymptotic standardized distribution of the total number of runs, as the length of the sequence increases, is normal.

Formulas for the mean and variance of the number of runs of a specified length were obtained by Bruns (1906); von Mises (1921) demonstrated that the distribution can be approximated by a Poisson distribution.

Results concerning the distribution of runs of like elements when the numbers n_0, n_1 of 0's and 1's in the sequence are fixed (i.e., the conditional distribution given n_0 and n_1) were published by Ising (1925) and Stevens (1939). Let the number of runs of i consecutive 1's be denoted by r_{1i} and the number of runs of j consecutive 0's be denoted by r_{0j}. Then the probability of obtaining such sets of runs of various lengths, conditional on $\sum_i ir_{1i} = n_1$, $\sum_j jr_{0j} = n_0$, with $\sum_i r_{1i} = r_1$, $\sum_j r_{0j} = r_0$, $n_1 + n_0 = N$, is

$$F(r_1, r_0)\binom{r_1}{r_{11}, r_{12}, \ldots, r_{1N}}\binom{r_0}{r_{01}, r_{02}, \ldots, r_{0N}}\Big/\binom{N}{n_1}, \tag{10.73}$$

where $F(r_1, r_0) = 0$, 1, 2 according as $|r_1 - r_0| > 1$, $|r_1 - r_0| = 1$, $r_1 = r_0$, respectively.

The joint distribution of the total numbers of runs R_1, R_0 is

$$\Pr[R_0 = r_0, R_1 = r_1] = F(r_1, r_0)\binom{n_1 - 1}{r_1 - 1}\binom{n_0 - 1}{r_0 - 1}\Big/\binom{N}{n_1}, \tag{10.74}$$

where $r_0 + r_1 \geq 2$, $r_1 \leq n_1$, $r_0 \leq n_0$. From this result it is straightforward to deduce that the marginal distribution of R_1 is

$$\begin{aligned}\Pr[R_1 = r_1] &= \binom{n_1 - 1}{r_1 - 1}\binom{n_0 + 1}{n_0 + 1 - r_1}\Big/\binom{n_0 + n_1}{n_1},\\ &= \binom{n_0}{r_1 - 1}\binom{n_1}{n_1 - r_1}\Big/\binom{n_0 + n_1}{n_1 - 1},\end{aligned} \tag{10.75}$$

where $r_1 = 1, 2, \ldots, \min(1 + n_0, n_1)$.

This is sometimes called the *Ising–Stevens distribution*. The pgf is

$$G(z) = z\frac{{}_2F_1[-n_0, 1 - n_1; 2; z]}{{}_2F_1[-n_0, 1 - n_1; 2; 1]}, \tag{10.76}$$

and hence the distribution is a shifted form of a special case of the classical hypergeometric distribution (see Section 6.2.1). The mean and variance are

$$E[R_1] = \frac{(n_0 + 1)n_1}{n_0 + n_1}, \qquad \operatorname{Var}(R_1) = \frac{n_0(n_0 + 1)n_1(n_1 - 1)}{(n_0 + n_1)^2(n_0 + n_1 - 1)}. \tag{10.77}$$

Finally the mean and variance of the *total* number of runs $R_0 + R_1$ are

$$E[R_0 + R_1] = 1 + \left(\frac{2n_0n_1}{n_0 + n_1}\right), \qquad \operatorname{Var}(R_0 + R_1) = \frac{2n_0n_1(2n_0n_1 - n_0 - n_1)}{(n_0 + n_1)^2(n_0 + n_1 - 1)}. \tag{10.78}$$

The asymptotic normality of the distribution of R_1 (or of $R_0 + R_1$) as $n_0 + n_1$ increases was established by Wald and Wolfowitz (1940).

Instead of assuming that n_0 and n_1 are known, one can suppose that n_0 and n_1 have arisen by some form of sampling. David and Barton (1962) investigated the following scenarios:

1. There is a finite population of n elements with np of one kind and $n - np$ of the other. A sample of size N is drawn with replacement from this population; n_0 and n_1 are now the numbers of the two kinds of elements in a binomial sample ($n_0 + n_1 = N$).
2. Alternatively, inverse sampling with replacement for a fixed number n_0 of elements of one kind can be used; n_1 is now the number of elements of the second kind that have been sampled.

Consider now arrangements around a circle. When a red objects and b blue objects are arranged at random in a circle, the number of red runs X is identically equal to the number of blue runs; the pmf is

$$\begin{aligned}\Pr[X = x] &= \binom{a-1}{x-1}\binom{b}{b-x}\Big/\binom{a+b-1}{b-1}\\ &= \binom{b-1}{x-1}\binom{a}{a-x}\Big/\binom{a+b-1}{a-1}, \qquad x = 1, 2, \ldots, \min(a, b),\end{aligned}$$

and the pgf is

$$G(z) = z\frac{{}_2F_1[1-a, 1-b; 2; z]}{{}_2F_1[1-a, 1-b; 2; 1]}. \tag{10.79}$$

This again is an Ising–Stevens distribution (Stevens, 1939). The mean and variance are

$$\mu = \frac{ab}{a+b-1}, \qquad \mu_2 = \frac{a(a-1)b(b-1)}{(a+b-1)^2(a+b-2)}. \tag{10.80}$$

A natural generalization of runs of two kinds along a line is to consider arrangements of $k > 2$ kinds of elements, conditional on there being $n_1, n_2, \ldots,$ n_k elements of the 1st, 2nd, ..., kth kind, respectively (with $\sum_j n_j = N$). Let r_{ij} now denote the number of runs of elements of the ith kind that are exactly of length j and put

$$r_i = \sum_{j=1}^{n_i} r_{ij}, \qquad i = 1, 2, \ldots, k. \tag{10.81}$$

Then the probability of obtaining the array $\{r_{ij}\}$ of runs is

$$F(r_1, r_2, \ldots, r_k)\left[\prod_{i=1}^{k}\binom{r_1}{r_{i1}, r_{i2}, \ldots, r_{in_i}}\right]\Big/\binom{N}{n_1, n_2, \ldots, n_k},$$

where $F(r_1, r_2, \ldots, r_k)$ is the coefficient of $\prod_{i=1}^{k} x_i^{r_i}$ in the expansion of

$$\left(\sum_{i=1}^{k} x_i\right)^k \prod_{j=1}^{k} \left(\sum_{i=1}^{k} x_i - x_j\right)^{r_i - 1}.$$

Various additional aspects of runs distributions are discussed in David and Barton (1962), in the references therein, and in Part IV of the Introduction to David et al. (1966).

An early example of the statistical use of distributions concerning runs is the two-sample test of Wald and Wolfowitz (1940). A more powerful test of this type is that of Weiss (1976). The theory of runs also provides tests of randomness; see Gibbons (1986) for a careful discussion and useful bibliography. Runs are also used for certain tests of adequacy of fit; for this and other applications, for example in quality control, see the helpful review article by Weiss (1988).

10.6.2 Runs Up and Down

Another kind of "run" that has received attention is a run of increasing or decreasing values in a sequence of N independent rv's each having the same distribution, usually assumed to be continuous (obviating the need to consider tied values). Evidently the numbers of such runs are unaltered if each value is replaced by its "rank order," that is, 1 for the smallest value, 2 for the next smallest, and so on, up to N for the largest value. Accordingly, it is only necessary to study the occurrence of runs by considering rearrangements of the integers $1, 2, \ldots, N$. The results will be valid, whatever the common continuous distribution of the rv's giving rise to the observations.

Wallis and Moore (1941) constructed a test of randomness of a sequence by considering the distribution of the number of *turning points* in a sequence of N values $x_1, x_2, \ldots, x_N$. The jth observation constitutes a turning point if either $x_j = \min(x_{j-1}, x_j, x_{j+1})$ or $x_j = \max(x_{j-1}, x_j, x_{j+1})$. Neither x_1 nor x_N can be turning points, but for each of the other $N - 2$ observations, the probability that it is a turning point is $\frac{2}{3}$. Hence the expected value of the number of turning points is $2(N-2)/3$. The variance and third and fourth moments are

$$\begin{aligned} \mu_2 &= \frac{16N - 29}{90}, \\ \mu_3 &= -\frac{16(N+1)}{945}, \\ \mu_4 &= \frac{448N^2 - 1976N + 2301}{4725}. \end{aligned} \tag{10.82}$$

For $N > 12$, the distribution can be treated as normal. The number of runs (up or down) is *one more* than the number of turning points.

Wallis and Moore (1941) have defined *phases* as runs excluding those beginning with x_1 or ending with x_N. There are no phases if there are no turning points; otherwise the number of phases is *one less* than the number of turning points. They defined the *duration* of a phase between turning points x_j and $x_{j'}$ (where $j < j'$) to be $j' - j$; they showed that the expected number of phases of duration d is

$$\frac{2(d^2 + 3d + 1)(N - d - 2)}{(d + 3)!}.$$

Wallis and Moore (1941) traced the history of these distributions back to work by Bienaymé (1874). They ascribed credit for the formula for the distribution of phase duration to Besson. Further references concerning statistical tests based on runs up and down are given in Gibbons (1986).

10.7 DISTRIBUTIONS OF ORDER *k*

10.7.1 Early Work on Success Runs Distributions

"The Probability of throwing a Chance assigned a given number of times without intermission, in any given number of Trials" (De Moivre 1738, p. 243) was interpreted by Todhunter (1865, pp. 184–186) to mean the probability that a run of r successes is completed at the nth trial in a sequence of Bernoulli trials each with probability of success p. Let the probability of this event be u_n, using Todhunter's notation. Todhunter argued that

$$u_{n+1} = u_n + (1 - u_{n-r})qp^r, \quad \text{where } q = 1 - p,$$

and hence found the generating function of u_n; he obtained a formula substantially in agreement with the result given by De Moivre.

Feller (1957, pp. 299–303) treated this problem as an application of the theory of recurrent events and hence showed that the distribution of the trial number X at which the *first* run of length r occurs has pgf

$$G(z) = \frac{p^r z^r (1 - pz)}{1 - z + qp^r z^{r+1}}, \tag{10.83}$$

where $x = r, r + 1, r + 2, \ldots, r = 1, 2, \ldots,$ and $0 < p < 1$. The mean and variance are

$$\mu = \frac{1 - p^r}{qp^r}, \qquad \mu_2 = \frac{1}{(qp^r)^2} - \frac{2r + 1}{qp^r} - \frac{p}{q^2}. \tag{10.84}$$

Clearly $r = 1$ gives an ordinary geometric distribution with support $x = 1, 2, 3, \ldots.$

De Moivre gave without proof a formula for the number of trials needed in order to have a probability of approximately $\frac{1}{2}$ of getting a run of r successes.

Feller proved that the probability q_r that there is no success run of length r in n trials is

$$q_n \approx \frac{1 - p\zeta}{\zeta^{n+1} q(r + 1 - r\zeta)}, \tag{10.85}$$

where ζ is the smallest in absolute value of the roots of the denominator of (10.83), and demonstrated that the formula gives a remarkably good approximation, even for n as low as 4. Feller (1957, pp. 303–304) also considered success runs of either kind and success runs occurring before failure runs.

Shane (1973) gave the name *Fibonacci distribution* to the distribution with pgf (10.83) when $r = 2$ and $p = q = \frac{1}{2}$. He showed that in this particular case

$$\Pr[X = x + 2] = 0.5 \Pr[X = x + 1] + 0.25 \Pr[X = x]$$

and hence that

$$\Pr[X = x] = 2^{-x} F_{x-1}, \qquad x = 2, 3, \dots, \tag{10.86}$$

where $\{F_1, F_2, F_3, F_4, F_5, \dots\}$ are the Fibonacci numbers $\{1, 1, 2, 3, 5, \dots\}$. Taillie and Patil (1986) related this distribution to a problem in disassortative mating. Patil, Boswell, Joshi, and Ratnaparkhi (1984) and Taillie and Patil (1986) named the distribution with pgf (10.83), where $r \geq 1$ and $p = q = \frac{1}{2}$, a *poly-nacci distribution* and expressed the probabilities in terms of rth-order Fibonacci numbers. For general values of both r and p, $r \geq 1$, $0 < p < 1$, Uppuluri and Patil (1983) and Patil et al. (1984) called the distribution (10.83) a *generalized poly-nacci distribution.*

The papers by Philippou and Muwafi (1982) and Philippou, Georghiou, and Philippou (1983) created an upsurge of interest in these distributions under the name *success runs distributions of order k*. From 1984 onward there has been a great number of papers. Many of these have been written by Aki and Hirano and by Philippou, Balakrishnan, and Koutras and their colleagues. Within the confines of this volume it has proved impossible to document all the work that has taken place. Philippou (1986), Charalambides (1986b), and Aki and Hirano (1988) included in their papers good overviews that were up to date at the time that they were written.

The formulas for the probabilities for these distributions are not at first sight very illuminating; also in some cases there is more than one correct published expression for the probabilities. In general the pgf's give more insight into the models that lead to these distributions; see Charalambides (1986b). Also the properties of these distributions can be obtained more readily from their pgf's than from their pmf's. We will quote the pgf's where we can.

Setting $r = k$ in (10.83) gives Philippou and Muwafi's (1982) *geometric distribution of order k*. Immediately related distributions are the *negative binomial of order k*, the *Poisson of order k*, the *logarithmic of order k*, and the *compound Poisson of order k*; see Philippou et al. (1983), Philippou (1983, 1984), Aki et al. (1984), Charalambides (1986b), Hirano (1986), and Philippou (1986).

These and other related distributions of order k are discussed in the following sections.

10.7.2 Geometric Distribution of Order *k*

Consider a sequence of iid Bernoulli trials $X_1, X_2, \ldots$. Then the distribution of T_k where

$$T_k = \min\{n : X_{n-k+1} = \cdots = X_n = 1\}$$

(the waiting time until a sequence of k consecutive successes first occurs) has the pgf (10.83) with $r = k$:

$$\begin{aligned} G(z) &= \frac{p^k z^k (1 - pz)}{1 - z + qp^k z^{k+1}} \\ &= \frac{p^k z^k}{1 - (1 - p^k) z\{[(1 - p^k z^k)/(1 - pz)]/[(1 - p^k)/(1 - p)]\}}. \end{aligned} \tag{10.87}$$

This is Philippou and Muwafi's (1982) *geometric distribution of order* k. Its pgf has the form $G(z) = z^k g_1(g(z))$, where

$$g_1(z) = \frac{p^k}{1 - (1 - p^k)z} \tag{10.88}$$

is the pgf of a geometric distribution and

$$g(z) = z \frac{(1 - p^k z^k)/(1 - pz)}{(1 - p^k)/(1 - p)} \tag{10.89}$$

is the pgf of a truncated geometric distribution with a different parameter and support $1, 2, \ldots, k$. The mean and variance are given by (10.84) with $r = k$. Philippou and Muwafi's expression for the pmf is

$$\Pr[T_k = x] = p^x \sum_{x_1, \ldots, x_k} \binom{x_1 + \cdots + x_k}{x_1, \ldots, x_k} \left(\frac{q}{p}\right)^{x_1 + \cdots + x_k} \tag{10.90}$$

for $x = k, k + 1, \ldots,$ where summation takes place over $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x - k$.

Graphs of some typical geometric distributions of order k are in Hirano et al. (1984) and Balakrishnan and Koutras (2002). Balakrishnan and Koutras also discuss estimation of the mean $\overline{T}_k$.

A closely related rv is the length L_n of the longest success run in n trials. Since $T_k \le n$ implies and is implied by $L_n \ge k$, we have $\Pr[T_k \le n] = \Pr[L_n \ge k]$.

Good bounds on $\Pr[T_k > x]$ were given by Feller (1968). Chen–Stein methods of obtaining bounds (Chen, 1975) have been used by Chryssaphinou and Papastavridis (1990) and Barbour, Holst, and Janson (1992); the latter obtained the very good bounds

$$|\Pr[T_k > x] - \exp[-(n - k + 1)qp^k]| \le (2kq + 1)p^k.$$

Good, much simpler bounds,

$$(1 - p^k)^{n-k+1} \le \Pr[T_k > x] \le (1 - qp^k)^{n-k+1},$$

have been obtained in reliability studies [e.g., by Fu and Koutras (1995); see also Muselli (2000)].

The geometric distribution of order k can be made more general in several ways by changing the basic Bernoulli assumptions. Consider relaxation of the independence assumption by allowing $X_1, X_2, \ldots$ to be a homogeneous Markov chain with transition probabilities

$$p_{i,j} = \Pr[X_t = j | X_{t-1} = i], \qquad t \geq 2, \qquad 0 \leq i, \qquad j \leq 1,$$

and initial probabilities $p_j = \Pr[X_1 = j]$, $j = 0, 1$. The outcome is the *Markov–geometric distribution of order k* with pgf

$$G(z) = \frac{[p_1 + (p_0 p_{01} - p_1 p_{00})z](p_{11}z)^k}{p_{11} + (p_{01}p_{10} - p_{00}p_{11})z - p_{01}p_{10}zA(z)}$$

where

$$A(z) = \frac{1 - (p_{11}z)^k}{1 - p_{11}z}.$$

The many papers on the distribution include Mohanty (1994), Antzoulakos and Philippou (1997), and Koutras (1997a); for Bernoulli trials with higher order dependence see Aki, Balakrishnan, and Mohanty (1996).

The *intervened geometric distribution of order k* (with single intervention) assumes that a single intervention takes place at time T_k when the first failure occurs, $k - 1$ successes having previously occurred. The probability of success in iid Bernoulli trials is assumed to be p_0 until the first failure takes place; thereafter it is assumed to be p_1. Balakrishnan, Balasubramanian, and Viveros (1995) showed that the pgf for T_k is

$$G(z) = (p_0 z)^k + \frac{q_0 p_1^k z^{k+1}(1 - p_1 z)[1 - (p_0 z)^k]}{(1 - p_0 z)(1 - z + q_1 p_1^k z^{k+1})},$$

with mean

$$E[T_k] = (1 - p_0^k)\left(\frac{1}{q_1 p_1^k} - \frac{1}{q_1} + \frac{1}{q_0}\right),$$

where $q_0 = 1 - p_0$ and $q_1 = 1 - p_1$. Balakrishnan and his co-authors discussed the use of their results for start-up demonstration tests in which corrective action is taken when the first failure occurs.

The *extended geometric distribution of order k* of Aki (1985) is a multiparameter distribution based on binary sequences of order k. The pgf is

$$G(z) = \frac{(\prod_{i=1}^k p_i)z^k}{1 - \sum_{i=1}^k (\prod_{j=0}^{i-1} p_j) q_i z^i};$$

see Balakrishnan and Koutras (2002) for further references. Balakrishnan (1997) has studied a more general extension of the geometric distribution of order k.

The geometric distribution of order k and its offshoots have many practical applications in areas as diverse as reliability, start-up demonstration testing, statistical quality control, nonparametric statistical inference, molecular biology, ecology, meteorology, and psychology. Balakrishnan and Koutras (2002) give a full account, including many references.

10.7.3 Negative Binomial Distributions of Order *k*

The *type I negative binomial distribution of order k* is also called the *type I waiting-time distribution of order k*. It is the waiting-time distribution for b runs of successes of length k, given nonoverlapping counting. Whenever a success run is completed in this kind of counting, the pattern of the successes and failures that have occurred becomes irrelevant and counting begins all over again. This distribution is the b-fold convolution of geometric distributions of order k (Philippou, 1984) and therefore has the pgf

$$G_{\mathrm{I}}(z) = [z^k g_1(g(z))]^b = z^{bk} g_2(g(z)) \tag{10.91}$$

with $g_1(z)$ and $g(z)$ as in (10.88) and (10.89), respectively; note that $g_2(z)$ is the pgf of a negative binomial distribution. Expression (10.91) can be rewritten as

$$G_{\mathrm{I}}(z) = \frac{(pz)^{bk}}{[1 - qzA(z)]^b} \tag{10.92}$$

with $A(z) = [1 - (pz)^k]/(1 - pz)$. The mean and variance are

$$\mu = \frac{b(1 - p^k)}{qp^k} \qquad \text{and} \qquad \mu_2 = \frac{b[1 - (2k + 1)qp^k - p^{2k+1}]}{q^2 p^{2k}}, \tag{10.93}$$

and the pmf can be expressed as

$$\Pr[X = x] = p^x \sum_{x_1,\ldots,x_k} \binom{x_1 + \cdots + x_k + b - 1}{x_1, \ldots, x_k, b - 1} \left(\frac{q}{p}\right)^{x_1+\cdots+x_k} \tag{10.94}$$

for $x = kr, kr + 1, \ldots,$ where summation is over $x_1, \ldots, x_k$ such that

$$x_1 + 2x_2 + \cdots + kx_k = x - kr.$$

Type II waiting-time distributions arise when a sequence of k or more successes is counted as a single run; counting does not begin again until the whole sequence is ended. This kind of counting gives rise to the *type II negative binomial distribution of order k* with pgf

$$G_{\mathrm{II}}(z) = [G_I(z)]^b \left(\frac{qz}{1 - pz}\right)^{b-1} = \frac{p^{bk} q^{b-1} z^{bk+b-1}}{[1 - qzA(z)]^b (1 - pz)^{b-1}}. \tag{10.95}$$

The mean and variance are

$$\mu = \frac{b - p^k}{qp^k},$$
$$\mu_2 = \frac{b[1 - (2k+1)qp^k - p^{2k+1}]}{q^2 p^{2k}} + \frac{(b-1)p}{q^2}. \tag{10.96}$$

Ling (1988, 1989) has introduced a third kind of counting called overlapping counting. Consider an uninterrupted sequence of $\ell > k$ successes which follow a failure and are followed by a failure. In overlapping counting this is deemed to give rise to $\ell - k + 1$ success runs of length k. This kind of counting yields the *type III negative binomial distribution of order k* with pgf

$$G_{\mathrm{III}}(z) = \frac{(1-pz)(pz)^{b+k-1}(1 - z + qp^{k-1}z^k)^{b-1}}{(1 - z + qp^k z^{k+1})^b}. \tag{10.97}$$

The mean and variance are now

$$\mu = \frac{bq + p - p^k}{qp^k},$$
$$\mu_2 = \big[-p^{2k+1} - (b-1)(2k-1)p^{k+2} + (4bk - 2k + 1)p^{k+1} \tag{10.98}$$
$$-b(2k+1)p^k - (b-1)p^2 + b\big] \big/ (q^2 p^{2k});$$

see Balakrishnan and Koutras (2002).

All three negative binomial distributions of order k are expounded in depth by these authors. Properties, asymptotics, and estimation are covered. The problems of deriving explicit expressions for the pmf's are discussed and appropriate references are provided. They give diagrams illustrating the shapes of the distributions; a more extensive set of diagrams is in Hirano et al. (1984).

Types I, II, and III Markov negative binomial distributions of order k arise from the nonoverlapping, greater-than-or-equal, and overlapping methods of counting. Their pgf's are convolutions of the corresponding Markov geometric distributions; see Koutras (1997a), Antzoulakos (1999), and Balakrishnan and Koutras (2002) for details.

Types I, II, and III intervened negative binomial distributions of order k can be constructed.

Types I, II, and III extended negative binomial distributions of order k have been researched by Aki (1985) and Hirano and Aki (1987).

10.7.4 Poisson and Logarithmic Distributions of Order k

The *Poisson distribution of order k* was derived as the limiting form as $b \to \infty$ of the Type I negative binomial distribution of order k shifted to the support

0, 1, ... Its pgf can be written as

$$G(z) = \exp\left[-\lambda\left(k - \sum_{i=1}^{k} z^i\right)\right] = e^{\lambda k[h(z)-1]}, \tag{10.99}$$

where $h(z)$ is the pgf of a *discrete rectangular distribution* (Section 6.10.1) with support $1, 2, \ldots, k$. This representation of the pgf shows that the distribution is a particular stuttering Poisson distribution. The mean and variance are

$$\mu = \frac{k(k+1)\lambda}{2} \qquad \text{and} \qquad \mu_2 = \frac{k(k+1)(2k+1)\lambda}{6}, \tag{10.100}$$

and the pmf is

$$\Pr[X = x] = e^{-k\lambda} \sum_{x_1,\ldots,x_k} \frac{\lambda^{x_1+\cdots+x_k}}{x_1!\cdots x_k!} \tag{10.101}$$

for $x = 0, 1, 2, \ldots,$ where summation is overall $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x$; see Tiwari, Tripathi, and Gupta (1992) and Balakrishnan and Koutras (2002), who give references to a number of earlier papers.

A similar result for the limiting form of the type III negative binomial distribution of order k was obtained by Hirano et al. (1991).

The *compound Poisson distribution of order k* is a gamma-mixed Poisson distribution of order k; see Panaretos and Xekalaki (1986a). The pgf is

$$\begin{aligned} G(z) &= \left[1 + \alpha^{-1}\left(k - \sum_{i=1}^{k} z^i\right)\right]^{-c} \\ &= \{1 - \alpha^{-1}k[h(z) - 1]\}^{-c}, \end{aligned} \tag{10.102}$$

where $h(z)$ is again the pgf of a discrete rectangular distribution with support $1, 2, \ldots, k$. The mean and variance are

$$\mu = \frac{k(k+1)c}{2\alpha} \qquad \text{and} \qquad \mu_2 = \frac{k(k+1)(2k+1)c}{6\alpha} + \frac{k^2(k+1)^2c}{4\alpha^2}, \tag{10.103}$$

and the pmf is

$$\Pr[X = x] = \left(\frac{\alpha}{k+\alpha}\right)^c \sum_{x_1,\ldots,x_k} \binom{x_1 + \cdots + x_k + c - 1}{x_1, \ldots, x_k, c-1} \left(\frac{1}{k+\alpha}\right)^{x_1+\cdots+x_k} \tag{10.104}$$

for $x = 0, 1, 2, \ldots,$ where again summation is overall $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x$. Panaretos and Xekalaki (1986a) referred to this distribution as another "negative binomial distribution of order k."

A *logarithmic distribution of order k* is obtained as a limiting form of a left-truncated type I negative binomial distribution of order k; the pgf is

$$G(z) = \ln\left(\frac{1-pz}{1-z+pq^k z^{k+1}}\right) \Big/ (-k\ln p); \qquad (10.105)$$

this has the form

$$G(z) = g_3(g(z)),$$

with $g(z)$ as in (10.89) and with $g_3(z)$ of the form $\ln(1-\alpha z)/\ln(1-\alpha)$, where $\alpha = 1 - p^k$. The mean and variance are

$$\mu = \frac{1-p^k-kqp^k}{qp^k(-k\ln p)}, \qquad \mu_2 = \frac{1-p^{2k+1}-(2k+1)qp^k}{qp^{2k}(-k\ln p)} - \mu^2. \qquad (10.106)$$

The pmf can be expressed as

$$\Pr[X=x] = \frac{p^x}{-k\ln p}\sum_{x_1,\ldots,x_k}\frac{(x_1+\cdots+x_k-1)!}{x_1!\cdots x_k!}\left(\frac{q}{p}\right)^{x_1+\cdots+x_k} \qquad (10.107)$$

for $x = 1, 2, \ldots$, where summation is over $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = x$.

A second *logarithmic distribution of order k* was derived as a limiting form of the gamma-mixed Poisson of order k by Panaretos and Xekalaki (1986a). The pgf is

$$G(z) = \frac{-\ln[1-\theta(z+z^2+\cdots+z^k)/(1+\theta k)]}{\ln(1+\theta k)}. \qquad (10.108)$$

Formulas have been given for the pmf's of some of these distributions [see, e.g., Godbole (1990a)], although they are generally not very friendly. Charalambides (1986b) has helped to clarify relationships between order-k distributions and stopped-sum distributions (Chapter 9) by writing the pgf's of the geometric, negative binomial, logarithmic, and Poisson distributions of order k in the form $g_u(g_v(z))$ (such relationships are not readily seen from the form of the pmf's). Charalambides recognized the relevance of Bell polynomials and hence the relevance of Faà di Bruno's formula (Section 9.1). He gave expressions for the probabilities and also the factorial moments of some of the order-k distributions in terms of truncated Bell and partial Bell polynomials. Aki, Kuboki, and Hirano (1984) and Hirano (1986) also understood the connection with stopped-sum distributions and gave symbolic representations as $\mathcal{F}_1 \bigvee \mathcal{F}_2$ distributions using the notation of Chapter 9.

10.7.5 Binomial Distributions of Order *k*

Consider now the number of occurrences X of the kth consecutive success in n independent Bernoulli trials with probability of success p. The distributions of X have been called the *binomial distribution of order k*. An important application is in the theory of consecutive-k-out-of-n:F failure systems [see, e.g., Shanthikumar (1982)], and hence it is relevant to certain telecommunication and oil pipeline systems and in the design of integrated circuitry (Chiang and Niu, 1981; Bollinger and Salvia, 1982). A consecutive-k-out-of-n:F system is one comprising n ordered components, each with independent probabilities q and p of operating or failing; the system fails when k consecutive components fail. Papers on such systems have appeared in the journal *IEEE Transactions on Reliability*, which has devoted entire sections to them.

The pmf for the *type I binomial distribution of order* k (using nonoverlapping counting) was given independently by Hirano (1986) and by Philippou and Makri (1986) as

$$\Pr[X = x] = p^n \sum_{i=0}^{k-1} \sum_{x_1,\ldots,x_k} \binom{x_1 + \cdots + x_k + x}{x_1, \ldots, x_k, x} \left(\frac{q}{p}\right)^{x_1+\cdots+x_k} \tag{10.109}$$

for $x = 0, 1, 2, \ldots, [n/k]$, where the inner summation is over $x_1, \ldots, x_k$ such that $x_1 + 2x_2 + \cdots + kx_k = n - i - kx$. Neither paper gave an elementary formula for the pgf. Recurrence relations for the probabilities were obtained by Aki and Hirano (1988). If the distribution of Y is negative binomial of order k with parameters b and p as in the previous section, then

$$\Pr[X \geq b] = \Pr[Y \leq n], \qquad b = 0, 1, \ldots, \left[\frac{n}{k}\right]; \tag{10.110}$$

this result appears in Feller (1957, p. 297) and parallels the well-known relationship between the tails of ordinary binomial and negative binomial distributions (Section 5.6). Assuming that $\sum_{x=0}^{[n/k]} \Pr[X = x] = 1$ and $k \leq n \leq 2n - 1$, Philippou (1986) found the mean and variance to be

$$\begin{aligned} \mu &= p^k[1 + (n-k)q], \\ \mu_2 &= p^k[1 + (n-k)q] - p^{2k}[1 + (n-k)q]^2. \end{aligned} \tag{10.111}$$

For the *type II binomial distribution of order* k (using greater-than-or-equal counting) the pmf is

$$\begin{aligned} \Pr[X = x] = {} & \sum_{i=0}^{k-1} \sum_{*} \binom{x_1 + \cdots + x_n}{x_1, \ldots, x_n} p^n \left(\frac{q}{p}\right)^{x_1+\cdots+x_n} \\ & + \sum_{i=k}^{n} \sum_{\dagger} \binom{x_1 + \cdots + x_n}{x_1, \ldots, x_n} p^n \left(\frac{q}{p}\right)^{x_1+\cdots+x_n}, \end{aligned} \tag{10.112}$$

where $\sum_*$ is over all nonnegative integers $x_1, \ldots, x_n$ satisfying

$$\sum_{j=1}^{n} jx_j = n - i \qquad \text{and} \qquad \sum_{j=k+1}^{n} x_j = x$$

and $\sum_{\dagger}$ is over all nonnegative integers $x_1, \ldots, x_n$ satisfying

$$\sum_{j=1}^{n} j x_j = n - i \qquad \text{and} \qquad \sum_{j=k+1}^{n} x_j = x - 1.$$

Ling (1988) gave the following summation expression for the pmf of the *type III binomial distribution of order* k (overlapping counting):

$$\Pr[X = x] = \sum_{i=0}^{n} \sum_{\ddagger} \binom{x_1 + \cdots + x_n}{x_1, \ldots, x_n} p^n \left(\frac{q}{p}\right)^{x_1 + \cdots + x_n}, \tag{10.113}$$

where $\sum_{\ddagger}$ is over all nonnegative integers $x_1, \ldots, x_n$ satisfying

$$\sum_{j=1}^{n} j x_j = n - i \qquad \text{and} \qquad \sum_{j=k+1}^{n} (j - k) x_j = x - \max(0, i - k + 1);$$

see Hirano et al. (1991) for an alternative expression.

Probability generating functions for the binomial distributions of order k are discussed by Balakrishnan and Koutras (2002), but they lack tractability. The Poisson distribution is a limiting form of the types I and II distributions. With the notable exception of Aki and Hirano (1989) there has not been much work on estimation for binomial distributions of order k. *Types I, II, and III Markov binomial distributions of order* k have recently been researched by Koutras (1997b) and Antzoulakos and Chadjiconstantinidis (2001).

10.7.6 Further Distributions of Order k

Many other distributions of order k have been created; we do not have space to give formulas for them.

Hirano (1986) obtained a (shifted) Pólya–Aeppli distribution of order k and a Neyman type A distribution of order k using stopped-sum models (Chapter 9).

The emphasis in Panaretos and Xekalaki (1986a) was on urn models for distributions of order k. They gave one such model for Philippou's (1983) gamma-mixed Poisson of order k. Hypergeometric, inverse-hypergeometric, Pólya, and inverse-Pólya distributions of order k were called cluster hypergeometric distributions by Panaretos and Xekalaki (1986a, 1989). A cluster negative binomial, a cluster binomial, and a cluster generalized Waring distribution of order k appear in Panaretos and Xekalaki (1986a,b, and c, respectively). Xekalaki, Panaretos, and Philippou (1987) dealt with mixtures of distributions of order k. Certain distributions of order k arising by inverse sampling were studied by Xekalaki and Panaretos (1989).

New Pólya and inverse-Pólya distributions of order k were examined by Philippou, Tripsiannis, and Antzoulakos (1989). Philippou (1989) investigated Poisson-of-order-k mixtures of binomial, negative binomial, and Poisson distributions, and their interrelationships. Godbole (1990b) reexamined the hypergeometric and related distributions of order k of Panaretos and Xekalaki (1986a) and derived a waiting-time distribution of order k.

The papers by Aki (1985) and Hirano and Aki (1987) on multiparameter distributions of order k (also called *extended distributions of order k*) have already been mentioned. We draw attention also to those by Philippou (1988), Philippou and Antzoulakos (1990), and Ling (1990).

The book by Balakrishnan and Koutras (2002) discusses many more run-related models than has been possible here. It contains very many more references.

10.8 q-SERIES DISTRIBUTIONS

Much of the research concerning *basic hypergeometric series* (*q-series*) distributions has taken place during the last 12 years. The adoption of the Gasper and Rahman (1990) notation for basic hypergeometric series has clarified interrelationships between *q-series* distributions and has simplified their handling.

We remind readers that the new definition is

$$\begin{aligned} {}_A\phi_B&(a_1, \ldots, a_A; b_1, \ldots, b_B; q, z) \\ &= \sum_{j=0}^{\infty} \frac{(a_1; q)_j \ldots (a_A; q)_j z^j}{(b_1; q)_j \ldots (b_B; q)_j (q; q)_j} \left[(-1)^j q^{\binom{j}{2}}\right]^{B-A+1}, \end{aligned} \tag{10.114}$$

where $(a; q)_0 = 1$, $(a; q)_j = (1-a)\cdots(1-aq^{j-1})$, and $|q| < 1$. The modification $[(-1)^j q^{\binom{j}{2}}]^{B-A+1}$ leaves the function unaltered when $A = B + 1$. This definition is used here, in the third edition of this book, but it was not used in the second edition. Section 1.1.12 gives definitions of other q-entities such as Gaussian binomial coefficients.

When $q \to 1$, an ${}_A\phi_B[\cdot]$ series tends to an ${}_AF_B[\cdot]$ series; it is called a q-analog of the ${}_AF_B[\cdot]$ series. Similarly, a distribution involving ${}_A\phi_B[\cdot]$ is called a q-analog of the ${}_AF_B[\cdot]$ distribution. Usually there is more than one q-analog of a classical discrete distribution.

Reversing a q-series distribution that terminates at $X = n$, that is, transforming the variable to $W = n - X$, may markedly alter the distribution's properties. For example, reversing the distribution with pgf

$$G_D(z) = (\tau; q)_m \, {}_2\phi_1(q^{-m}, 0; q^{1-m}\tau^{-1}; q, qz), \qquad m \in \mathbb{Z}^+,$$

gives the distribution with pgf

$$G_{\text{DR}}(z) = \tau^m \, {}_2\phi_0\left(q^{-m}, \tau; -; q, \frac{q^m z}{\tau}\right).$$

Note the change in the expression for $\Pr[X = 0]$.

It has not proved helpful to classify basic hypergeometric distributions by the numbers (A and B) of numerator and denominator parameters. The section begins by considering distributions with finite support. These are followed by ones with infinite support. We then consider a few with bilateral support (on $\ldots, -2, -1, 0, 1, 2, \ldots$) and finally some that involve q-entities but not q-series directly.

10.8.1 Terminating Distributions

Absorption Distribution The name *absorption distribution* was given to the distribution that represents the number of individuals that fail to cross a specified region containing hazards of a certain kind (such as a minefield) by Blomqvist (1952), Borenius (1953), and Zacks and Goldfarb (1966). More recently Dunkl (1981) described the model in terms of a manuscript containing m errors. A proofreader reads through the manuscript line by line looking for errors. If he or she finds an error, he or she corrects it, returns to the beginning of the manuscript, and repeats the process. If he or she reaches the end without finding an error, he or she likewise goes back to the beginning and repeats the process. The probability of finding any particular error is assumed to be constant and equal to $1 - q$. Interest lies in the number of errors X that are found in n attempts to read through the manuscript. Kemp (1998) interpreted the distribution in terms of the capture of rare animals for an endangered species breeding program.

The sampling model implies that it is an analog of the binomial distribution. However, when $q \to 1$, it becomes a degenerate distribution with $G_A(z) = 1$. There are several q-binomial analogs; to distinguish it from these, it is sometimes called the *Dunkl q-binomial distribution.*

Blomqvist (1952) showed that

$$\begin{aligned}
&\Pr[X = 0] = q^{mn},\\
&\Pr[X = x] = q^{(m-x)(n-x)}\\
&\quad\times \frac{(1-q^m)(1-q^{m-1})\cdots(1-q^{m-x+1})(1-q^n)(1-q^{n-1})\cdots(1-q^{n-x+1})}{(1-q)(1-q^2)\cdots(1-q^x)},
\end{aligned} \tag{10.115}$$

where $x = 1, 2, \ldots, \min(m, n)$ and $0 < q < 1$. The symmetry of the pmf with respect to m and n is unexpected.

Borenius (1953) standardized the distribution using his approximations for the mean and variance. Plots of the resultant cdf indicated that when $q \geq 0.9$ his standardized distribution tended rapidly to normality as m and n increase.

Dunkl (1981) restated the pmf in terms of q-binomial coefficients. He made the assumption that $q^m = (1 - \theta/m)^m \doteqdot e^{-\theta} = c$. This enabled him to obtain much closer approximations for the mean and variance. He also considered inverse absorption sampling; see Section 10.8.2.

By stating the pgf in the form

$$G_A(z) = q^{mn}\,{}_2\phi_1(q^{-n}, q^{-m}; 0; q, qz), \qquad m, n \in \mathbb{Z}^+, \qquad 0 < q < 1, \tag{10.116}$$

Kemp (1998) was able to show that, when $n \le m$,

$$\mu = \sum_{j=1}^{n} \frac{(q^{-m}; q)_j (q^{-n}; q)_j}{1 - q^j} q^{j(m+n-j+1)},$$
$$\mu_2 = 2\sum_{j=2}^{n} \frac{(q^{-m}; q)_j (q^{-n}; q)_j}{1 - q^j} q^{j(m+n-j+1)} \sum_{u=1}^{j-1} \frac{1}{1 - q^u}. \qquad (10.117)$$

For $n > m$ the roles of m and n are reversed.

Newby (1999) adopted a different approach to the distribution, based on the use of two shift operators. He used these to develop recursions for the expected values of general functions of a variable with an absorption distribution and hence to obtain generating functions and exact moments for the distribution.

Assuming that $n \le m$, the reversed form of the absorption distribution has the pgf

$$\begin{aligned} G_{\text{AR}}(z) &= (q^{m+1-n}; q)_n \times {}_1\phi_1(q^{-n}; q^{m+1-n}; q, q^{m+1}z) \\ &= (q^{m+1-n}; q)_n \sum_{x=0}^{n} \frac{(1 - q^n) \cdots (1 - q^{n-x+1}) q^{x(m-n+x)} z^x}{(q^{m+1-n}; q)_x \ (q; q)_x}. \end{aligned}$$

Rawlings (1997, 2002) generalized the absorption distribution via a model in which j particles are propelled through a chamber m units long (a j particle is a sequence of j dots joined by a line $j - 1$ units long). Identical independent Bernoulli trials with parameter q decide whether or not the j particle moves forward by one unit. If it fails to move, it is absorbed. If the number of j particles is n and $\ell \ge jn$, then the number of absorptions has the pgf

$$\begin{aligned} G_{A^*}(z) &= (q^{\ell-j+1}; q^{-j})_n z^n \, {}_1\phi_1\left(q^{-jn}; q^{\ell-jn+1}; q^j, \frac{q^{\ell+1}}{z}\right) \\ &= (Q^{(\ell+1)/j-n}; Q)_n z^n \, {}_1\phi_1\left(Q^{-n}; Q^{(\ell+1)/j-n}; Q, \frac{Q^{(\ell+1)/j}}{z}\right), \end{aligned} \qquad (10.118)$$

where $q^j = Q$. Compared with $G_A(z) = z^n G_{\text{AR}}(1/z)$, we have $m + 1$ replaced by $(\ell + 1)/j$; the integer restriction on m has been removed.

In Kemp (2001b) the absorption distribution was characterized as the distribution of $U|(U + V = n)$, where n is constant, $n \le m$, and U and V are independent, iff U has the Kemp q-binomial distribution (see below) with pgf

$$G_{KB}(z) = \frac{{}_1\phi_0(q^{-n}; -; q, -\theta z)}{{}_1\phi_0(q^{-n}; -; q, -\theta)}$$

and V has a Heine distribution (see Section 10.8.2) with pgf

$$G_H(z) = \frac{{}_0\phi_0(-; -; q, -\lambda z)}{{}_0\phi_0(-; -; q, -\lambda)},$$

where $\theta = \lambda$. Kemp also showed that the distribution arises from (a) an equilibrium birth–death process and (b) a one-dimensional random walk in which the further that a particle is from the origin, the easier it is for the particle to move toward the origin and the more difficult it is for it to move away.

The probabilities of the absorption distribution are logconcave and therefore the distribution is IFR (increasing failure rate) and strongly unimodal (A. W. Kemp, 2002a).

Convolutions of Bernoulli rv's Kemp's (1987a) weapon defense model (Section 3.12.3) assumed a form of Poissonian binomial sampling in which the Bernoulli probabilities p_i, $i = 1, \ldots, n$, have the form $\ln p_i = \ln c + (i-1)\ln q$, that is, satisfy a *loglinear relationship*. The pgf for the number of hits out of n is

$$G_1(z) = \prod_{i=0}^{n-1}[1 + c\,q^i(z-1)] \tag{10.119}$$

$$= {}_1\phi_0(q^{-n}; -; q, q^n c\,(1-z)), \tag{10.120}$$

which is the pgf for a terminating q-series distribution. Because it is a convolution of Bernoulli pgf's, it can be expanded using Heine's theorem (Section 1.1.12); the resultant expressions for the probabilities are complicated, however. The factorial moment generating function is

$$G_1(1+t) = {}_1\phi_0(q^{-n}; -; q, -q^n c\,t),$$

giving the following simple expressions for the factorial moments:

$$\mu = \frac{(1-q^n)c}{1-q}, \qquad \mu_2 = \mu - \frac{(1-q^{2n})c^2}{1-q^2}. \tag{10.121}$$

As n becomes large, the distribution tends to an Euler distribution (Section 10.8.2).

A. W. Kemp and C. D. Kemp's (1991) Poissonian binomial sampling model for Weldon's dice data assumed $p_i = cq^{i-1}/(1+cq^{i-1})$, that is, a *loglinear-odds relationship*, $\ln[p_i/(1-p_i)] = \ln c + (i-1)\ln q$. It gives rise to the *Kemp q-binomial distribution*. Kemp and Kemp suggested that it would be a realistic model for multiple-channel production processes such as plastic intrusion molding. It also arises from a stationary stochastic process; see Kemp and Newton (1990), who were interested in a situation involving a dichotomy between parasites on hosts with and without open wounds resulting from previous parasite attacks. Jing and Fan (1994) examined the distribution as a possibility for modeling the q-deformed binomial state in quantum physics. A. W. Kemp (2002a) gave a random-walk model.

The pgf is

$$G_{KB}(z) = \prod_{i=0}^{n-1}\frac{1+cq^i z}{1+cq^i} \tag{10.122}$$

$$= \frac{{}_1\phi_0(q^{-n}; -; q, -cq^n z)}{{}_1\phi_0(q^{-n}; -; q, -cq^n)}. \tag{10.123}$$

As $q \to 1$, (10.119) tends to (3.3) with $p = c$ and (10.122) tends to (3.2) with $p = c$, showing that both convolutions of Bernoulli rv's tend to the binomial distribution. Also as $n \to \infty$ the Kemp q-binomial distribution tends to a Heine distribution (Section 10.8.2).

From Heine's theorem

$$\Pr[X_{\mathrm{KB}} = 0] = \left(\prod_{i=0}^{n-1}(1+cq^i)\right)^{-1},$$

$$\Pr[X_{\mathrm{KB}} = x] = \frac{(1-q^n)\cdots(1-q^{n-x+1})}{(1-q)\cdots(1-q^x)}q^{x(x-1)/2}(cz)^x \times \Pr[X_{\mathrm{KB}} = 0]. \tag{10.124}$$

The cumulants are the sums of the individual Bernoulli cumulants and hence

$$\mu = \kappa_1 = \sum_{i=0}^{n-1}\frac{cq^i}{1+cq^i} \qquad \text{and} \qquad \mu_2 = \kappa_2 = \sum_{i=0}^{n-1}\frac{cq^i}{(1+cq^i)^2}.$$

Rao and Shanbhag (1994) proved a far-reaching theorem related to the Kemp q-binomial distribution; see also Shanbhag and Kapoor (1993). In particular, it characterizes the Kemp q-binomial distribution as the distribution of $U|(U + V = n)$, where U has a Heine and V has an independent Euler distribution; see Section 10.8.2 concerning the Heine and Euler distributions.

The probabilities of the Kemp q-binomial distribution, like those of the absorption distribution, are logconcave and hence the distribution is IFR (increasing failure rate) and strongly unimodal. The reversed distribution has the same form as the unreversed distribution (A. W. Kemp 2002a).

q-Deformed Binomial Distribution The q-deformed oscillator and its applications have excited considerable interest in research into subatomic physics. Jing (1994) wanted to create a q-analog of the binomial distribution in order to extend the concept of the ordinary binomial state to the q-deformed binomial state. He called his distribution the *q-deformed binomial distribution*; see also Jing and Fan (1994). It has the pmf

$$\Pr[X_{\mathrm{QD}} = x] = \frac{(q;q)_m(\tau;q)_{m-x}\tau^x}{(q;q)_x(q;q)_{m-x}}, \qquad x = 0, 1, \ldots, \tag{10.125}$$

where $0 < q < 1$, $0 < \tau < 1$. The pgf for the distribution is

$$G_{\mathrm{QD}}(z) = (\tau;q)_m\, {}_2\phi_1\left(q^{-m}, 0; \frac{q^{1-m}}{\tau}; q; qz\right);$$

for the reversed distribution it is

$$G_{\mathrm{QDR}}(z) = \tau^m\, {}_2\phi_0\left(q^{-m}, \tau; -; q; \frac{q^m z}{\tau}\right).$$

The distribution is logconcave for $\tau \le q$; when $\tau > q$, there is the possibility that it is neither concave nor convex. As $m \to \infty$, it tends to the Euler distribution (Section 10.8.2).

Jing (1994) approached the moment properties of the distribution by considering $E[q^{rx}]$. The usual uncorrected moments $E[x^r]$ are not known.

Kemp (2003) characterized the distribution as that of $U|(U + V = n)$, where U has an Euler distribution and V has an independent Dunkl q-binomial distribution.

Other q-Binomial Analogs Crippa and Simon (1997) studied stochastic processes leading to terminating q-series distributions. Let $P_{n,\ell}$ be the probability that a process is in state ℓ at epoch n with $P_{1,1} = 1$. The Markov chains that they studied have the very general recursive relationship

$$P_{n,\ell} = (1 - \lambda_{n,\ell})P_{n-1,\ell} + \lambda_{n,\ell-1}P_{n-1,\ell-1} \quad \text{for } n \ge 2,$$

where (i) $\lambda_{n,\ell} = q^{\alpha\ell+\beta n+\gamma}$ and (ii) $\lambda_{n,\ell} = 1 - q^{\alpha\ell+\beta n+\gamma}$ and α, β, and γ are such that $0 \le \lambda_{n,\ell} \le 1$ for all $0 \le \ell \le n$. They obtained expressions for the pgf's and the first two moments and gave some graph-theoretical models for these distributions.

A. W. Kemp (2002a, 2003) investigated four further q-binomial analogs. The Kemp q-binomial pmf (10.124) can be reformulated as

$$\Pr[X_{\text{KB}} = x] = C_{\text{KB}} \begin{bmatrix} n \\ x \end{bmatrix}_q q^{x(x-1)/2}\theta^x, \tag{10.126}$$

where C_{KB} is the normalizing constant and $\left[\begin{smallmatrix} n \\ x \end{smallmatrix}\right]_q$ is a Gaussian (basic) binomial coefficient (see Section 1.1.12). Weighting (10.126) with the weighting factor $q^{x(1-x)/2}$ gives a distribution that Kemp called the *Rogers–Szegö distribution* (R-SD) because the pgf is a Rogers–Szegö polynomial (Andrews, 1998). Weighting (10.126) with the weighting factor $q^{x(x-1)/2}$ gives Kemp's *Stieltjes–Wigert distribution* (S-WD), so-called because the pgf is a Stieltjes–Wigert polynomial (Gasper and Rahman, 1990).

Both distributions retain the same form when reversed. Their rv's can be shown to be conditional rv's of the kind $(U|(U + V = n))$, where U and V are (i) both Euler rv's (R-SD) and (ii) both Heine rv's (S-WD). Both distributions are logconcave, IFR, and strongly unimodal. Kemp (2003) fitted the R-SD to overdispersed data on student absences and the S-WD to very underdispersed word count data, with very good results compared with binomial fits.

The other two q-binomial analogs in A. W. Kemp (2002a, 2003) are the *Jackson distribution* [named after F. H. Jackson, who wrote a very long series of papers (1904–1954) on the theory of q-series] and the *Wall distribution*, for which the pgf is a Wall polynomial (Gasper and Rahman, 1990).

In both cases the pgf has a different form when reversed, both are logconcave for only certain combinations of their parameters, and both rv's can be interpreted as conditional $(U|(U + V = n))$ rv's, where U and V have discrete q-series distributions.

10.8.2 q-Series Distributions with Infinite Support

Euler and Heine Distributions The simplest nonterminating q-series distribution appeared almost simultaneously in (i) the physics literature as Biedenhahn's (1989) *q-Poisson energy distribution* in the theory of the quantum harmonic oscillator and (ii) the statistics literature as Benkherouf and Bather's (1988) *Euler distribution* which they obtained as a feasible prior distribution for the number of undiscovered sources of oil; see also Macfarlane (1989). Benkherouf and Bather obtained an alternative distribution which they called the *Heine distribution*. As $q \to 1$, both distributions tend to the Poisson distribution (Kemp, 1992a,b), so they are both q-Poisson analogs.

The pgf of the Euler distribution is

$$G_E(z) = \frac{{}_1\phi_0(0; -; q, \eta z)}{{}_1\phi_0(0; -; q, \eta)} = \prod_{j=0}^{\infty}\left(\frac{1-\eta q^j}{1-\eta q^j z}\right), \qquad 0 < q < 1, \qquad 0 < \eta < 1, \tag{10.127}$$

that is, an infinite convolution of geometric pgf's. Its pmf is

$$\Pr[X_E = x] = \frac{\eta^x}{(q;q)_x}\Pr[X_E = 0], \quad \text{where } \Pr[X_E = 0] = \prod_{j=0}^{\infty}(1-\eta q^j),$$

and the mean and variance are

$$\mu = \sum_{x=0}^{\infty}\left(\frac{\eta q^x}{1-\eta q^x}\right), \qquad \mu_2 = \sum_{x=0}^{\infty}\left(\frac{\eta q^x}{(1-\eta q^x)^2}\right). \tag{10.128}$$

For the Heine distribution, on the other hand, the pgf is

$$G_H(z) = \frac{{}_0\phi_0(-; -; q, -\lambda z)}{{}_0\phi_0(-; -; q, -\lambda)} = \prod_{j=0}^{\infty}\left(\frac{1+\lambda q^j z}{1+\lambda q^j}\right), \qquad 0 < q < 1, \qquad 0 < \lambda. \tag{10.129}$$

This is an infinite convolution of Bernoulli pgf's. The pmf is

$$\Pr[X_H = x] = \frac{\eta^x q^{x(x-1)/2}}{(q;q)_x}\Pr[X_H = 0], \quad \text{where } \Pr[X_H = 0] = \prod_{j=0}^{\infty}(1-\eta q^j)^{-1},$$

and the mean and variance are

$$\mu = \sum_{x=0}^{\infty}\left(\frac{\lambda q^x}{1+\lambda q^x}\right), \qquad \mu_2 = \sum_{x=0}^{\infty}\left(\frac{\lambda q^x}{(1+\lambda q^x)^2}\right). \tag{10.130}$$

Kemp (1992a) studied the pgf's, moments, cumulants, and other properties, as well as maximum-likelihood and other methods of estimation, for both distributions. Both have infinite support, are unimodal, and have increasing failure rates. The Euler distribution is overdispersed and infinitely divisible, whereas the Heine distribution is underdispersed and not infinitely divisible. Kemp found, moreover, that $-1 < q < 0$, $0 < \eta < 1$ in (10.127) gives rise to a third member of this family. She called it the pseudo-Euler distribution; it is infinitely divisible and overdispersed but, unlike the others, it can be multimodal.

Kemp (1992b) investigated steady-state Markov chain models for the Heine and Euler distributions; these include current-age models for discrete renewal processes, success-runs processes with nonzero probabilities that a trial is abandoned, Foster processes, and equilibrium random walks corresponding to M/M/1 queues with elective customer behavior.

[Note that this account of the Euler and Heine distributions differs from that in the second edition of this book. Here we use the modern Gasper and Rahman (1990) notation for basic hypergeometric series. Also certain errors in the previous edition have been corrected.]

Inverse-Absorption Distribution The Heine pgf is an infinite convolution of Bernoulli pgf's; it is therefore an obvious limiting form of the Kemp q-binomial distribution. Dunkl (1981) commented that there is a q-analog for the negative binomial distribution where the pgf is a finite convolution of geometric pgf's; the Euler distribution is its limiting form. Dunkl obtained it as the distribution of Y, the number of attempts to read through a manuscript in order to detect exactly k of the m errors that it contains (each time that an error is found the proofreader starts again at the beginning of the manuscript). Clearly $Y \geq k$.

This *inverse-absorption distribution* has the pgf

$$G_{\mathrm{IA}}(z) = z^k \prod_{i=1}^{k} \frac{1 - q^{m-i+1}}{1 - q^{m-i+1}z} = z^k \frac{{}_1\phi_0(q^k; -; q, q^{m-k+1}z)}{{}_1\phi_0(q^k; -; q, q^{m-k+1})}, \qquad (10.131)$$

where $m \in \mathbb{Z}^+$, $k \in \mathbb{Z}^+$, $0 < q < 1$. The pmf is

$$\Pr[Y = k + x] = \frac{(1 - q^k) \cdots (1 - q^{k+x-1}) q^{(m-k+1)x}}{(1 - q) \cdots (1 - q^x)}, \qquad x = 0, 1, \ldots,$$

where $k = 0, 1, 2, \ldots$ and $y \geq k$.

The cumulants are the sums of the corresponding geometric cumulants, and $\mu = \kappa_1, \;\; \mu_2 = \kappa_2$.

Newby (1999) investigated further the moments for this inverse-absorption analog of the negative binomial distribution.

Rawlings (1997) studied the inverse sampling distribution arising from his generalized absorption model [for which the direct sampling pgf is (10.118)]. Given a chamber of length $\ell \geq j$, let N be the number of j particles needed to obtain k absorptions, where $k \le \ell/j$. If $0 < q < 1$, then the pgf for n is

$$\begin{aligned} G_{\mathrm{IA}^*}(z) &= (q^{\ell-j+1}; q^{-j})_k z^k \, {}_1\phi_0(q^{jk}; -; q^j, q^{\ell-jk+1} z) \\ &= (Q^{(\ell+1)/j-k}; Q)_k z^k \, {}_1\phi_0(Q^k; -; Q, Q^{(\ell+1)/j-k} z) \\ &= \frac{z^k \, {}_1\phi_0(Q^k; -; Q, Q^{(\ell+1)/j-k} z)}{{}_1\phi_0(Q^k; -; Q, Q^{(\ell+1)/j-k})}. \end{aligned} \tag{10.132}$$

Comparison with $G_{\mathrm{IA}}(z)$ shows that the integer restriction on m can again be relaxed.

Generalized Euler Distributions Benkherouf and Alzaid (1993) proposed a more general model for oil exploration than the Euler and Heine models of Benkherouf and Bather (1988). It has the pmf

$$\begin{aligned} \Pr[X = x] &= \frac{(1 - aq^{x-1})\lambda}{1 - q^x} \Pr[X = x - 1] \\ &= \frac{(a; q)_x \lambda^x}{(q; q)_x} \Pr[X = 0], \qquad x = 0, 1, \ldots, \end{aligned}$$

where $0 < q < 1$, $0 \le a < 1$, $0 < \lambda < 1$. The pgf is

$$G(z) = \prod_{j=1}^{\infty} \left(\frac{1 - a\lambda q^j z}{1 - a\lambda q^j} \right) \left(\frac{1 - \lambda q^j}{1 - \lambda q^j z} \right) = \frac{{}_1\phi_0(a; -; q, \lambda z)}{{}_1\phi_0(a; -; q, \lambda)}. \tag{10.133}$$

Benkherouf and Alzaid found that *this generalized Euler distribution* is unimodal, infinitely divisible, and IFR when $a < q$. It is an infinite convolution of modified geometric distributions. The inverse absorption is the special case $a = q^k$, $\lambda = q^{m-k+1}$.

A. W. Kemp (2002c) relaxed the conditions $0 \le a < 1$ and $0 < \lambda < 1$ and made a thorough study of the existence conditions for the resultant family of distributions. It includes q-analogs of the binomial, Poisson, negative binomial, and other distributions. For different members of the family their logconvexity, logconcavity, over- and undispersion, failure rate behavior, strong unimodality, and infinite divisibility properties were explored using the methods of Gupta, Gupta, and Tripathi (1997).

Benkherouf's (1996) short paper looked at a generalization of the Heine distribution which can serve as a conjugate prior for modeling the number of undiscovered oil fields in a Bayesian search model.

Subcritical Branching Process with Geometric Offspring Given a branching process with Bernoulli offspring in the first generation, $g_1(z) = 1 - p + pz$, $0 < p < 1$, then the number of offspring in the nth generation has the pgf $g_n(z) = g_1(g_{n-1}(z)) = 1 - p^n + p^n z$. If T is the time to extinction, then

$$\Pr[T = n] = g_n(0) - g_{n-1}(0) = (1 - p)p^{n-1}, \qquad n = 1, 2, \ldots,$$

and the pgf for time to extinction is geometric with $G_T(z) = z(1-p)/(1-pz)$; extinction is certain since $m = [dg_1(z)/dz]_{z=1} = p < 1$.

Suppose instead that the number of first-generation offspring is geometric with $g_1(z) = (1+q-qz)^{-1}$. If $0 < q < 1$, then $m = [dg_1(z)/dz]_{z=1} = q < 1$, the branching process is subcritical, and extinction is certain to occur. The pgf for the number of offspring in the nth generation is

$$g_n(z) = \frac{(1-q^n) - qz(1-q^{n-1})}{(1-q^{n+1}) - qz(1-q^n)}, \qquad n = 1, 2, \ldots,$$

compare Feller (1968), who used a notation with $q \to p^*/(1-p^*)$, $q^* = 1 - p^*$. Farrington and Grant (1999) showed that

$$\Pr[T = n] = g_n(0) - g_{n-1}(0) = \frac{(1-q)^2 q^{n-1}}{(1-q^n)(1-q^{n+1})}, \qquad n = 1, 2, \ldots,$$

and the pgf for time to extinction is

$$G_T(z) = \frac{z}{1+q}\left[1 + \frac{(1-q)qz}{1-q^3} + \frac{(1-q)(1-q^2)q^2z^2}{(1-q^3)(1-q^4)} + \cdots\right].$$

This is a q-series pgf which can be restated as

$$G_T(z) = \frac{z}{1+q}\, {}_2\phi_1(q, q; q^3; q, qz). \tag{10.134}$$

Farrington and Grant applied the results of their studies on time to extinction for subcritical branching processes with Bernoulli, geometric, and Poisson offspring distributions to the spread of infection in highly vaccinated populations, outbreaks of enteric fever, and person-to-person transmission of human monkeypox.

q-Confluent Hypergeometric Distributions C. D. Kemp (2002) derived two analogs of the hyper-Poisson distribution by truncating the Heine and Euler distributions at the rth probability, normalizing the tail, and shifting the resultant distributions to support $0, 1, \ldots$. He then removed the constraint $r \in \mathbb{Z}^+$ and obtained the distributions with pgf's

$$G_{\text{qHP1}}(z) = \frac{{}_1\phi_1(q; q^c; q, -(1-q)\theta z)}{{}_1\phi_1(q; q^c; q, -(1-q)\theta)}, \qquad 0 < c, \qquad 0 < \theta, \tag{10.135}$$

$$G_{\text{qHP2}}(z) = \frac{{}_2\phi_1(q, 0; q^c; q, -(1-q)\theta z)}{{}_2\phi_1(q, 0; q^c; q, -(1-q)\theta)}, \qquad 0 < c, \qquad 0 < \theta < 1, \tag{10.136}$$

respectively. He found that both are logconcave and studied their modal properties. His paper gives mixing and group size models for both distributions.

The first (qHP1) is a member of a family of q-confluent hypergeometric distributions; see A. W. Kemp (2005). This includes a third q-Poisson analog, the Morse (1958) balking distribution, and other equilibrium distributions for

M/M/1 queues with various kinds of balking. The family also contains the *O/U distribution*, so-called because it is overdispersed for $0.5+\alpha < q < 1$ and underdispersed for $0 < q < 0.5+\alpha$, $0 < \alpha < 0.5$ (the value of α giving equidispersion increases as θ increases).

Kemp (2004b) also gave success runs models for q-confluent distributions.

Other Nonterminating q-Hypergeometric Distributions There are at least two *q-logarithmic analogs*. The first has the pgf

$$G_1(z) = z\frac{{}_2\phi_1(q, q; q^2; q, \theta z)}{{}_2\phi_1(q, q; q^2; q, \theta)}, \qquad 0 < q < 1, \qquad 0 < \theta < 1. \tag{10.137}$$

Its properties and models for the distribution were explored by C. D. Kemp (1997). As $q \to 1$, it tends to the usual logarithmic distribution.

The second q-logarithmic analog arises as the cluster size distribution for the (infinitely divisible) Euler distribution. Its pgf cannot be expressed straightforwardly as a q-hypergeometric series; see Section 10.8.4.

Kemp (2001a) revisited Weinberg and Gladen's (1986) analysis of a retrospective study concerning the number of cycles to conception. In a retrospective study the number of sterile couples who will never conceive is necessarily zero. For a prospective study this is not so; sterile couples will create infinite data values. Replacing Weinberg and Gladen's beta-geometric distribution with a *q-beta-geometric distribution* with pmf

$$\begin{aligned} \Pr[X=1] &= \frac{1-\rho}{1-\rho/u}, \\ \Pr[X=x] &= \frac{(1-\rho)(1-1/u)(1-m/u)\cdots(1-m^{x-2}/u)\rho^{x-1}}{(1-\rho/u)(1-m\rho/u)(1-m^2\rho/u)\cdots(1-m^{x-1}\rho/u)} \end{aligned} \tag{10.138}$$

for $x = 2, 3, \ldots$, created a model that is suitable for prospective studies as well as retrospective ones. The new fits resembled the previous ones; the data, however, are thought to contain misrecorded values.

The *Bailey–Daum distribution* (A. W. Kemp 2002b) arises from a random-walk model; it provides a probabilistic proof of the Bailey–Daum theorem (Gasper and Rahman, 1990).

10.8.3 Bilateral q-Series Distributions

Bilateral q-series distributions have support $\ldots, -2, -1, 0, 1, 2, \ldots$. Their pgf's may be expressible in terms of the bilateral basic hypergeometric function

$$\begin{aligned} &{}_r\psi_s(a_1, \ldots, a_r; b_1, \ldots, b_s; q, z) \\ &\quad = \sum_{j=-\infty}^{\infty} \frac{(a_1;q)_j \ldots (a_r;q)_j z^j}{(b_1;q)_j \ldots (b_s;q)_j}(-1)^{(s-r)j}\, q^{(s-r)j(j-1)/2}\, z^j; \end{aligned}$$

see Section 1.1.12.

Maximum-entropy distributions (MEDs) have maximum entropy subject to specific constraints. Usually Shannon's entropy is used (this entropy measure is $H_1 = -\sum_x p_x \ln p_x$, where $p_x = \Pr[X = x]$). It is a concave function which can be maximized subject to linear constraints by Lagrange's method; see Kapur (1989).

Lisman and van Zuylen (1972) obtained "the most probable distribution" (the MED) with support $\ldots, -2, -1, 0, 1, 2, \ldots$, given $E[(X-a)^2]$ (i.e., specified variance about a known point), but they thought the distribution intractable.

Kemp (1997a) proved that the discrete MED with specified mean μ and variance σ^2 has the pmf

$$\Pr[X = x] = \frac{\lambda^x q^{x(x-1)/2}}{\sum_{x=-\infty}^{\infty} \lambda^x q^{x(x-1)/2}}, \qquad x = \ldots, -2, -1, 0, 1, 2, \ldots. \quad (10.139)$$

Dasgupta (1993) had previously obtained the one-parameter distribution with $\lambda = q^2$, $q = \exp(-2\beta)$. The normal distribution is the continuous MED for given mean and given variance. It is appropriate to call (10.139) the *discrete normal distribution*, not to be confused with the discrete normal distributions of Bowman, Shenton, and Kastenbaum (1991) and Roy (2003). It is analogous to the normal distribution in that it is the only discrete distribution on $(-\infty, \infty)$ for which the maximum-likelihood equations are the first two moment equations. It is logconcave, unimodal, and IFR. Kemp (1997a) characterized the distribution as the distribution of the difference of two independent Heine rv's. She also obtained expressions for the cumulants.

Liang (1999) constructed monotone empirical Bayes tests for the distribution. Szabłowski (2001) gave the natural reparameterization $q \to \exp(-1/\sigma^2)$, $\lambda \to \exp(-1/2\sigma^2 + \alpha/\sigma^2)$, whence

$$\Pr[X = x] = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\alpha)^2}{2\sigma^2}\right\} \Big/ \hat{U}(\alpha, \sigma^2), \quad x = \ldots, -1, 0, 1, \ldots,$$

where $\hat{U}(\alpha, \sigma^2)$ is a normalizing constant. This enabled Szabłowski to obtain simple approximate formulas for the mean and variance and to show that the sum of two discrete normal distributions is approximately a discrete normal distribution with parameters $(\alpha_1 + \alpha_2, \sigma_1^2 + \sigma_2^2)$. He also discussed the relationships between the distribution and Jacobi theta functions.

Kemp (1997b) reexamined the *difference of two Heine variables* (the *discrete normal distribution*) and also studied the *difference of two Euler variables*. She formulated their pmf's in terms of modified q-Bessel functions $J_\nu^{(1)}(z; q)$ and $J_\nu^{(2)}(z; q)$; see Gasper and Rahman (1990) and Kemp (1997b) for definitions. Both distributions were found to be logconcave, unimodal, and IFR. Their factorial cumulant generating functions were obtained and hence their means and variances. She also obtained formulas for the difference of two generalized Euler rv's. Special cases of these have pgf's such as

$$G_a(z) = \frac{{}_1\psi_1(A; qB^{-1}; q, \lambda z)}{{}_1\psi_1(A; qB^{-1}; q, \lambda)},$$

$$G_b(z) = \frac{{}_1\psi_1(A; 0; q, \lambda z)}{{}_1\psi_1(A; 0; q, \lambda)},$$

$$G_c(z) = \frac{{}_0\psi_1(-; qB^{-1}; q, -\alpha z)}{{}_0\psi_1(-; qB^{-1}; q, -\alpha)};$$

these yield summation formulas for certain bilateral q-hypergeometric series.

10.8.4 q-Series Related Distributions

Consider the $n!$ equiprobable permutations of the numbers $1, 2, \ldots, n$ and let Q be the number of inversions in one of the permutations. For example the permutation 3,2,1,4 of the numbers 1,2,3,4 is the outcome of $Q = 3$ inversions, $3 \rightleftharpoons 2$, $3 \rightleftharpoons 1$, $2 \rightleftharpoons 1$. Given $n = 4$, the rv Q can take the values $0, 1, \ldots, 6$.

Pólya (1970) [see also Kendall and Stuart (1979)] showed that the pgf for the *number of inversions* in the equiprobable permutations of the numbers $1, 2, \ldots, n$ has the form

$$G(q; n) = (n!)^{-1} \prod_{j=1}^{n} \left(\frac{1 - q^n}{1 - q} \right) = \frac{(q; q)_n}{n!(1 - q)^n}, \tag{10.140}$$

where q is the generating variable. The cumulants for the distribution are

$$\kappa_1 = \frac{n(n-1)}{4}, \qquad k_{2j} = \frac{B_{2j}}{2j} \left(\sum_{i=1}^{n} i^{2j} - n \right), \qquad j \geq 1,$$

$$k_{2j+1} = 0,$$

where B_{2j} is a nonzero, even-order Bernoulli number (Section 1.1.9).

The distribution studied by Di Bucchianico (1999) and Prodinger (2004) also involves the use of a q-entity. Di Bucchianico obtained new combinatorial proofs of the recurrences satisfied by the *null distribution of the Wilcoxon–Mann-Whitney test statistic* and hence found that the pgf is

$$\begin{aligned} G(q) &= \frac{\begin{bmatrix} m+n \\ m \end{bmatrix}_q}{\begin{pmatrix} m+n \\ m \end{pmatrix}} \\ &= \frac{m!n!(1-q)(1-q^2)\cdots(1-q^{m+n})}{(m+n)!(1-q)(1-q^2)\cdots(1-q^m)(1-q)(1-q^2)\cdots(1-q^n)}, \end{aligned} \tag{10.141}$$

where $m \in \mathbb{Z}^+$, $n \in \mathbb{Z}^+$, $0 < q < 1$, q is the generating variable, and $\begin{bmatrix} m+n \\ m \end{bmatrix}_q$ is a Gaussian binomial coefficient; see Section 1.1.12. He discussed ways to compute the moments of the distribution.

Prodinger realized that

$$G(q) = \frac{g_1(q) \cdots g_{m+n}(q)}{g_1(q) \cdots g_m(q)\, g_1(q) \cdots g_n(q)},$$

where $g_k(q) = k^{-1}(1 - q^k)/(1 - q)$ is a pgf (of a discrete rectangular distribution) and hence obtained the distribution's cumulants from the discrete rectangular cumulants. His paper gives general expressions for the cumulants, also formulas for the first three uncorrected moments. In particular,

$$\begin{aligned}
\mu &= \kappa_1 = \tfrac{1}{2} mn, \\
\mu_2 &= \kappa_2 = \tfrac{1}{12}[mn(m + n + 1)], \\
\mu_3 &= \kappa_3 = \tfrac{1}{120}\left\{mn(m + n + 1)[m(m + 1) + mn + n(n + 1)]\right\}.
\end{aligned}$$

The pgf of the *second q-logarithmic distribution* (Section 10.8.2) cannot be expressed as a q-hypergeometric series satisfying the definition (10.114) (although its derivative with respect to z is more tractable). The Euler distribution (see the previous section) is infinitely divisible; it therefore has a pgf of the form

$$G_E(z) = \exp\{\lambda[g(z) - 1]\}.$$

Given a Poisson distribution for the number of clusters, an Euler distribution is the outcome when the cluster size distribution has the pgf

$$g(z) = \frac{\sum_{i=1}^{\infty} \left\{a^i z^i/[i(1 - q^i)]\right\}}{\sum_{i=1}^{\infty} \left\{a^i/[i(1 - q^i)]\right\}}, \qquad 0 < q < 1, \qquad 0 < a < 1. \tag{10.142}$$

C. D. Kemp (2004) has commented that the cluster size distribution tends to the usual logarithmic distribution when $q \to 0$ (not when $q \to 1$) and has compared its properties with those of (10.137).

CHAPTER 11

Parametric Regression Models and Miscellanea

11.1 PARAMETRIC REGRESSION MODELS

11.1.1 Introduction

The increasing availability of cheap computer memory has led to the use of computers in the collection and analysis of large data sets. Count data regression is concerned with the effects of explanatory variables on a discrete rv with nonnegative support. Winkelmann (2000, Chapter 7) gives bibliographic details concerning a large number of discrete regression applications in areas as diverse as accidents, health economics, demography, marketing and management, and labor mobility.

Univariate Poisson regression assumes that the response (endogenous) variable has a Poisson(θ) distribution and seeks to describe the impact of the explanatory (exogenous) variables on the mean:

$$\theta = \exp(\mathbf{X}'\boldsymbol{\beta}). \tag{11.1}$$

It has been used to analyze, for example,

- time-series data on numbers of bank failures (Davutyan, 1989)
- the effect of actions by the firm and by federal regulators on takeover bids (Jaggia and Thosar, 1993; Cameron and Trivedi, 1998)
- the effect of temperature on sudden infant deaths (SIDS) (Campbell, 1994)
- the Common Birds Census data (Morgan, 2000)
- mortality from coronary heart disease (Dobson, 2002).

Lindsey (1996, Chapter 9) gives a worked example concerning the number of people recalling a stressful event and the number of months since its occurrence.

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

Much of the underlying theory has been developed by McCullagh and Nelder (1983) and Jørgensen (1997); Seeber (1998) gives an elementary account.

The assumption of iid endogenous variables is often not justified, however. Moreover a very common feature of biometric and biostatistical data is overdispersion—Poisson regression does not allow for this.

Much effort has been put into the development of alternative models: nonparametric, semiparametric, and parametric. This has led to a large literature in biometry, biostatistics, and econometrics; see, for example, Morgan (2000), Everitt (2003), Cameron and Trivedi (1996, 1998), Winkelmann and Zimmerman (1995), and Winkelmann (2000). Nonparametric models, such as the linear regression model, make minimal assumptions about an underlying probability model, whereas a parametric model, such as negative binomial regression, is probabilistically fully specified. Semiparametric models assume a partial parametric probability model but also incorporate linear or other functions with unknown coefficients (see, e.g., Sasieni, 1998). In this book we are interested almost exclusively in fully parametric models.

Much use has been made of the discrete univariate exponential family of distributions; these have a probability mass function (pmf) of the form

$$\Pr[Y = y; \theta] = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right]. \tag{11.2}$$

If $a(y) = y$, the distribution is said to be in canonical (standard) form; $b(\theta)$ is called the natural parameter of the distribution. Many well-known distributions belong to this family.

For the Poisson distribution, $\Pr[Y = y; \theta] = \exp[y \ln(\theta) - \theta - \ln(y!)]$; this is in canonical form and the natural parameter is $b(\theta) = \ln(\theta)$. The Poisson loglinear regression model (11.1) assumes that the link function $\ln(\theta)$ is a linear function of the explanatory variables (Lindsey, 1996, Section 1.3).

The binomial pmf can be rewritten as

$$\Pr[Y = y; \pi] = \exp\left[y \ln \pi - y \ln(1 - \pi) + n \ln(1 - \pi) + \ln\binom{n}{y}\right], \tag{11.3}$$

where $a(y) = y$ and $b(\pi) = \ln[\pi/(1 - \pi)]$. In logistic regression (frequently used for binary-type data with support $0, 1, 2, \ldots, n$, n known) the link function $\ln[\pi/(1 - \pi)]$ is assumed to be a linear function of the explanatory variables.

However $\mu_2 = \mu$ for the Poisson distribution and $\mu_2 = \mu(1 - \mu/n)$ for the binomial. Very often count and binary data are overdispersed, that is, they have variances that are greater than can be accounted for by these models. For count data this has led to the use of mixed Poisson regression models belonging to the Tweedie–Poisson family; see Section 11.1.2. It includes

the Poisson $\bigwedge$ Gamma $\equiv$ negative binomial model (see Section 11.1.3)

the Poisson $\bigwedge$ Lognormal model (see Section 11.1.4)

the Poisson $\bigwedge$ Inverse Gaussian model (see Section 11.1.5).

The normalization coefficients for some of these distributions do not have closed forms; this necessitates the use of computer-intensive methods.

Other count regression models that have been used include finite mixtures of Poisson distributions (Section 8.2.5; see McLachlan and Peel, 2000); Consul's generalized Poisson distribution (Section 7.2.6; see Consul and Famoye, 1992; Santos Silva, 1997); and the Lagrangian generalized negative binomial distribution (Section 7.2.7; see Consul and Famoye, 1995; Famoye, 1995).

The zero-truncated Poisson is widely used for zero-absent count data. Zero-altered regression models have been studied by Heilbron (1994) and Böhning (1998). Hurdle formulations (Section 8.2.3) have been found advantageous compared with zero-modified formulations as the zero frequency can be regressed separately from the remaining frequencies.

Two novel count regression models are the low-order Poisson polynomial model of Cameron and Johansson (1997; see Section 11.1.6) and the weighted Poisson model of Castillo and Pérez-Casany (1998) which permits both over- and underdispersion (see Section 11.1.7).

Another parametric approach is the use of Efron's (1986) double-exponential family. Here the existence of the second parameter, ψ, in

$$\Pr[Y = y; \theta, \psi] = c(\theta, \psi)\sqrt{\psi}\, f(y; \theta)^{\psi}[f(y; y)]^{1-\psi} \tag{11.4}$$

[where $c(\theta, \psi)$ is a normalizing constant] allows for control of the variance independently of the mean; see Section 11.1.8. Both over- and underdispersion are possibilities. There are double-Poisson and double-binomial members of this family.

The binomial $\bigwedge$ beta distribution (Section 6.9.2) is a popular choice for binary data that are overdispersed relative to the binomial distribution. Brooks, Morgan, Ridout, and Pack (1997) have compared the fits of this distribution with fits by (i) a mixture of the beta–binomial and a binomial, (ii) a mixture of two binomials, (iii) Paul's (1985, 1987) beta–correlated binomial, and (iv) the correlated-binomial distribution (Section 3.12.6), to data on the numbers of dead implants in treated and untreated mice. Finite binomial mixtures have also been studied by McLachlan and Peel (2000). A further possibility is the simplex–binomial model; see Section 11.1.9.

11.1.2 Tweedie–Poisson Family

This family is closely related to the generalized Pólya–Aeppli (Poisson–Pascal) distribution; see Sections 8.3.3 and 9.8.

A Tweedie–Poisson distribution is a mixture of Poisson distributions where the mixing distribution belongs to the Tweedie family of exponential dispersion distributions. The Laplace transform of a Tweedie(α, δ, θ) distribution is

$$E[e^{-sX}] = \int_0^{\infty} e^{-sx} f(x)\, dx = \exp\left[\frac{\delta}{\alpha}[\theta^{\alpha} - (\theta + s)^{\alpha}]\right], \tag{11.5}$$

where $\alpha \le 1, 0 < \delta, 0 \le \theta$; it is a continuous distribution that was first introduced into the literature by Tweedie (1947, 1984) and independently by Nelder and Wedderburn (1972). It has since been studied extensively (see, e.g., Bar Lev and Enis, 1986; Hougaard, 1986; Jørgensen, 1997).

From (11.5), the probability generating function (pgf) of a Poisson(λ) $\bigwedge_{\lambda}$ Tweedie(α, δ, θ) mixed distribution is

$$G(z) = \int_0^\infty e^{x(z-1)} f(x)\, dx = \exp\left[\frac{\delta}{\alpha}[\theta^\alpha - (\theta + 1 - z)^\alpha]\right]. \tag{11.6}$$

For $\alpha = 0, \frac{1}{2}, 1$, the Tweedie mixing distribution is gamma, inverse Gaussian, and deterministic, respectively. The Tweedie–Poisson distribution is correspondingly negative binomial, Poisson–inverse Gaussian, and Poisson. At first only the parameter space $0 \le \alpha \le 1, 0 < \delta, 0 \le \theta$ was explored. Later it was realized that the distribution also exists when $\alpha < 0$.

The Tweedie distribution (11.5) is infinitely divisible, and therefore, by Gurland's theorem (Section 8.3.2), a Poisson(λ) $\bigwedge_{\lambda}$ Tweedie(α, δ, θ) mixture is a generalized Poisson distribution with pgf of the form

$$G(z) = \exp\{\lambda[g(z) - 1]\}$$

where $g(z)$ is a valid pgf. The Tweedie–Poisson family is therefore infinitely divisible.

If $-\alpha = k > 0$, then

$$G(z) = \exp\left\{\frac{\delta}{k\theta^k}\left[\left(\frac{\theta}{\theta + 1 - z}\right)^k - 1\right]\right\},$$

and

$$g(z) = \left(\frac{1}{1 + 1/\theta - z/\theta}\right)^k,$$

which is the pgf of a negative binomial distribution. For $\alpha < 0$, therefore,

$$\begin{aligned} \text{Poisson}(\lambda) \bigwedge_{\lambda} \text{Tweedie}(\alpha, \delta, \theta) &\sim \text{Poisson}\left(\frac{\delta}{k\theta^k}\right) \bigvee \text{Negative binomial}\left(k, \frac{1}{\theta}\right) \\ &\sim \text{Poisson–Pascal} \\ &\sim \text{generalized Pólya–Aeppli}. \end{aligned}$$

The special case $\alpha = -1$ gives the Poisson $\bigvee$ Geometric, that is, the Pólya–Aeppli, distribution. Properties and estimation procedures for the Pólya–Aeppli and generalized Pólya–Aeppli distributions have received much attention in the biometric literature (see Sections 9.7 and 9.8).

If $0 < \alpha < 1$, then the generalizing distribution is

$$g(z) = \frac{[1 - z/(\theta + 1)]^\alpha - 1}{[1 - 1/(\theta + 1)]^\alpha - 1},$$

which is the pgf for Engen's extended negative binomial distribution (Section 5.12.2), with $w = \theta/(1+\theta)$ in his notation. So for $0 < \alpha < 1$,

$$\text{Poisson}(\lambda) \bigwedge_{\lambda} \text{Tweedie}(\alpha, \delta, \theta) \sim \text{Poisson}\left(\frac{\delta\theta^{\alpha}}{\alpha}\right) \bigvee \text{Engen}\left(\alpha, \frac{\theta}{1+\theta}\right).$$

The factorial moment generating function of a mixture of Poisson distributions is equal to the mgf of the mixing distribution. The Tweedie–Poisson fmgf is therefore

$$G(t) = \exp\left\{\frac{\delta\theta^{\alpha}}{\alpha}\left[1 - \left(1 - \frac{t}{\theta}\right)^{\alpha}\right]\right\}$$

with mean and variance

$$\mu = \delta\theta^{\alpha-1}, \qquad \mu_2 = (\theta + 1 - \alpha)\delta\theta^{\alpha-2},$$

as in Hougaard, Lee, and Whitmore (1997).

These authors have suggested the following recursion method for calculating the probabilities:

$$p_0 = \exp\left[\frac{\delta}{\alpha}[\theta^{\alpha} - (\theta+1)^{\alpha}]\right],$$

$$p_x = \frac{p_0}{x!}\left(\sum_{i=1}^{x} c_{x,i}(\alpha)\delta^i(\theta+1)^{i\alpha-x}\right),$$

where

$$c_{n,1}(\alpha) = \frac{\Gamma(n-\alpha)}{\Gamma(1-\alpha)},$$

$$c_{n,i}(\alpha) = c_{n-1,i-1}(\alpha) + c_{n-1,i}(\alpha)[(n-1) - i\alpha] \quad \text{for } 2 \le i \le n-1,$$

$$c_{n,n}(\alpha) = 1.$$

This involves a double recursion, unlike Katti and Gurland's (1961) single-recurrence equation in Section 9.8.

Hougaard, Lee, and Whitmore (1997) commented that the sum of two Tweedie–Poisson rv's with parameters $(\alpha, \delta_i, \theta)$, where $i = 1, 2$, is a Tweedie–Poisson$(\alpha, \delta_1 + \delta_2, \theta)$ rv and that a Tweedie–Poisson distribution belongs to the natural exponential family. They proved that these distributions are unimodal and described the application of the distributions with $\alpha = 0, \frac{1}{2}, 1$ to data on epileptic seizures in treated and untreated patients.

11.1.3 Negative Binomial Regression Models

The negative binomial distribution is a power series distribution (PSD) and so belongs to the exponential family. Various parameterizations have been used for regression models.

The NB1 of Cameron and Trivedi (1986) arises when a Poisson(θ) distribution is mixed using a gamma distribution with the parameterization

$$f(\theta; \lambda, P) = \frac{\theta^{\lambda/P-1}}{\Gamma(\lambda/P)P^{\lambda/P}} e^{-\theta/P},$$

giving

$$\Pr[X = x; \lambda, P] = \frac{\Gamma(\lambda/P + x)}{\Gamma(\lambda/P)\, x!} \left(\frac{1}{1+P}\right)^{\lambda/P} \left(\frac{P}{1+P}\right)^{x}, \qquad \lambda > 0,\ P > 0, \tag{11.7}$$

$x = 0, 1, \ldots$. In this parameterization, $\mu = \lambda$, $\mu_2 = \lambda(1 + P)$, and the variance is a linear function of the mean.

The NB2 of Cameron and Trivedi (1986) is the outcome when a Poisson(θ) distribution is mixed using a gamma distribution with the parameterization

$$f(\theta; \alpha, \lambda) = \frac{(\alpha/\lambda)^{\alpha}\, \theta^{\alpha-1}}{\Gamma(\alpha)} e^{-\theta\alpha/\lambda}.$$

This gives

$$\Pr[X = x; \alpha, \lambda] = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\, x!} \left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha} \left(\frac{\lambda}{\alpha+\lambda}\right)^{x}, \qquad \alpha > 0, \qquad \lambda > 0, \tag{11.8}$$

$x = 0, 1, \ldots$. Now $\mu = \lambda$, $\mu_2 = \lambda + \lambda^2/\alpha$, and the variance is a quadratic function of the mean.

The Negbin$_k$ distribution of Winkelmann, Signorino, and King (1995) [see also Winkelmann (2000), pp. 127–128] uses the parameterization (λ, σ^2), where α in (11.8) is set equal to λ^{1-k}/σ^2, giving $\mu = \lambda$ and $\mu_2 = \lambda(1 + \lambda^k \sigma^2)$. Taking $k = 0$ gives NB1, while $k = 1$ gives NB2. Note that in this parameterization σ^2 is a parameter and $\mu_2 \neq \sigma^2$.

11.1.4 Poisson Lognormal Model

Kemp and Kemp (1966) found that, if mixing is treated purely as a formal process with the Poisson parameter θ taking negative values, then a Poisson(θ) $\bigwedge_{\theta}$ Normal(μ, θ) distribution exists with pgf

$$\begin{aligned} G_N(z) &= \int_{-\infty}^{\infty} \frac{e^{\theta(z-1)} e^{-(\theta-\mu)^2/(2\sigma^2)}\, d\theta}{(2\pi\sigma^2)^{1/2}} \\ &= \exp\left[\tfrac{1}{2}(\mu - \sigma^2)(z - 1) + \sigma^2(z^2 - 1)\right]. \end{aligned} \tag{11.9}$$

provided that $\mu - \sigma^2 \geq 0$. This is known as the Hermite distribution. The Poisson(θ) $\bigwedge_{\theta}$ Truncated normal(μ, σ) model, with the normal distribution truncated at zero, is a meaningful discrete model; see Section 9.4 concerning both distributions.

If the rv X has a normal distribution with mean α and variance ρ^2 and $X = \ln Y$, then Y is said to have a lognormal distribution (Johnson, Kotz, and Balakrishnan, 1994, Chapter 14). The pdf of Y is

$$f(y) = \frac{1}{y(2\pi\rho^2)^{1/2}} \exp\left[-\frac{(\ln y - \alpha)^2}{2\rho^2}\right], \qquad y \geq 0,$$

and the uncorrected moments are $E[Y^k] = \exp(k\alpha + k^2\rho^2/2)$. The books by Aitchison and Brown (1957) and Crow and Shimizu (1988) give good accounts of the properties and applications of the lognormal distribution.

The distribution obtained by assigning a lognormal distribution to the Poisson parameter, that is, Poisson$(\theta) \bigwedge_{\theta}$ Lognormal(α, ρ), is termed a *discrete lognormal distribution*. Equivalently, it is the distribution that arises when the Poisson natural parameter $\ln(\theta)$ has a normal distribution. It was originally derived as a competitor to the logarithmic distribution for species abundance data (Anscombe, 1950); see Section 7.1.2.

The pmf is

$$p_x = \int_0^\infty \frac{e^{-\theta}\theta^x}{x!} \cdot \frac{1}{\theta(2\pi\rho^2)^{1/2}} \exp\left(\frac{-(\ln\theta - \alpha)^2}{2\rho^2}\right) d\theta \tag{11.10}$$

$$= \int_{-\infty}^\infty \frac{\exp(-\xi e^u)\xi^x e^{xu}}{x!} \cdot \frac{1}{\theta(2\pi\rho^2)^{1/2}} \exp\left(\frac{-u^2}{2\rho^2}\right) du, \tag{11.11}$$

where $u = \ln\theta - \alpha$ and $\xi = e^\alpha$. The pgf is

$$G_{\text{LN}}(z) = \int_{-\infty}^\infty \exp[\xi e^u(z-1)] \frac{1}{\theta(2\pi\rho^2)^{1/2}} \exp\left(\frac{-u^2}{2\rho^2}\right) du. \tag{11.12}$$

This distribution is not very tractable. The factorial moments are, however, equal to the uncorrected moments of the lognormal distribution, giving

$$\mu = \xi e^{\rho^2/2}, \qquad \mu'_{[2]} = e^{2\alpha+2\rho^2}, \quad \text{i.e., } \mu_2 = \mu + \mu^2(e^{\rho^2} - 1). \tag{11.13}$$

Summaries of the properties of the discrete lognormal distribution appear in Reid (1981) and Shaban (1988). For its use in regression modeling see Hinde (1982) and Hougaard, Lee, and Whitmore (1997).

11.1.5 Poisson–Inverse Gaussian (Sichel) Model

The *Poisson–inverse Gaussian distribution* is a two-parameter mixture of Poisson distributions, obtained by allowing the Poisson parameter λ to have an inverse-Gaussian distribution with density function

$$f(\lambda) = \frac{(1-\theta)^{-1/4}[2/(\alpha\theta)]^{-1/2}\lambda^{-3/2}}{2K_{1/2}(\alpha\sqrt{1-\theta})} \exp\left[\left(1 - \frac{1}{\theta}\right)\lambda - \frac{\alpha^2\theta}{4\lambda}\right], \qquad \lambda > 0, \tag{11.14}$$

where $0 < \alpha$, $0 < \theta < 1$, and $K_\nu(\cdot)$ is a modified Bessel function of the third kind (Section 1.1.5). Since $K_{-1/2}(y) = K_{1/2}(y) = \sqrt{\pi/(2y)}e^{-y}$, the pmf is

$$\Pr[X = x] = \sqrt{\frac{2\alpha}{\pi}} \frac{\exp(\alpha\sqrt{1-\theta})(\alpha\theta/2)^x}{x!} K_{x-1/2}(\alpha), \qquad x = 0, 1, \ldots. \tag{11.15}$$

The distribution was introduced by Holla (1966) as a useful model in studies of repeated accidents and recurrent disease symptoms. It has also been researched by Sankaran (1968) and Sichel (1971). Shaban (1981) looked at limiting cases and approximations.

Sichel (1982b) investigated the following parameter estimation methods: sample moments, using the sample mean and variance; use of the mean and zero frequency; and maximum likelihood. For $1.5 < \alpha < 20.0$ and $0.90 \leq \theta$, he found the first two of these methods to be very inefficient and therefore recommended the use of maximum-likelihood estimation. Willmot (1987b) has discussed the use of the Poisson–inverse Gaussian distribution as an alternative to the negative binomial. Its use in regression models has been studied by Dean, Lawless, and Willmot (1989).

The two-parameter Poisson–inverse Gaussian distribution is a special case of the more general *Sichel distribution*; this is a three-parameter mixture of Poisson distributions, obtained by allowing the Poisson parameter λ to have the following generalization of an inverse-Gaussian distribution [see Sichel (1974, 1975)]:

$$f(\lambda) = \frac{(1-\theta)^{\gamma/2}[2/(\alpha\theta)]^\gamma \lambda^{\gamma-1}}{2K_\gamma(\alpha\sqrt{1-\theta})} \exp\left[\left(1 - \frac{1}{\theta}\right)\lambda - \frac{\alpha^2\theta}{4\lambda}\right], \qquad \lambda > 0,$$

where $0 < \alpha, 0 < \theta < 1, -\infty < \gamma < \infty$, and $K_\nu(\cdot)$ is a modified Bessel function. An alternative parameterization with $\beta = \alpha\theta/2$ (Atkinson and Yeh, 1982) leads to the following expression for the pmf:

$$\Pr[X = x] = \frac{(\alpha^2 - 2\alpha\beta)^{\gamma/2}}{\alpha^\gamma K_\gamma(\sqrt{\alpha^2 - 2\alpha\beta})} \frac{\beta^x K_{\gamma+x}(\alpha)}{x!}, \qquad x = 0, 1, \ldots, \tag{11.16}$$

with $0 < \beta < \alpha/2$. Stein, Zucchini, and Juritz (1987) recommended a further reparameterization with $\xi = \beta(1 - 2\beta/\alpha)^{-1/2}$ and $\omega = (\xi^2 + \alpha^2)^{1/2} - \xi$. This gives

$$\Pr[X = 0] = \frac{(\omega/\alpha)^\gamma K_\gamma(\alpha)}{K_\gamma(\omega)},$$

$$\Pr[X = 1] = \frac{(\xi\omega/\alpha)K_{\gamma+1}(\alpha)}{K_\gamma(\alpha)} \Pr[X = 0],$$

and in general

$$\Pr[X = x] = \frac{(\omega/\alpha)^\gamma}{K_\gamma(\omega)} \frac{(\xi\omega/\alpha)^x K_{\gamma+x}(\alpha)}{x!}, \qquad x = 0, 1, \ldots. \tag{11.17}$$

Use of the recurrence relationship $K_{n+1}(y) = (2n/y)K_n(y) + K_{n-1}(y)$, together with $K_{-n}(y) = K_n(y)$, yields the useful recurrence formula

$$\begin{aligned}\Pr[X = x] &= \frac{2\beta}{\alpha}\left(\frac{\gamma + x - 1}{x}\right)\Pr[X = x - 1] + \frac{\beta^2}{x(x-1)}\Pr[X = x - 2] \\ &= \frac{2\xi\omega}{\alpha^2}\left(\frac{\gamma + x - 1}{x}\right)\Pr[X = x - 1] + \frac{(\xi\omega/\alpha)^2}{x(x-1)}\Pr[X = x - 2]\end{aligned} \tag{11.18}$$

for $x = 2, 3, \ldots$.

The pgf is

$$G(z) = \sum_{j \geq 0} \frac{\xi^j K_{\gamma+j}(\omega)(z-1)^j}{K_\gamma(\omega)j!} \tag{11.19}$$

$$= \frac{K_\gamma[\omega\sqrt{1 - 2\beta(z-1)}]}{K_\gamma(\omega)[1 - 2\beta(z-1)]^{\gamma/2}}; \tag{11.20}$$

see Willmot (1986). The factorial moments are $\mu'_{[r]} = \xi^r K_{\gamma+r}(\omega)/K_\gamma(\omega)$; the mean and variance are therefore

$$\begin{aligned}\mu &= \frac{\xi\, K_{\gamma+1}(\omega)}{K_\gamma(\omega)}, \\ \mu_2 &= \frac{\xi\, K_{\gamma+1}(\omega)}{K_\gamma(\omega)}\left[1 + \xi\left(\frac{K_{\gamma+2}(\omega)}{K_{\gamma+1}(\omega)} - \frac{K_{\gamma+1}(\omega)}{K_\gamma(\omega)}\right)\right].\end{aligned} \tag{11.21}$$

The three-parameter Sichel distribution becomes the two-parameter Poisson–inverse Gaussian distribution when $\gamma = -\frac{1}{2}$. The formulas above simplify considerably in this special case; for instance, the mean and variance simplify to $\mu = \xi$ and $\mu_2 = \xi(1 + \xi/\omega)$. For this special case the parameters ξ and ω can be regarded as mean and shape parameters; we refer the reader to Stein et al. (1987).

An important feature of the three-parameter distribution is its flexibility, as well as its very long positive tail combined with finite moments of all orders. This makes the distribution particularly suitable for modeling highly skewed data.

The three-parameter form was introduced by Sichel (1971) and studied further by him in Sichel (1973a,b) in the context of diamondiferous deposits. In Sichel (1974, 1975) he applied the distribution to sentence length and to word frequencies and in Sichel (1982a) he examined its use for modeling repeat buying. Ord and Whitmore (1986) have used the two-parameter form as a model for species abundance data. Its application to the number of insurance claims per policy has been studied by Willmot (1987b).

Atkinson and Yeh (1982) developed an approximate maximum-likelihood method for the three-parameter distribution using a grid of half-integer values of γ. This removed a major numerical handicap regarding the use of the more flexible three-parameter form of the distribution. Ord and Whitmore (1986) investigated minimum χ^2 estimation as well as maximum-likelihood estimation. Stein et al. (1987) have reexamined estimation for both the two-parameter and three-parameter forms. Their reparameterization enabled them to give an algorithm for an exact maximum-likelihood method in the general case. They pointed out that as $\alpha \to \infty$ the distribution tends to a modified logarithmic distribution.

11.1.6 Poisson Polynomial Distribution

Cameron and Johansson's (1997) Poisson polynomial distributions have pmf's that are constructed from a Poisson pmf with a polynomial adjustment. The pmf for the ℓth-order distribution is

$$\Pr[X = x] = C\,\frac{e^{-\theta}\theta^x}{x!}\left(\sum_{i=0}^{\ell} a_i x^i\right)^2, \qquad x = 0, 1, \ldots, \qquad 0 < \theta,$$
$$a_0 = 1, \qquad a_i \in \mathbb{R}, \tag{11.22}$$

where C is the normalizing constant; see Cameron and Trivedi (1998). The squared factor ensures that the probabilities are nonnegative. This is a semiparametric regression model for $\ell > 2$. Here we are concerned only with the *Poisson polynomial distributions of order* $\ell = 1, 2$, having two and three unknown parameters, respectively.

For the first-order Poisson polynomial distribution the pgf is

$$\begin{aligned} G_1(z) &= C\sum_{x=0}^{\infty}\frac{e^{-\theta}\theta^x}{x!}(1 + 2ax + a^2x^2)z^x \\ &= C\,e^{\theta(z-1)}[1 + a(a+2)\theta z + a^2\theta^2 z^2] \qquad (11.23) \\ &= e^{\theta(z-1)}\frac{(1 + a\theta z)^2 + a^2\theta z}{(1 + a\theta)^2 + a^2\theta}, \qquad (11.24) \end{aligned}$$

where $C = (1 + a\theta)^2 + a^2\theta$. When $a < -2$ or $a > 0$, (11.23) shows that the distribution is a mixture of three Poisson distributions, with support $x = 0, 1, \ldots,$ $x = 1, 2, \ldots,$ and $x = 2, 3, \ldots,$ all with the same parameter θ. From (11.24), when $a < -2$ or $a > 0$, it is also a mixture of Ong's (1988) discrete Charlier distribution and a shifted Poisson distribution.

The moments can be derived from Poisson moments; the first two are

$$\begin{aligned} \mu &= C[\theta + 2a(\theta + \theta^2) + a^2(\theta + 3\theta^2 + \theta^3)], \\ \mu_2 &= C[(\theta + \theta^2) + 2a(\theta + 3\theta^2 + \theta^3) + a^2(\theta + 7\theta^2 + 6\theta^3 + \theta^4)] - \mu^2. \end{aligned} \tag{11.25}$$

The second-order Poisson polynomial distribution has the pmf

$$\Pr[X = x] = C^* \frac{e^{-\theta}\theta^x}{x!}(1 + ax + bx^2)^2, \qquad x = 0, 1, \ldots, \qquad 0 < \theta. \tag{11.26}$$

The distribution certainly exists for $a \geq 0$, $b \geq 0$, and the moments can be found from the Poisson moments, as in Cameron and Trivedi (1998).

Cameron and Johansson (1997) found that, for a regression application to recreational boating trips with mildly overdispersed data, they needed a Poisson polynomial model with $\ell = 5$ in order to outperform a negative binomial model. On the other hand, for regression data on (underdispersed) takeover bids, they found that a Poisson polynomial model with $\ell = 1$ outperformed the hurdle Poisson, double-Poisson, and certain other distributions capable of modeling underdispersion.

11.1.7 Weighted Poisson Distributions

Del Castillo and Pérez-Casany's (1998) weighted Poisson family was deliberately developed for use in generalized linear models with covariates, as an alternative to the negative binomial distribution. It is a regular exponential family with the advantage that the index of dispersion can be greater, equal to, or (unlike the negative binomial) less than unity.

Consider the effect of the weight function $w(x) = (x + a)^r$, $a \geq 0$, $r \in \mathbb{R}$, on the Poisson variable X with parameter λ. The outcome has the pmf

$$\Pr[X^w = x; \lambda, r, a] = \frac{C(\lambda, r, a)(x + a)^r \lambda^x e^{-\lambda}}{x!}, \tag{11.27}$$

where $C(\lambda, r, a) = \{\sum_{x=0}^{\infty}[(x + a)^r \lambda^x e^{-\lambda}]/x!\}^{-1}$ is the normalizing constant. The Poisson distribution is the special case $r = 0$. For a given value of λ the mean increases with r.

Del Castillo and Pérez-Casany commented that if X has a weighted Poisson distribution (WPD) with parameters λ, r, a, then its size-biased form X^* [with weight function $w(x) = x$] is WPD$(\lambda, r, a + 1)$. More generally, when $w(x) = x + a$,

$$\Pr[X^w = x] = \left(1 - \frac{\lambda}{\lambda + a}\right)\Pr[X = x] + \frac{\lambda}{\lambda + a}\Pr[X^* = x],$$

and when $w(x) = (x + a)^r$, $r \in \mathbb{Z}^+$, X^w is a mixture of the successive size-biased versions of order s, $s = 0, 1, 2, \ldots, r$. The parameter a is therefore a measure of closeness to the Poisson distribution.

Del Castillo and Pérez-Casany proved that the moments about the origin are nondecreasing functions of λ and r, where $\lambda > 0$, $r \in \mathbb{R}$. Also they are nondecreasing in a provided that $r \leq 0$. They showed that

$$E[(X + a)^\ell] = \frac{C(\lambda, r + \ell, a)}{C(\lambda, r, a)}, \qquad \ell \in \mathbb{R},$$

and hence that the mean and variance are

$$\mu = \lambda \frac{C(\lambda, r+1, a)}{C(\lambda, r, a)},$$
$$\mu_2 = \frac{C(\lambda, r+2, a)C(\lambda, r, a) - C^2(\lambda, r+1, a)}{C^2(\lambda, r, a)}. \tag{11.28}$$

They proved furthermore that the index of dispersion μ_2/μ is greater or less than unity iff r is less or greater than zero. They demonstrated graphically that the index of dispersion is almost constant when r is fixed, in which case the variance is approximately a linear function of the mean. Their paper gives two examples fitted using profile log-likelihood, one with overdispersed data and the other with underdispersed data.

11.1.8 Double-Poisson and Double-Binomial Distributions

These distributions belong to the double-exponential family of distributions that was developed by Efron (1986) in the context of loglinear regression when the Poisson and binomial mean/variance relationships are too restrictive.

Let $f(y; \theta)$ be a member of a one-parameter exponential family. Then the presence of the second parameter, ψ, in

$$f(y; \theta, \psi) = c(\theta, \psi)\sqrt{\psi}\; f(y; \theta)^{\psi}[f(y; y)]^{1-\psi}, \tag{11.29}$$

where $c(\theta, \psi)$ is a normalizing constant, allows for overdispersion in the data to be fitted.

The double-Poisson distribution has the pmf

$$p(x; \mu, \psi) = \frac{c(\mu, \psi)\psi^{1/2}e^{-\mu\psi}(x/e)^x(\mu e/x)^{x\psi}}{x!}$$
$$= \frac{c(\mu, \psi)\psi^{1/2}e^{-\mu\psi - x}x^x(\mu e/x)^{x\psi}}{x!}. \tag{11.30}$$

Efron showed that

$$\frac{1}{c(\mu, \psi)} \simeq 1 + \frac{1-\psi}{12\mu\psi}\left(1 + \frac{1}{\mu\psi}\right),$$

which is approximately unity. Fitting the distribution with both μ and ψ to data while ignoring this normalizing constant gives approximate maximum-likelihood estimation. When $\psi = 1$ the distribution becomes a Poisson distribution; for $\psi < 1$ there is overdispersion and for $\psi > 1$ there is underdispersion.

Mosteller and Parunak (1985) were interested in using EDA to identify outliers in a two-way table of counts of 19 kinds of archaeological artifacts found at varying distances from permanent water in Nevada. Efron (1986) used the double-Poisson distribution to reanalyze these data.

The double-binomial pdf is

$$p(x;\pi,\psi) = c(\pi,\psi)\binom{n}{x}\frac{x^x(n-x)^{n-x}}{n^n}\cdot\frac{n^{n\psi}\pi^{\psi x}(1-\pi)^{\psi(n-x)}}{x^{\psi x}(n-x)^{\psi(n-x)}}. \tag{11.31}$$

This distribution was used by Efron to analyze numbers of people in differing-sized samples who tested positive for toxoplasmosis in 34 cities in El Salvador; the predictor variable was rainfall. Lindsey (1995) gave a worked example which contrasts the fits of a binomial and a double-binomial distribution to Geissler's classical data on family composition.

The difficulty in using the double-binomial distribution is the intractability of the normalizing constant $c(\pi,\psi)$.

11.1.9 Simplex–Binomial Mixture Model

The simplex–binomial distribution is similar to the beta–binomial distribution, but it has the Barndorff-Neilsen and Jørgensen (1991) simplex distribution as the mixing distribution instead of the beta distribution. The pdf of the standard (two-parameter) univariate simplex distribution is

$$f(y;\mu,\sigma) = \left[\frac{\lambda}{2\pi[y(1-y)]^3}\right]^{1/2}\exp\left(-\frac{1}{2\sigma^2}d(y;\mu)\right), \qquad 0<y<1,$$

where $\mu\in(0,1)$, $\sigma^2 = 1/\lambda > 0$, and

$$d(y;\mu) = \frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}$$

is the unit deviance [see Jørgensen (1997, Section 2.3) for a definition of unit deviance]. It is the special case $(\alpha_1,\alpha_2) = \left(-\frac{1}{2},-\frac{1}{2}\right)$ of the more general four-parameter simplex distribution with pdf

$$f(y;\alpha_1,\alpha_2,\mu,\sigma) = \frac{C(\alpha_1,\alpha_2,\mu,\sigma)}{(2\pi\sigma^2)^{1/2}}y^{\alpha_1-1}(1-y)^{\alpha_2-1}\exp[\lambda t_{\alpha_1,\alpha_2}(y;\mu)],$$

where $C(\alpha_1,\alpha_2,\mu,\sigma)$ is part of the normalizing constant and

$$t_{\alpha_1,\alpha_2}(y;\mu) = -\frac{(y-\mu)^2}{y(1-y)}\,\mu^{2\alpha_1-1}(1-\mu)^{2\alpha_2-1}.$$

This distribution is very much more flexible than the beta distribution; Jørgensen (1997) illustrated this using diagrams for 18 combinations of the parameters. He commented on five special cases, $(\alpha_1,\alpha_2) = \left(\frac{1}{2},\frac{1}{2}\right)$, $(0,0)$, $\left(-\frac{1}{2},-\frac{1}{2}\right)$, $\left(-\frac{1}{2},\frac{1}{2}\right)$, $\left(\frac{1}{2},-\frac{1}{2}\right)$, that all give exponential family dispersion models and for which the expressions for $C(\alpha_1,\alpha_2,\mu,\sigma)$ simplify; in three of these, when $\alpha_1+\alpha_2=0$, they become generalized inverse-Gaussian distributions.

The pmf of the mixed binomial(n, p) distribution, using the simplex distribution as the mixing distribution, is

$$\begin{aligned}\Pr[X=x] &= \binom{n}{x}\frac{C(\alpha_1,\alpha_2,\mu,\sigma)}{(2\pi\sigma^2)^{1/2}}\int_0^1 p^{\alpha_1+x-1}(1-p)^{\alpha_2+n-x-1}e^{\lambda t_{\alpha_1,\alpha_2}(p;\mu)}\,dp\\ &= \binom{n}{x}\mu^x(1-\mu)^{n-x}\frac{C(\alpha_1,\alpha_2,\mu,\sigma)}{C(\alpha_1+x,\alpha_2+n-x,\mu,\sigma\mu^x(1-\mu)^{n-x})},\end{aligned} \tag{11.32}$$

where

$$\begin{aligned}t_{\alpha_1,\alpha_2}(p;\mu) &= -\frac{(p-\mu)^2}{2p(1-p)}\mu^{2\alpha_1-1}(1-\mu)^{2\alpha_2-1}\\ &= \mu^{-2x}(1-\mu)^{-2(n-x)}t_{\alpha_1+x,\alpha_2+n-x}(p;\mu).\end{aligned}$$

The simplex–binomial pmf is therefore equal to the original binomial pmf times a factor that tends to 1 as σ^2 tends to 0.

When $(\alpha_1,\alpha_2)=\left(-\frac{1}{2},-\frac{1}{2}\right)$, $C(\alpha_1,\alpha_2,\mu,\sigma)=1$ and the pmf simplifies to

$$\Pr[X=x]=\frac{\binom{n}{x}\mu^x(1-\mu)^{n-x}}{C\left(x-\frac{1}{2},n-x-\frac{1}{2},\mu,\sigma\mu^x(1-\mu)^{n-x}\right)}. \tag{11.33}$$

In general there is no closed expression for $C(\alpha_1,\alpha_2,\beta,\gamma)$; similarly there are no known closed expressions for the moments. This is a computer-intensive distribution that has been put forward for use in binomial regression modeling when a beta mixing distribution would be insufficiently flexible.

11.2 MISCELLANEOUS DISCRETE DISTRIBUTIONS

This section contains descriptions of various distributions that have not fitted straightforwardly into the previous chapters. In some cases, for example, the lost-games distribution, an earlier section would have been made unwieldy by their inclusion. However, most of the distributions in this chapter, such as the discrete Adés and the Gram–Charlier type B distributions, do not relate naturally to the distributions discussed earlier in the book. A few of the distributions that we include have lain dormant for many years; we hope that these have potential, either for modeling purposes or for the further development of the theory of discrete distributions.

The arrangement within this section is alphabetical.

11.2.1 Dandekar's Modified Binomial and Poisson Models

Consider a sequence of n trials where the probability of a success in a single trial is p (constant), with the proviso that, if a trial produces a success, then the probability of success for the next $m-1$ trials is zero. Dandekar's (1955) *first modified binomial distribution* is obtained by supposing in addition that the probability of success in the first observed trial is p, not zero. Let X be the number of successes in the n trials; then

$$\Pr[X \le x] = q^{n-xm} \sum_{j=0}^{x} \binom{n - xm + j - 1}{j} p^j, \tag{11.34}$$

where $q = 1 - p$ and $x = 0, 1, 2, \ldots$ such that $x \le [n/m]$, with $[\cdot]$ denoting the integer part.

Let $n \to \infty$, $m \to \infty$, and $p \to 0$ so that $np \to \lambda$ and $m/n \to k < 1$, where λ and k are constant. Then Dandekar's first modified binomial distribution tends to his *first modified Poisson distribution*, for which

$$\Pr[X \le x] = e^{\lambda(kx-1)} \sum_{j=0}^{x} \frac{(1 - kx)^j \lambda^j}{j!}, \tag{11.35}$$

where $0 < \lambda$, $0 < k < 1$, and $x = 0, 1, 2, \ldots$ such that $x \le [1/k]$, where $[\cdot]$ denotes the integer part.

Dandekar's (1955) *second modified binomial distribution* arises when the first observed trial is assumed to occur at a random point in an infinite sequence of such trials; his *second modified Poisson distribution* is then the result of a limiting process analogous to that for his first modified Poisson distribution; see Dandekar (1955) and Patil et al. (1984).

Dandekar used these distributions to fit data on the numbers of pregnancies experienced by 369 women over a five-year period. He reasoned that conception is not possible during pregnancy or for a short time afterward. Neither of the modified binomial fits was better than the binomial fit. He also argued that when a person has a major accident then he or she becomes incapacitated for a period during which another accident will not take place. The fit of his first modified Poisson distribution to the classical data on accidents to 647 women working on high explosive shells was much better than the Poisson fit. It also gave a markedly better fit than the Poisson to data on numbers of lost articles. The problem remained, however, that in both cases his estimated values of k were negative and therefore incompatible with his derivation of the model.

Dandekar's distributions were reexamined by Basu (1955). Simple forms for the pgf's, means, and variances of these four distributions appear to be unknown.

11.2.2 Digamma and Trigamma Distributions

The psi (digamma) function is defined as $\psi(z) = d \ln \Gamma(z)/dz$ (see Section 1.1.2). Hence

$$\Delta\psi(z) = \psi(z + 1) - \psi(z) = \frac{1}{z},$$

$$\Delta^n \psi(z) = \Delta^{n-1}\left(\frac{1}{z}\right) = \frac{(-1)^{n-1}(n - 1)!(z - 1)!}{(z + n - 1)!},$$

and Newton's forward-difference formula (Section 1.1.3) gives

$$\psi(z+n) = \sum_{j=0}^{n} \binom{n}{j} \Delta^j \psi(z) = \psi(z) + \sum_{j=1}^{n} \frac{(-1)^{j-1} n!(z-1)!}{j(n-j)!(z+j-1)!}. \qquad (11.36)$$

More generally, when $\operatorname{Re}(z+\nu) > 0$, z not a negative integer,

$$\psi(z+\nu) - \psi(z) = \sum_{j \ge 1} \frac{(-1)^{j-1} \nu(\nu-1)\cdots(\nu-j+1)}{jz(z+1)\cdots(z+j-1)} \qquad (11.37)$$

is a convergent series; see Nörlund (1923). Taking $\nu = -\alpha$ and $z = \alpha + \gamma$, where α and γ are real and $\gamma > 0$, $\alpha > -1$, $\alpha + \gamma > 0$, the series becomes

$$\psi(\alpha+\gamma) - \psi(\gamma) = \sum_{x \ge 1} \frac{(\alpha+x-1)!(\alpha+\gamma-1)!}{x(\alpha-1)!(\alpha+\gamma+x-1)!}, \qquad (11.38)$$

with all terms having the same sign. Moreover, for the trigamma function $\psi'(\gamma)$,

$$\psi'(\gamma) = \lim_{\alpha \to 0} \frac{\psi(\alpha+\gamma) - \psi(\gamma)}{\alpha} = \sum_{x \ge 1} \frac{(x-1)!(\gamma-1)!}{x(\gamma+x-1)!}. \qquad (11.39)$$

Sibuya's (1979) *digamma and trigamma distributions* have probabilities that are proportional to the terms in (11.38) and (11.39), respectively. For the digamma distribution therefore

$$\Pr[X = x] = \frac{1}{\psi(\alpha+\gamma) - \psi(\gamma)} \times \frac{(\alpha+x-1)!(\alpha+\gamma-1)!}{x(\alpha-1)!(\alpha+\gamma+x-1)!}, \qquad (11.40)$$

and for the trigamma distribution

$$\Pr[X = x] = \frac{1}{\psi'(\gamma)} \times \frac{(x-1)!(\gamma-1)!}{x(\gamma+x-1)!} \qquad (11.41)$$

(in both cases the support is $x = 1, 2, \ldots$). When $\gamma = 1$, the trigamma distribution becomes a zeta distribution (see Section 11.2.20).

Sibuya obtained these distributions as limiting cases of the zero-truncated hypergeometric type IV (beta–negative binomial) distribution (Section 6.2.3), with pgf

$$\frac{{}_2F_1[\alpha, \beta; \alpha+\beta+\gamma; z] - 1}{{}_2F_1[\alpha, \beta; \alpha+\beta+\gamma; 1] - 1} \qquad (11.42)$$

and probabilities

$$\Pr[X = x] = \frac{\left(\dfrac{(\alpha + x - 1)!(\beta + x - 1)!(\alpha + \beta + \gamma - 1)!}{x!(\alpha - 1)!(\beta - 1)!(\alpha + \beta + \gamma + x - 1)!}\right)}{\left(\dfrac{(\alpha + \beta + \gamma - 1)!(\gamma - 1)!}{(\alpha + \gamma - 1)!(\beta + \gamma - 1)!} - 1\right)}, \tag{11.43}$$

$x = 1, 2, \ldots$. The digamma distribution is the outcome as $\beta \to 0$ with $\alpha > 0$; the trigamma is obtained when $\alpha \to 0$ and $\beta \to 0$.

Bernardo (1976) has written a computer algorithm for the psi (digamma) function which can be used to facilitate calculation of the probabilities for the digamma distribution. Schneider (1978) and Francis (1991) have developed an algorithm for the trigamma function.

11.2.3 Discrete Adès Distribution

Perry and Taylor (1985) put forward the *discrete Adès distribution* in order to model counts of individuals per unit of habitat in population ecology. There is much empirical evidence, especially in entomology, to support the view that, in a data set containing several samples of data, the theoretical mean and variance vary from sample to sample in such a way that

$$\mu_2 = \alpha\mu^{\beta}, \tag{11.44}$$

that is, $\ln \mu_2 = \ln \alpha + \beta \ln \mu$; this is known as *Taylor's power law*.

Suppose that X has a gamma distribution with density function

$$f(x) = \frac{\lambda^r e^{-\lambda x} x^{r-1}}{\Gamma(r)}, \qquad x \geq 0, \qquad \lambda > 0, \qquad r > 0, \tag{11.45}$$

and that

$$Y = \begin{cases} 0 & \text{if } 0 \leq x \leq 1, \\ (\ln x)^b & \text{if } x \geq 1. \end{cases} \tag{11.46}$$

Then Y is deemed to have an Adès distribution with parameters r, λ, b; sometimes $c = 1/b$ has been used instead of b. Perry and Taylor (1985) checked both theoretically and numerically that (11.44) holds to a very good degree of approximation given a family of Adès distributions with constant b and constant coefficient of variation.

The Adès distribution defined in this way is a continuous distribution with an atom of probability at the origin. It does not have closed forms for its mean and variance. Holgate (1989) has obtained asymptotic formulas for the central moments and hence has been able to reexamine the relationship between the variance and the mean. Holgate also looked at estimation by the method of moments.

The *discrete Adès distribution* of Perry and Taylor (1985, 1988) is the distribution of W, where

$$\begin{aligned} \Pr[W=0] &= \Pr[0 \le Y < 0.5], \\ \Pr[W=x] &= \Pr[x-0.5 \le Y < x+0.5], \qquad x=1,2,\ldots. \end{aligned} \tag{11.47}$$

Perry and Taylor (1988) fitted this discrete Adès distribution to 22 entomological data sets made up altogether of 215 samples, with results that they found very encouraging. To do this, they used the maximum likelihood program (MLP) of Ross (1980).

Kemp (1987b) pointed out that constrained (one-parameter) families of the negative binomial, Neyman type A, Pólya–Aeppli, and inflated Poisson distributions can be constructed so that they obey Taylor's power law. There has been controversy between Perry and Taylor (1988) and Kemp (1988a) as to whether these constrained distributions are able to exhibit the same flexibility of form as the discrete Adès distribution.

11.2.4 Discrete Bessel Distribution

The *discrete Bessel distribution* of Yuan and Kalbfleisch (2000) is a two-parameter power series distribution with pmf

$$\Pr[X=x] = \frac{1}{I_\nu(a)\, x!\Gamma(x+\nu+1)}\left(\frac{a}{2}\right)^{2x+\nu}, \qquad x=0,1,2,\ldots, \tag{11.48}$$

where $\nu > -1$, $a > 0$, and $I_\nu(a)$ is the modified Bessel function of the first kind (Section 1.1.5):

$$I_\nu(a) = \left(\frac{a}{2}\right)^\nu \sum_{n=0}^{\infty} \frac{1}{n!\Gamma(n+\nu+1)}\left(\frac{a}{2}\right)^{2n}.$$

The pgf is

$$G(z) = \frac{z^{-\nu/2}\, I_\nu(a\sqrt{z})}{I_\nu(a)}, \tag{11.49}$$

and the mean and variance are

$$\mu = \frac{a\, I_{\nu+1}(a)}{2\, I_\nu(a)} \qquad \text{and} \qquad \mu_2 = \frac{a^2\, I_{\nu+2}(a)}{4\, I_\nu(a)} + \mu(1-\mu); \tag{11.50}$$

higher moments are obtainable from the factorial moments

$$\mu'_{[r]} = \left(\frac{a}{2}\right)^r \frac{I_{\nu+r}(a)}{I_\nu(a)}, \qquad r=1,2,3,\ldots. \tag{11.51}$$

The mode is $[\{(a^2+\nu^2)^{1/2}-\nu\}/2]$, where $[\cdot]$ denotes the integer part; when $m = \{(a^2+\nu^2)^{1/2}-\nu\}/2$ is an integer, there are joint modes at m and $m-1$.

Yuan and Kalbfleisch derived the discrete Bessel distribution as an inverse probability distribution by assigning a Poisson prior to a gamma distribution.

They considered the number of customers using a laundromat with respect to electricity consumption. Suppose that the power consumption for each customer is an iid exponential rv with parameter θ and that there is also an independent overhead power consumption having a gamma distribution with scale parameter θ and shape parameter $\nu + 1$. If the number of customers is x, then the total power consumption Y has a gamma distribution with scale parameter θ and shape parameter $x + \nu + 1$.

Suppose now that customers arrive according to a Poisson process with parameter λ. Then the prior distribution of the number of customers X arriving in time $(0, t)$ is Poisson(λt). Given an observation $Y = y$, the posterior distribution of the number of customers is

$$\Pr[X = x] = \frac{C(\theta\lambda t y)^x}{x!\Gamma(x + \nu + 1)}, \tag{11.52}$$

where C is a normalizing constant. The posterior distribution of X is therefore (11.48) with $a = 2\sqrt{(\theta\lambda t y)}$.

Yuan and Kalbfleisch commented on the relationship between the discrete Bessel distribution and the conditional distribution of $Y|(X - Y = \nu)$, where X and Y are independent Poisson rv's with parameters λ_1 and λ_2 and $\lambda_1\lambda_2 = a^2/4$; see Section 4.12.3. They also gave a derivation of the distribution as the conditional distribution of $Y|(X - Y = [\nu])$, where Y is Poisson(λ_2) and X has a mixed Poisson distribution.

The modified Bessel function of the first kind satisfies the recurrence relation

$$I_\nu(y) = I_{\nu+2}(y) + 2(\nu + 1)y^{-1}I_{\nu+1}(y).$$

The distribution with parameters (ν, a) is therefore a mixture of one with parameters $(\nu + 1, a)$ and a right-shifted one with support $1, 2, \ldots$ and parameters $(\nu + 2, a)$, given the mixing weights $2\nu R_\nu(a)/a$ and $R_\nu(a)R_{\nu+1}(a)$, respectively. Here $R_\nu(a) = I_{\nu+1}(a)/I_\nu(a)$ is called a Bessel quotient.

Yuan and Kalbfleisch also discussed simulation of the discrete Bessel distribution and explored its relationship to the von Mises distribution. Relevant properties of the Bessel quotients, including bounds, are given in an appendix to their paper.

11.2.5 Discrete Mittag–Leffler Distribution

The relationships between this distribution and certain continuous distributions has aroused the interest of probabilists, but the distribution has received scant attention from statisticians.

Pillai (1990) has shown that

$$F_a(x) = 1 - E_a(-x^a), \qquad 0 < a \le 1, \qquad 0 < x, \tag{11.53}$$

where $E_a(z) = \sum_{k=0}^{\infty} z^k/\Gamma(1 + ak)$, the Mittag–Leffler function, is the cdf of a continuous distribution. He called it the Mittag–Leffler distribution, explored its properties, and found that its Laplace transform is

$$\Psi(s) = (1 + s^a)^{-1}, \qquad 0 < a < 1, \qquad 0 < s.$$

A mixture of independent Poisson distributions with parameter $\theta\lambda$, where $0 < \theta$, θ constant, and λ has the continuous Mittag–Leffler distribution (11.53), therefore has the pgf

$$G_a(z) = [1 + \theta^a(1 - z)^a]^{-1}.$$

This distribution with $\theta^a = c$ is known as the *discrete Mittag–Leffler distribution*; see Pillai (1990) and Pillai and Jayakumar (1995).

Pillai and Jayakumar gave the following derivation. Consider a sequence of independent Bernoulli trials with probabilities of success $a/1$, $a/2$, $a/3, \ldots$, $0 < a < 1$, and suppose that the first success occurs at trial N. Then

$$\begin{aligned} \Pr[N = n] = p_n &= (1 - a)(1 - a/2) \cdots \{1 - a/(n - 1)\}a/n, \\ &= (-1)^{n-1} \frac{a!}{n!(a - n)!}, \qquad n = 1, 2, \ldots \end{aligned} \tag{11.54}$$

and the pgf is $G_N(z) = \sum_{n=1}^{\infty} p_n z^n = 1 - (1 - z)^a$. This is the limiting form of Engen's extended negative binomial distribution (Section 5.12.2), with parameters (k, w), where $k = -a$ lies in the interval $(-1, 0)$ and $w \to 0$.

Suppose now that M iid sequences of this kind are conducted, where M is a geometric rv with $\Pr[M = m] = (1 - q)q^m$, $0 < q < 1$. Then $X = N_1 + N_2 + \cdots + N_m$ has the discrete Mittag–Leffler pgf

$$G_X(z; 1) = \frac{1}{1 + c(1 - z)^a}, \tag{11.55}$$

where $c = q/(1 - q)$, $0 < c$.

More generally, if M has a negative binomial (b, q) distribution, then X has the pgf

$$G_X(z; k) = \left[\frac{1}{1 + c(1 - z)^a}\right]^b, \qquad 0 < a < 1, \qquad 0 < c, \qquad 0 < b. \tag{11.56}$$

This is the discrete Linnik distribution that has been studied by Devroye (1990) for the case $c = 1$ and Pakes (1995) for $c > 0$.

The limiting form of (11.56) as $b \to \infty$, $c \to 0$ such that $bc = \lambda$ has the pgf

$$G_X(z) = \exp[-\lambda(1 - z)^a], \qquad 0 < a < 1, \qquad 0 < \lambda; \tag{11.57}$$

it is the discrete stable distribution of Steutel and van Harn (1979) and Devroye (1993). Christoph and Schreiber (1998) have explored its properties (including infinite divisibility) in depth.

11.2.6 Discrete Student's *t* Distribution

Type IV of Ord's family of distributions (Section 2.3.3) satisfies the recurrence relationship

$$\Pr[X = x] = P_x = \left(\frac{(x+a)^2 + d^2}{(x+k+a)^2 + b^2} \right) P_{x-1}, \qquad k > 0, \tag{11.58}$$

for all integer values of x (positive, negative, or zero). A special case of this when $d = b$ is the type VII (*discrete Student's t distribution*) for which

$$P_x = \alpha_k \left[\prod_{j=1}^{k} [(j+x+a)^2 + b^2] \right]^{-1}, \qquad -\infty < x < \infty, \tag{11.59}$$

where $0 \le a \le 1$, $0 < b < \infty$, k is a nonnegative integer, and α_k is a normalizing constant.

Ord (1967b,c, 1968b, 1972) has shown that

$$\alpha_k = b \prod_{j=1}^{k} (j^2 + 4b^2) \left[\binom{2k}{k} w(a,b) \right]^{-1}, \tag{11.60}$$

where $w(a,b)$ can be expressed in terms of a and b only, using imaginary parts of the digamma function $\psi(\cdot)$ with a complex argument. For $b = 1$, $w(a,b)$ varies from 3.12988 for $a = 0.5$ to 3.15334 for $a = 0$, 1; for $b \ge 2$, Ord found $|w(a,b) - \pi| < 0.0001$.

Ord found that

$$\begin{aligned}
\mu &= \mu_1' = -\left(\tfrac{1}{2}k + a\right), \\
\mu_2 &= \frac{k^2/4 + b^2}{2k-1}, \\
\mu_4 &= \frac{k^3(k-4)/16 + (3k-2)kb^2/2 + 3b^4}{(2k-1)(2k-3)}.
\end{aligned} \tag{11.61}$$

For every k, there exist moments up to order $2k$.

Ord commented that the discrete Student's t distribution has interest because of its curious property that all finite odd moments about the mean are zero, although the distribution is generally asymmetric; it is symmetric only when $a = 0$, 0.5, 1. The asymmetry has the unusual form

$$P_0 > P_1 > P_{-1} > P_2 > P_{-2} > \cdots.$$

Unlike Roy's (2003) discretized normal distribution, it is not a discretized form of the t distribution. No natural interpretation of the distribution has been found.

11.2.7 Feller–Arley and Gegenbauer Distributions

The Feller–Arley and Gegenbauer distributions have very different modes of genesis, yet mathematically they are closely related,

The *Feller–Arley distribution* arises from a number of stochastic processes and is so called because of the influence of Feller (1950) and Arley (1943) and their researches on the theory of stochastic processes. It is the distribution of the number of individuals in the nth generation for the simple Galton–Watson branching process; in the usual parameterization (e.g., Karlin and Taylor, 1975)

$$G(z) = \left[\frac{nc + (1 - nc)z}{1 + nc - ncz}\right]^k. \tag{11.62}$$

The Feller–Arley distribution also arises from the simple homogeneous birth–death process with k initial individuals; in the usual parameterization (e.g., Bailey, 1964)

$$G(z) = \left[\frac{\mu w(z,t) - 1}{\lambda w(z,t) - 1}\right]^k, \quad \text{where } w(z,t) = \frac{(z-1)e^{(\lambda-\mu)t}}{\lambda z - \mu}. \tag{11.63}$$

In both cases the pgf has the form

$$G(z) = \left(\frac{1 - \theta^* z}{1 - \theta^*}\right)^{U_1} \cdot \left(\frac{1 - \theta z}{1 - \theta}\right)^{U_2}, \tag{11.64}$$

where $U_1 = k = -U_2$, that is, a k-fold convolution of the pgf (3.88) (Kemp, 1979). The formulas (3.82) for the probabilities apply with n_1 replaced by k and n_2 replaced by $-k$. The mean and variance are

$$\mu = k\left(\frac{\theta}{1-\theta} - \frac{\theta^*}{1-\theta^*}\right) \qquad \text{and} \qquad \mu_2 = k\left(\frac{\theta}{(1-\theta)^2} - \frac{\theta^*}{(1-\theta^*)^2}\right); \tag{11.65}$$

see Kemp (1979). The distribution is a convolution of inflated geometric distributions (Section 8.2.3); see Phillips (1978).

The simple homogeneous birth–death–immigration process (Bailey, 1964) has a pgf of the form (11.64), but with exponents $U_1 = k$, $U_2 = -k - \nu/\lambda$; the equations (3.82) for the probabilities and the associated moment properties again apply. The distribution has a very long history dating back to McKendrick (1926); see Irwin (1963).

Plunkett and Jain's (1975) derivation of the *Gegenbauer distribution* mixes a Hermite distribution with pgf

$$G(z) = \exp[\gamma(z-1) + \gamma\rho(z^2 - 1)]$$

(Section 9.4) using a gamma distribution for γ. The outcome has the pgf

$$\begin{aligned}
G(z) &= \int_0^\infty e^{\gamma(z-1)+\gamma\rho(z^2-1)} \frac{\exp(-\gamma\delta)\gamma^{(\ell-1)}\delta^\ell\, d\gamma}{\Gamma(\ell)} \\
&= (1-\xi-\eta)^\ell(1-\xi z-\eta z^2)^{-\ell} \qquad (11.66)\\
&= (1-\xi-\eta)^\ell \sum_{x=0}^\infty \frac{G_x^\ell(\xi,\eta)z^x}{x!}, \qquad (11.67)
\end{aligned}$$

where the support is $x = 0, 1, 2, \ldots$, $\xi = (1+\rho+\delta)^{-1}$, $\eta = \rho(1+\rho+\delta)^{-1}$ and $G_x^\ell(\xi,\eta)$ is a modified Gegenbauer polynomial of order x; for Gegenbauer polynomials, see Rainville (1960, Chapter 17). Plunkett and Jain expressed the probabilities in terms of Gegenbauer polynomials:

$$\begin{aligned}
\Pr[X=x] &= \frac{(1-\xi-\eta)^\ell G_x^\ell(\xi,\eta)}{x!} \\
&= (1-\xi-\eta)^\ell \sum_{j=0}^{[x/2]} \frac{\Gamma(x+\ell-j)\xi^{x-2j}\eta^j}{j!\Gamma(x+1-2j)\Gamma(\ell)}. \qquad (11.68)
\end{aligned}$$

Factorizing the quadratic expressions in the pgf gives

$$\begin{aligned}
G(z) &= (1-\xi-\eta)^\ell(1-\xi z-\eta z^2)^{-\ell} \\
&= \left[\frac{(1-a)(1+b)}{(1-az)(1+bz)}\right]^\ell, \qquad \xi>0, \qquad \eta>0, \qquad \ell>0, \qquad (11.69)
\end{aligned}$$

where $a-b=\xi$, $ab=\eta$, and $0<b<a<1$. The Gegenbauer pgf therefore has the form (11.64) with $U_1 = U_2 = -\ell$, that is, an ℓ-fold convolution of the pgf (3.89), and the equations (3.82) for the probabilities are applicable. The "factorized" notation makes the distribution easy to handle, especially its moment properties; see Kemp (1979).

This is a three-parameter distribution with parameters $\delta, \rho, \ell > 0$, that is, with $\xi, \eta > 0$ and $\xi + \eta < 1$ (since $\delta > 0$). Plunkett and Jain obtained some of the distribution's probability and moment properties as well as some of its limiting forms. They also fitted the distribution to a set of accident data using the method of moments.

Borah (1984) has also studied the probability and moment properties of (11.66) and has used estimation via the first two sample moments and the ratio of the first two sample frequencies (i.e., $\overline{x}$, s^2, and f_1/f_0).

Medhi and Borah's (1984) four-parameter *generalized Gegenbauer distribution* has the pgf

$$G(z) = (1-\xi-\eta)^\ell(1-\xi z-\eta z^m)^{-\ell}; \qquad (11.70)$$

they studied moment estimation and estimation via $\overline{x}$, s^2, and f_1/f_0, assuming a known small integer value for m; see Wimmer and Altmann's (1995) comments on their paper.

Other generalizations of the Gegenbauer distribution are those of Patil and Raghunandanan (1990), who examined, for example, a gamma mixture of stuttering Poisson distributions. This gives the pgf

$$G(z) = \prod_{j=1}^{k} \left(\frac{a_j}{1 + a_j - z^j} \right), \tag{11.71}$$

where $a_j > 0,\ j = 1, 2, \ldots, k$.

Kemp (1992c) has derived (11.69) by a quite different field observation model in which entities occur either as singlets or as doublets and become aware of an observer according to a nonhomogeneous stochastic process; visibility is then determined by Rao damage processes (see Section 9.2 for Rao's damage process).

When $\ell = 1$, various formulas simplify very considerably; the case $\ell = 1$ corresponds to an exponential mixing distribution in Plunkett and Jain's mixed Hermite model and to a nonhomogeneous stochastic awareness process leading to a geometric awareness distribution in Kemp's (1992c) model. Kemp studied various forms of estimation for this special case, including maximum-likelihood estimation.

These distributions are closely related to the convolution of two binomial variables. Ong (1995a) has given references to earlier work on this distribution and obtained three new stochastic formulations.

Minkova (2002) has investigated (in isolation, in the context of risk theory) the sums of a fixed number of iid zero-inflated Poisson, binomial, negative binomial, logarithmic, and geometric variables.

11.2.8 Gram–Charlier Type B Distributions

Helpful approximations to discrete distributions can occasionally be obtained by using distributions formed from the first few terms of a Gram–Charlier type B expansion. [The type A expansions, applicable to continuous distributions, are of considerably wider use; see Johnson, Kotz, and Balakrishnan (1994, Chapter 12).] The basic idea is to express the pmf $\Pr[X = x] = p_x$ as a linear series in the backward differences (with respect to x) of the Poisson probabilities

$$\omega_x = \begin{cases} \dfrac{e^{-\theta}\theta^x}{x!}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{11.72}$$

Thus

$$p_x = \sum_{i=0}^{\infty} a_i \, \nabla^i \omega_x, \tag{11.73}$$

where $\nabla\omega_x = \omega_x - \omega_{x-1}$. Note that, if (11.73) is summed from $x = c$ to $x = \infty$, then

$$\sum_{x=c}^{\infty} p_x = \sum_{i=0}^{\infty} a_i \nabla^i \left(\sum_{x=c}^{\infty} \omega_x \right) \tag{11.74}$$

(i.e., the cumulative sum of the probabilities p_x is expressed in terms of the corresponding cumulative sums of Poisson probabilities). With the definition (11.72), $\sum_{x=c}^{\infty} \omega_x = 1$ for all $c \le 0$.

Since the factorial moments are determined from the pgf, and conversely (provided that the moments exist), we might expect values of p_x to be determined to an adequate degree of approximation by a sufficient number of sample values of $\mu'_{[1]}, \mu'_{[2]}, \ldots$. In fact, if $p_x = 0$ for $x > m$, then

$$p_k = \sum_{x=0}^{m-k} \frac{(-1)^x \mu'_{[k+x]}}{(x!k!)}, \qquad k = 0, 1, 2, \ldots, m,$$

and the value of p_k lies between any two successive partial sums obtained by terminating summation at $x = s$ and $x = s + 1$.

Kendall (1943) demonstrated the use of Gram–Charlier type B series with two, three, and four terms by quoting calculations from Aroian (1937) on the emission of α-particles. The representation of quasi–binomial, Lagrangian Poisson, and quasi–negative binomial distributions by means of modified Gram–Charlier expansions has been studied by Berg (1985).

11.2.9 "Interrupted" Distributions

Interrupted distributions arise under circumstances in which some values of a rv are not observed because of their juxtaposition in time or space to some other value. Consider, for example, a Geiger–Muller counter recording the arrival of radioactive particles. Suppose that there is a constant resolving "dead" time D during which the counter is unable to record any arriving particles. The distribution of the number of arrivals *recorded* in a fixed period of time of length τ will not be the same as that of the actual number of arrivals. Assume that the actual arrivals occur randomly in time at a rate λ (so that the actual number of arrivals in time period τ has a Poisson distribution with mean $\lambda\tau$), that the times T between successive arrivals are independent, and that each interarrival time has the same density function

$$f(t) = \lambda e^{-\lambda t}, \qquad t \ge 0. \tag{11.75}$$

This is the case examined in connection with telephone calls by Erlang (1918) and later by Giltay (1943), Feller (1948), and Feix (1955).

An especially simple case is that in which no particle has been recorded in time D preceding the beginning of a time period of length τ. A full analysis of this case appeared in the first edition of this book. Given that there are $K \ge 1$ actual arrivals [the probability of this is $e^{-\lambda\tau}(\lambda\tau)^K/K!$], then the probability

$p_{k,K}$ that exactly k will be recorded was shown to be

$$p_{k,K} = \binom{K}{k} \left[\left(\frac{kD}{\tau} \right)^{K-k} g^k + k \int_g^{1-(k-1)D/\tau} y^{k-1}(1-y)^{K-k}\, dy \right], \quad (11.76)$$

where

$$g = \begin{cases} 0 & \text{if } k \geq \dfrac{\tau}{D}, \\ 1 - \dfrac{kD}{\tau} & \text{if } k \leq \dfrac{\tau}{D}. \end{cases}$$

The overall probability of recording exactly $k(\geq 1)$ arrivals is

$$p_k = e^{-\lambda\tau} \sum_{K=k}^{\infty} \frac{(\lambda\tau)^K}{K!} p_{k,K}. \quad (11.77)$$

Oliver (1961) has used a model leading to the same distribution to represent traffic counts over a fixed interval of time τ. In this model the distribution of times between successive arrivals of vehicles is assumed to have the density function

$$f(t) = \begin{cases} 0 & \text{if } t < D, \\ \lambda \exp[-\lambda(t-D)] & \text{if } t \geq D. \end{cases}$$

The number of arrivals in time τ then has the distribution (11.77) if the time period starts later than time D after a vehicle passes. Oliver also considered the situation in which the interval starts immediately after the arrival of a vehicle and so obtained a different distribution, which can, however, be obtained from (11.77) by replacing τ by $\tau - D$.

A model of a similar kind has been used by Singh (1964) to represent the distribution of numbers of conceptions over a fixed interval of time τ to married couples in a specified population. Singh assumed that for a time D following a conception no further conception is possible. Instead of representing the time between conceptions as a continuous rv, however, he divided τ into m subintervals of length τ/m. He supposed that, provided that at least $mD\tau^{-1} - 1$ subintervals have elapsed since a subinterval in which there was a conception, the probability of conception in a subinterval is p. ($mD\tau^{-1}$ is taken to be an integer.) The distribution that Singh obtained tends to distribution (11.77) as m increases and p decreases, with $mp = \lambda\tau$. For further work on this topic, see Singh (1968) and Singh, Bhattacharya, and Yadava (1974).

11.2.10 Lost-Games Distributions

The distribution of the number of games lost by the ruined gambler in the classical gambler's ruin problem has the pgf

$$\begin{aligned} G(z) &= z^a \left(\frac{1 - (1-4pqz)^{1/2}}{2qz} \right)^a \\ &= z^a p^a\, {}_2F_1\left[\tfrac{1}{2}a, \tfrac{1}{2}(a+1); a+1; 4pqz\right], \end{aligned} \quad (11.78)$$

$x = a, a+1, \ldots$, where a is the gambler's initial capital (a positive integer) and $q = 1 - p$ is the probability that he or she wins an individual game. This is also the pgf for the distribution of the number of customers served during a busy period of an M/M/1 queue (Poissonian arrivals, exponential service times, and one server), starting with a customers; see Takács (1955) and Haight (1961a). Moreover it is the pgf for the total size of an epidemic; see McKendrick (1926). The relationships between the random walks corresponding to these three stochastic processes were studied by Kemp and Kemp (1968).

A more general form of the *lost-games distribution* has the pgf

$$\begin{aligned} K(z|p, j, a) &= z^j \left(\frac{1 - (1 - 4pqz)^{1/2}}{2qz} \right)^a \\ &= z^j p^a {}_2F_1 \left[\tfrac{1}{2}a, \tfrac{1}{2}(a+1); a+1; 4pqz \right], \end{aligned} \tag{11.79}$$

where $x = j, j+1, \ldots$ and the parameter constraints are $0.5 < p < 1$, $q = 1 - p$, $0 < a$, and j is a nonnegative integer. From (11.79) the pmf is

$$p_x = \Pr[X = x|p, j, a] = \frac{(2x + a - 2j - 1)! a p^{a+x-j} q^{x-j}}{(x + a - j)!(x - j)!}, \qquad x = j, j+1, \ldots, \tag{11.80}$$

where $y!$ is taken to mean $\Gamma(y + 1)$ when y is not an integer. The probabilities can be computed very easily by means of the recurrence relation

$$(x - j)(x + a - j)p_x = (2x + a - 2j - 1)(2x + a - 2j - 2)pqp_{x-1}, \qquad x > j, \tag{11.81}$$

with $p_x = 0$ for $x < j$, $p_j = p^a$. The unimodality of the distribution follows from (11.81).

The mean, variance, and third corrected moment are

$$\mu = j + \frac{aq}{p - q}, \qquad \mu_2 = \frac{apq}{(p - q)^3}, \qquad \mu_3 = \frac{apq(1 + 2pq)}{(p - q)^5}; \tag{11.82}$$

see Haight (1961a) and Kemp and Kemp (1968).

The distribution is a member of the class of modified power series distributions (Section 2.2.2); see Gupta (1984). It is a (possibly shifted) generalized hypergeometric probability distribution (Section 2.4.1) of the kind considered in Section 6.11.2; see Kemp (1968b). Also it belongs to the Gould series family (Section 2.6); see Charalambides (1986a). When $j = 0$, the distribution is infinitely divisible and hence is a Poisson–stopped sum distribution (Section 9.3); the pgf for the distribution whose sum is Poisson–stopped (the cluster size distribution) is given in Kemp and Kemp (1969a).

Let $K(p, j, a)$ denote the lost-games distribution with the parameters p, j, and a as in (11.79). Then, if $X_1 \sim K(p, m, a_1)$ and $X_2 \sim K(p, n, a_2)$ and if X_1 and X_2 are independent, it follows that $X_1 + X_2 \sim K(p, m + n, a_1 + a_2)$.

There are a number of models for the general lost-games distribution. Otter (1949) investigated a multiplicative process characterized by the equation $G(z, w)$

$\equiv zf(w) - w = 0$, where $w = P(z)$ is the pgf of interest; this is the transformation that defines a Lagrangian distribution (see Section 2.5). Example 1 in Otter's paper, with $f(w) = (p + qz)^2$, yields the distribution with pgf $K(z|p, 2, 1)$. The relationships between Otter's branching model, the epidemiological model of Neyman and Scott (1964), and the lost-games distribution were discussed by Kemp and Kemp (1969a), who also studied clustering models where the distribution of clusters is (1) binomial, (2) negative binomial, and (3) Poisson. Kemp and Kemp (1971) also investigated mixing processes; these included mixed negative binomial, mixed Poisson, and mixed confluent hypergeometric models.

The name *inverse binomial distribution* was used by Yanagimoto (1989) for the lost-games distribution with pgf $K(z|p, 0, a)$ because its cgf can be regarded as the inverse function of the cumulant generating function (cgf) of the binomial distribution. If the parameter space for this form of the distribution is extended to include $p < 0.5$, then there is a positive probability that the variate value is infinite. Yanagimoto used this probability to estimate the proportion of discharged patients who can be expected to stay completely free from some disease; in the context of the M/M/1 queue it is the probability that the busy period will never end.

Kemp and Kemp (1992) have developed a state-dependent equilibrium birth-and-death process for the distribution and related this to the size of naturally occurring groups of individuals; they used maximum-likelihood estimation to fit the distribution to data on group sizes.

11.2.11 Luria–Delbrück Distribution

Luria and Delbrück (1943) set up the Luria–Delbrück experiment in a study of "random mutation" versus "directed adaptation." They thought that on the hypothesis of directed adaptation the number of mutants in the replicate cultures that they grew would be Poissonian. On the hypothesis of random mutation, however, some mutations would take place before plating out and consequently the counts would have high variability and skewness. (Their experimental results were markedly non-Poissonian.)

Lea and Coulson (1949) derived the distribution in two ways. First, they assumed that the total population at time t is N_t; they also assumed that during the interval $(t, t + \partial t)$ a nonmutant gives rise to a new nonmutant with probability $\alpha\,\partial t + O(\partial t^2)$ and a nonmutant gives rise to a new mutant with probability $g\,\partial t + O(\partial t^2)$. Each mutant is assumed to give rise to a clone of mutants in such a way that each member of the clone divides into two mutants during the interval $(t, t + \partial t)$ with probability $(\alpha + g)\,\partial t + O(\partial t^2)$. This process yields the differential equation

$$m\frac{\partial G(z)}{\partial t} = m(z-1)G(z) + (z^2 - z)\frac{\partial G(z)}{\partial z},$$

where $G(z)$ is the pgf of the number of mutants per plate. The solution to this equation is

$$G(z) = (1-z)^{m(1-z)/z} = e^{m(1-z)[\ln(1-z)]/z}, \tag{11.83}$$

where m is the product of the population size and the relative mutation rate. This one-parameter distribution with infinite support is called the *Luria–Delbrück distribution*, after the original experimenters. It has interested many probabilists and statisticians.

Lea and Coulson's second derivation makes an extra assumption about the way in which the total population size is increasing; the pgf (11.83) is obtained as a limiting form. Bartlett (1952) obtained partial differential equations for the pgf's $\Pi_1(z_1, z_2)$ and $\Pi_2(z_1, z_2)$ for the normal and mutant populations; these have an exact solution. The pgf for the number of mutants starting with N_0 normal bacteria is $[\Pi(1, z)]^{N_0}$; the Luria–Delbrück pgf is a limiting form of this. Kemp (1994) assumed that the number of mutations that take place before plating out is Poissonian and that each mutant produces a clone of mutants according to a Yule process, where the process parameter varies in a certain way. If the resultant clone size pgf is $g(z) = \sum_{i=1}^{\infty} z^i/[i(i+1)]$, then the pgf for the total number of mutants is (11.83).

Data quoted by Ryan (1952) [in the discussion on Armitage's (1952) read paper] were fitted well by the distribution, but a good fit to data cannot discriminate between models.

Direct computation of the probabilities was awkward before Ma, Sandri, and Sarkar (1991, 1992) found a simple recursive relationship:

$$x \Pr(X = x) = m\left[\frac{\Pr(X = x-1)}{2} + \frac{\Pr(X = x-2)}{3} + \cdots + \frac{\Pr(X = 0)}{x+1}\right], \qquad x \geq 1, \tag{11.84}$$

where $\Pr(X = 0) = e^{-m}$. Derivation of an upper bound for the probabilities has attracted much attention. Ma et al. found that $\Pr[X = x] \approx c/x^2$ for large x and gave an expression for c.

Pakes (1993) commented on work on the distribution by a number of earlier researchers. He pointed out that the distribution is infinitely divisible and that it has infinite moments. Moreover, he proved that $c = m$.

Kemp (1994) showed that

$$x(x+1)\Pr[X = x] < m\left(1 + \frac{11m}{30}\right), \qquad x = 1, 2, \ldots,$$

and hence (inelegantly) found that

$$x \Pr[X \geq x] \approx m. \tag{11.85}$$

Interest in the distribution had resurfaced in *Nature* (Cairns, Overbaugh, and Miller, 1988). Equation (11.85) supported their reasoning about the mutation process in *Escherichia coli*.

Goldie (1995) applied a simple general result from the theory of regularly varying sequences to obtain the asymptotic result (11.85). Prodinger (1996), using

Maple, derived the expansion of $G(z)$ about $z = 1$ and hence deduced that

$$\Pr[X = x] = \frac{m}{x} + m^2 \frac{\log x}{x^2} + \frac{m^2\gamma - m^2 - m}{x^2} + O\left(\frac{\log^2 x}{x^3}\right), \tag{11.86}$$

where γ is Euler's constant.

11.2.12 Naor's Distribution

Naor (1957) studied the following urn model: Suppose that an urn contains n balls of which one is red and the remainder are white. Sampling, with replacement of a white ball (if drawn) by a red ball, continues until a red ball is drawn. Let Y be the requisite number of draws. Then

$$\Pr[Y = y] = \frac{(n-1)\cdots(n-y+1)y}{n^y} = \frac{(n-1)!y}{(n-y)!n^y} \tag{11.87}$$

(after $n - 1$ white draws the urn contains only red balls and so no more than n draws are required). This problem is closely related to problem 6.24 in Wilks (1962, pp. 152–153). Consider a collection of n objects labeled $1, 2, \ldots, n$. An object is chosen and returned, another is chosen and returned, and so on, until an object is chosen that has already been chosen before. Sampling then ceases. The number of draws that are required is the rv $Y + 1$, with Y as above; at least two draws are needed and at most $n + 1$.

The pmf for the number of draws $X = n - Y$ *not* required is

$$\Pr[X = x] = \frac{(n-1)!(n-x)}{x!n^{n-x}}, \qquad x = 0, 1, \ldots, n-1. \tag{11.88}$$

Naor (1957) showed that the pgf of X is

$$G(z) = \frac{n!}{n^n}\left((1-z)\sum_{j=0}^{n-1}\frac{(nz)^j}{j!} + \frac{(nz)^n}{n!}\right) \tag{11.89}$$

and found that when n is large

$$\begin{aligned} \mu &\approx n + \frac{1}{3} - \sqrt{\frac{n\pi}{2}}, \\ \mu_2' &\approx n^2 + \frac{8n}{3} + \frac{1}{3} - (2n+1)\sqrt{\frac{n\pi}{2}}. \end{aligned} \tag{11.90}$$

Kemp (1968b) showed that the pgf of X can be rewritten as

$$G(z) = \frac{{}_1F_1[-n+1; -n; nz]}{{}_1F_1[-n+1; -n; n]} \tag{11.91}$$

and that the corrected moments satisfy the recurrence relation

$$\begin{aligned}\mu_{r+2} + (2\mu - n - 1)\mu_{r+1} + (\mu - n - 1)\mu\mu_r \\ = n\sum_{j=0}^{r}\binom{r}{j}[\mu_{r+1-j} + (\mu - n + 1)\mu_{r-j}];\end{aligned} \tag{11.92}$$

hence

$$\mu_2 = (n + 1 - \mu)(\mu - n) + 2n, \qquad \mu'_{[2]} = n(2\mu + 1 - n).$$

Naor's interest in the distribution arose from a machine-minding problem with n repairmen (Naor, 1956, 1957); his solution of the machine-minding problem involved the rv W that is a convolution of X with a Poisson rv with mean $n\xi$. He showed that W has mean and variance

$$\begin{aligned}E[W] &\approx n\xi + n + \frac{1}{3} - \sqrt{\frac{n\pi}{2}}, \\ \mathrm{Var}(W) &\approx n\xi + \frac{n(4-\pi)}{2} + \frac{2}{9} - \frac{\sqrt{2n\pi}}{6}\end{aligned} \tag{11.93}$$

for n sufficiently large. He hence derived a normal approximation.

11.2.13 Partial-Sums Distributions

Consider a "parent" distribution with support $0, 1, 2, \ldots$ and probabilities $p_0, p_1, p_2, \ldots$ such that $\sum_{j=0}^{\infty} p_j = 1$. From this parent a new distribution can be derived with probabilities proportional to the complement of the parent cdf. This distribution has pmf

$$\Pr[X = x] = (\mu'^{*}_1)^{-1} \sum_{j=x+1}^{\infty} p_j, \qquad x = 0, 1, \ldots. \tag{11.94}$$

(The rth moment about zero and the rth factorial moment of the parent distribution are denoted by μ'^{*}_r and $\mu'^{*}_{[r]}$, respectively.)

The mgf is

$$(\mu'^{*}_1)^{-1}(1 - e^t)^{-1}[1 - \phi^*(t)], \tag{11.95}$$

where

$$\phi^*(t) = \sum_{j=0}^{\infty} p_j e^{jt}. \tag{11.96}$$

The rth moment about zero is

$$\begin{aligned}\mu'_r &= (\mu'^{*}_1)^{-1}\sum_{j=1}^{\infty}\left(p_j\sum_{k=0}^{j-1}k^r\right) \\ &= (\mu'^{*}_1)^{-1}\sum_{i=1}^{\infty}\frac{\Delta^i 0^r \mu'^{*}_{[i+1]}}{(i+1)!}, \qquad r = 1, 2, \ldots.\end{aligned} \tag{11.97}$$

A necessary and sufficient condition for the parent distribution and the descendant distribution to be identical under (11.94) is for them to be geometric; see Wimmer and Altmann (2001).

The distribution that is obtained when the parent distribution is Poisson, with $p_j = e^{-\theta}(\theta^j/j!)$, has been called *Poisson's exponential binomial limit*. For this distribution $\mu'^{*}_{[j]} = \theta^j$ and

$$\begin{aligned} \mu &= \frac{\theta}{2}, \\ \mu_2 &= \frac{\theta}{2} + \frac{\theta^2}{12}, \\ \mu_3 &= \frac{\theta}{2} + \frac{\theta^2}{4}, \\ \mu_4 &= \frac{\theta}{2} + \frac{4\theta^2}{3} + \frac{\theta^3}{4} + \frac{\theta^4}{80}. \end{aligned} \tag{11.98}$$

As $\theta \to \infty$, the moment ratios $\sqrt{\beta_1}$ and β_2 tend to 0 and 1.8, respectively.

Gold (1957) and Gerstenkorn (1962), however, defined *Poisson's exponential binomial limit distribution* as the special case $q = 1$ of the more general distribution with pmf

$$\Pr[X = x] = \frac{q^x e^{-q\omega}}{c} \sum_{j=x}^{\infty} \frac{\omega^j}{j!}, \qquad x = 0, 1, \ldots, \tag{11.99}$$

where c is a normalizing constant.

Bissinger (1965) has constructed systems of distributions derived from a parent distribution by another type of summation process. The *Bissinger system distributions* are defined by

$$\Pr[X = x] = (1 - p_0)^{-1} \sum_{j=x+1}^{\infty} j^{-1} p_j, \qquad x = 0, 1, \ldots. \tag{11.100}$$

It is easily verified that $\sum_{x=0}^{\infty} \Pr[X = x] = 1$, so (11.100) defines a proper distribution. Bissinger gave this class of distributions the name *STER distributions*, from the phrase "Sums of Truncated forms of the Expected value of the Reciprocal" (of a variable having the parent distribution). These distributions arose in connection with an inventory decision problem, with the distribution of the number of demands as the parent distribution. From (11.100),

$$\Pr[X = x - 1] - \Pr[X = x] = (1 - p_0)^{-1} x^{-1} p_x \geq 0, \qquad x \geq 1. \tag{11.101}$$

Hence the successive values of $\Pr[X = x]$ are nonincreasing as x increases.

Bissinger obtained the following relationship with the moments ($\mu_r'^*$) about zero of the parent distribution:

$$\mu_r'^* = (1 - p_0) \sum_{j=0}^{r} \binom{r+1}{j} \mu_j'. \tag{11.102}$$

In particular

$$\begin{aligned} \mu_1' &= \tfrac{1}{2}[(1 - p_0)^{-1}\mu_1'^* - 1], \\ \mu_2' &= \tfrac{1}{6}[(1 - p_0)^{-1}(2\mu_2'^* - 3\mu_1'^*) + 1]. \end{aligned} \tag{11.103}$$

Xekalaki (1983c) has shown that a necessary and sufficient condition for the parent distribution and the descendant distribution to be identical under (11.100) is for them to be the Yule distribution.

Patil and Joshi (1968) gave the following results. Let $G^*(z)$, μ^*, and μ_2^* be the pgf, mean, and variance, respectively, of the parent distribution. Then for the corresponding STER distribution,

$$G(z) = \frac{\theta}{1 - z} \int_z^1 \frac{[G^*(u) - p_0]\,du}{u}, \tag{11.104}$$

$$\mu = \tfrac{1}{2}(\theta\mu^* - 1) \qquad \text{and} \qquad \mu_2 = \tfrac{1}{3}\{\theta[\mu_2^* + (\mu^*)^2] - 1\} - \mu(1 + \mu), \tag{11.105}$$

where $\theta = (1 - p_0)^{-1}$. For work on STER distributions for hypergeometric-type demand distributions, see Kemp and Kemp (1969b) and Section 2.4.2.

We further note that

$$\Pr[X = x] = \sum_{j=x}^{\infty} (j + 1)^{-1} p_j, \qquad x = 0, 1, 2, \ldots, \tag{11.106}$$

defines a proper distribution and that Haight (1961b) listed the distribution with pmf

$$\Pr[X = x] = \frac{\sum_{j=x+k}^{\infty} [p_j/(j - k + 1)]}{\sum_{j=k}^{\infty} p_j}, \tag{11.107}$$

where $x = 0, 1, \ldots$ and $k = 1, 2, \ldots$.

Another distribution that is defined in terms of partial sums is the *discrete Weibull distribution* of Nakagawa and Osaki (1975), who considered the distribution to have useful potential as a failure–time distribution; see Section 11.2.15. Here

$$\Pr[X \geq x] = (q)^{x^\beta}, \qquad 0 < \beta, \qquad 0 < q < 1, \qquad x = 0, 1, 2, \ldots, \tag{11.108}$$

giving

$$\Pr[X = x] = (q)^{x^\beta} - (q)^{(x+1)^\beta}. \tag{11.109}$$

Nakagawa and Osaki noted that when $\beta = 1$ the distribution becomes a geometric distribution. They commented that it is very difficult to obtain useful expressions for the mean and other characteristics of the distribution. Their calculations suggested that it can be J-shaped or unimodal.

Wimmer and Altmann (2001) have investigated a further partial-sums distribution for use in linguistics and musicology. Here the parent distribution has probabilities $p_0, p_1, p_2, \ldots$ and

$$\Pr[X = x] = \frac{k}{x}\sum_{j=x}^{\infty} p_j, \qquad x = 1, 2, \ldots. \tag{11.110}$$

If $\mu^* = \sum_{j=0}^{\infty} jp_j\,(<\infty)$ and $\mu = \sum_{x=0}^{\infty} x\Pr[X = x]\,(<\infty)$, then $k = \mu^*/\mu$. Moreover, if $G^*(z)$ and $G(z)$ are the pgf's of the parent and the descendant distribution, respectively, then

$$G(z) = k\int_0^z \frac{1 - G^*(z)}{1 - \mu}\,du. \tag{11.111}$$

Also

$$\mu'_{[r]} = \frac{k\mu'^{*}_{[r]}}{r} \quad \text{for } r = 1, 2, \ldots. \tag{11.112}$$

Wimmer and Altmann also proved that the only parent distribution that is identical to the descendant distribution, subject to condition (11.110), is the 1-displaced Salvia–Bollinger distribution (Section 5.12.2) with pmf

$$\begin{aligned}\Pr[X = x] &= (-1)^{x-1}\binom{\alpha}{x},\\ &= \frac{\alpha(1-\alpha)(2-\alpha)\cdots(x-1-\alpha)}{x!}, \qquad x = 1, 2, \ldots,\end{aligned}$$

where $0 < \alpha < 1$ and $\Pr[X = 0] = 0$ (Salvia and Bollinger, 1982).

When the parent distribution is geometric, the descendant distribution is the logarithmic distribution.

For a given parent distribution, the corresponding descendant distribution is given unambiguously by (11.110). Wimmer and Altmann commented that the reverse is not true; they pointed out that all parent distributions with probabilities $p_0, p_1, p_2, \ldots$ equal to

$$\left\{\alpha,\ \frac{(1-\alpha)(P_1 - 2P_2)}{P_1},\ \frac{(1-\alpha)(2P_2 - 3P_3)}{P_1}, \ldots\right\},$$

where α is arbitrary, $0 < \alpha < 1$, correspond to the same descendant distribution with probabilities $\Pr[X = 1] = P_1$, $\Pr[x = 2] = P_2, \ldots$.

In the theory of discrete renewal processes the iid random variables $T_1, T_2, \ldots$ can be assumed to take values $1, 2, \ldots$ with $\Pr[T_i = n] = p_n$ and

$$\mu^* = E[T_i] = \sum_{n=1}^{\infty} n p_n,$$

as in Lawler (1995). Let N_j be the number of events that have occurred up to and including time j, that is, $N_j = 0$ if $j < T_1$ and $N_j = \max\{n : T_1 + \cdots + T_n \leq j\}$ otherwise. Let the time elapsed since the last renewal be $A_j = j$ if $j < T_1$ and $A_j = j - (T_1 + \cdots + T_n)$ otherwise. Then A_j is a Markov chain with transition probabilities

$$\Pr[T_i = n | T_i > n-1] = \frac{p_n}{p_n + p_{n+1} + \cdots},$$

that is,

$$p(n, 0) = \frac{p_{n+1}}{p_{n+1} + p_{n+2} + \cdots} \qquad \text{and} \qquad p(n, n+1) = 1 - p(n, 0).$$

The *renewal distribution* is given by

$$\begin{aligned} \Pr[R = n] = \pi_n = p(n-1, n)\pi_{n-1} &= \left(\frac{p_{n+1} + p_{n+2} + \cdots}{p_n + p_{n+1} + \cdots} \right) \pi_{n-1} \\ &= (p_{n+1} + p_{n+2} + \cdots)\pi_0, \qquad n = 0, 1, \ldots. \end{aligned}$$

From $\sum_{n=0}^{\infty} \pi_n = 1$, we have $\pi_0 = 1/\mu^*$ and hence

$$\pi_n = \frac{1}{\mu^*} \sum_{j=n+1}^{\infty} p_j, \qquad n = 0, 1, \ldots. \tag{11.113}$$

This is (11.94) with $\Pr[X = x] = \pi_x$ and $p_0 = 0$.

Nair and Hitha (1989) and Hitha and Nair (1989) have investigated the mutual characterization of R and T; see also the clarification of certain ambiguities in their definition of the failure rate for discrete distributions by Kotz and Johnson (1991).

11.2.14 Queueing Theory Distributions

Here we are concerned not with the more general aspects of the theory of queues but with certain distributions arising naturally from simple models of queueing situations. There is a very extensive literature concerning stochastic processes, with a great amount devoted to queueing theory. For lack of space we mention here only the very accessible introduction to applied stochastic processes by Bhat (2002) and the two substantial volumes by Kleinrock (1975, 1976). Chapter 11 of Bhat (2002) contains a useful short bibliography; Bhat (1978) gave a comprehensive bibliography.

Consider first a stochastic process in which customers arrive, wait in a single queue before being served, take a finite length of time to be served, and then depart. Let $p_n(t)$ denote the probability that at time t the queue size is n (including the customer being served). Suppose that the customers' arrival rate is α_n and that their departure rate is β_n, where α_n and β_n are functions of the queue size n. Then

$$\frac{dp_n(t)}{dt} = \alpha_{n-1}p_{n-1}(t) - (\alpha_n + \beta_n)p_n(t) + \beta_{n+1}p_{n+1}(t). \tag{11.114}$$

When the time-dependent instantaneous probability distribution $\{p_0(t), p_1(t), \dots\}$ tends to an equilibrium distribution as $t \to \infty$, with probabilities $\{p_0, p_1, \dots\}$ that do not depend on the initial probability distribution $\{p_0(0), p_1(0), \dots\}$, we have

$$\begin{aligned} 0 &= -\alpha_0 p_0 + \beta_1 p_1, \\ 0 &= \alpha_{n-1}p_{n-1} - (\alpha_n + \beta_n)p_n + \beta_{n+1}p_{n+1}, \qquad n \geq 1. \end{aligned} \tag{11.115}$$

These equations have the solution

$$p_n = \frac{\alpha_0\alpha_1 \dots \alpha_{n-1}}{\beta_1\beta_2 \dots \beta_n} p_0 = \frac{\alpha_0\alpha_1 \dots \alpha_{n-1}}{\beta_1\beta_2 \dots \beta_n} \Bigg/ \left\{1 + \sum_{n=1}^{\infty} \frac{\alpha_0\alpha_1 \dots \alpha_{n-1}}{\beta_1\beta_2 \dots \beta_n}\right\}. \tag{11.116}$$

This queueing model is closely related to the generalized birth-and-death process with birth rate α_n and death rate β_n, where n is the current population size and the initial population size is unity. If

$$G(z) = \left\{1 + \sum_{n=1}^{\infty} \frac{\alpha_0\alpha_1 \dots \alpha_{n-1}}{\beta_1\beta_2 \dots \beta_n} z^n\right\} \Bigg/ \left\{1 + \sum_{n=1}^{\infty} \frac{\alpha_0\alpha_1 \dots \alpha_{n-1}}{\beta_1\beta_2 \dots \beta_n}\right\} \tag{11.117}$$

is the pgf for the equilibrium population size distribution given one initial individual at time $t = 0$, then $[G(z)]^M$ is the pgf for the corresponding equilibrium population size distribution given M, instead of one, initial individuals. Many of the distributions considered elsewhere in this volume, such as the Poisson, negative binomial, Yule, and lost-games distributions, have modes of genesis of this kind.

Exact results in the nonequilibrium theory of queues and birth-and-death processes are remarkably intransigent to obtain, even for the simplest of situations. Consider, for example, the very simple M/M/1 queueing system with random arrivals (that is, constant arrival rate α), one server, first come–first served queueing discipline, and exponential service times (that is, constant service rate β). Then $\rho = \alpha/\beta$ is said to be the traffic intensity. We have

$$\frac{dp_0(t)}{dt} = -\alpha\, p_0(t) + \beta\, p_1(t), \tag{11.118}$$

$$\frac{dp_n(t)}{dt} = \alpha\, p_{n-1}(t) - (\alpha + \beta)p_n(t) + \beta\, p_{n+1}(t), \qquad n \geq 1. \tag{11.119}$$

The difficulty in solving these equations is that (11.118) is different from the general form (11.119). Nevertheless, much effort has been expended, and several methods of solution have been devised, notably by Bailey (1954), Champernowne (1956), Cox and Smith (1961), and Cohen (1982); for more work on the transient M/M/1 queue, see Parthasarathy (1987), Abate and Whitt (1988), and Syski (1988).

The pmf satisfying (11.118) and (11.119) can be stated in several ways; one of the simpler statements is as an infinite series of modified Bessel functions of the first kind (Section 1.1.5):

$$p_n(t) = e^{-(\alpha+\beta)t}\left(\rho^{(n-M)/2} I_{n-M}(2t\sqrt{\alpha\beta}) + \rho^{(n-M-1)/2} I_{M+n+1}(2t\sqrt{\alpha\beta}) + (1-\rho)\rho^n \sum_{k=n+M+2}^{\infty} \rho^{-k/2} I_k(2t\sqrt{\alpha\beta})\right), \qquad (11.120)$$

where M is the initial queue size and $\rho = \alpha/\beta$.

Mohanty and Panny (1990) created a discrete-time M/M/1 model and obtained its transient solution. They showed that in the limit it agrees with the transient solution for the continuous-time model.

Another type of discrete distribution arising in queueing theory is the distribution of the number of customers served during a busy period for a queue in equilibrium. Consider, first, the distribution of X, the total number of customers served before a queue vanishes, given a single server queue with random arrival of customers at constant rate α and a constant length of time β occupied in serving each customer (i.e., the M/D/1 queue). The assumptions are that the probability of the arrival of a customer during the interval $(t, t + \Delta t)$ is $\alpha\,\Delta t + o(\Delta t)$ and that the probability of the arrival of two (or more) customers in this period is $o(\Delta t)$. If there are initially M customers in the queue, then

$$\Pr[X = x] = \frac{M}{(x-M)!} x^{x-M-1} (\alpha\beta)^{x-M} e^{-\alpha\beta x}, \qquad (11.121)$$

where $x = M, M+1, \ldots$. This is the Borel–Tanner distribution (see Section 7.2.2). For (11.121) to represent a proper distribution, it is necessary to have $\alpha\beta \leq 1$; if $\alpha\beta > 1$, then $\sum_{j=M}^{\infty} \Pr[X = j] < 1$. The distribution was obtained by Borel (1942) for the case $M = 1$ and for general values of M by Tanner (1953).

The corresponding busy-period discrete distribution for the M/M/1 queue described earlier in this section is the lost-games distribution; see Haight (1961a) and Section 11.2.10.

11.2.15 Reliability and Survival Distributions

The variable of interest in survival studies is time, such as time to failure of a piece of equipment in operational research, time to death in actuarial studies, or time to recovery in biostatistics. Very often time is measured as a continuous variable (or is assumed to be a continuous variable when recorded in hours or days). Consequently survival studies have overwhelmingly been concerned with continuous distributions.

Some authors, however [e.g., Barlow and Proschan (1965, 1975), Kalbfleisch and Prentice (1980), Lawless (1982), and Høyland and Rausand (1994)], have recognized the need for discrete survival distributions to model situations where a machine is inspected for failure only once per time period, where equipment such as a light switch is operated discretely, or where time to death of a tagged animal is recorded at the discrete points of time when surveys are made.

Here, as in Section 1.2.4, we assume that lifetimes are nonnegative integers and we adopt the definitions and terminology of Leemis (1995).

Patel (1973) calls these failure distributions. In his little-known paper he catalogs both continuous and discrete distributions according to the nature of their failure rates (hazard functions, forces of mortality). As discrete distributions he lists:

$$\begin{array}{llll}
\text{1. Binomial} & p_t = \binom{n}{t} p^t(1-p)^{n-t} & 0 < p < 1, & \\
 & & \quad t = 0, 1, 2, \ldots, n & \\
\text{2. Poisson} & p_t = \dfrac{\theta^t e^{-\theta}}{t!} & 0 < \theta, t = 0, 1, 2, \ldots & \\
\text{3. Geometric} & p_t = p(1-p)^t & 0 < p < 1, t = 0, 1, 2, \ldots & \\
\text{4. Pascal} & p_t = \binom{-\alpha}{t} p^\alpha (p-1)^t & 0 < p < 1, \alpha > 0, & \\
 & & \quad t = 0, 1, 2, \ldots & (11.122) \\
\text{5. Log series} & p_t = -\dfrac{1}{\ln(1-\theta)} \dfrac{\theta^t}{t} & 0 < \theta < 1, t = 1, 2, \ldots & \\
\text{6. Zeta} & p_t = C \cdot (t)^{-\alpha+1} & \alpha > 0, t = 1, 2, \ldots & \\
 & & \quad (C \text{ is the normalizing} & \\
 & & \quad \text{constant}) & \\
\text{7. Uniform} & p_t = \dfrac{1}{N} & t = 1, 2, \ldots, N & \\
\text{8.} & p_t = \dfrac{1}{t(t+1)} & t = 1, 2, \ldots &
\end{array}$$

He remarked that the *binomial, Poisson*, and *uniform* distributions are IFR (increasing failure rate), the *geometric* has a constant FR, and the *logarithmic, zeta*, and (8) (not named) distributions are DFR (decreasing failure rate); the *Pascal* (negative binomial) is IFR for $\alpha > 1$ and DFR for $0 < \alpha < 1$. Patel gave the failure rates for (7) and (8) as $h_t = 1/(N - t + 1)$ and $h_t = 1/(t + 1)$, respectively, but did not say how he ascertained the failure rate behavior of distributions (1)–(6).

The *discrete Weibull distribution* of Nakagawa and Osaki (1975) has already been mentioned in Section 11.2.13. Nakagawa and Osaki remarked on its suitability for failure data measured in discrete time (e.g., cycles, blows, shocks, and revolutions) and also as an approximation to the continuous Weibull distribution. For this distribution

$$\Pr[T \geq t] = \sum_{j \geq t} p_j = (q)^{t^\beta}, \qquad t = 0, 1, 2, \ldots, \tag{11.123}$$

where $\beta > 0$ and $0 < q < 1$; this gives

$$\Pr[T = t] = p_t = (q)^{t^\beta} - (q)^{(t+1)^\beta}. \tag{11.124}$$

The failure rate is

$$h_t = \frac{p_t}{p_t + p_{t+1} + \cdots} = 1 - \frac{(q)^{(t+1)^\beta}}{(q)^{t^\beta}}; \tag{11.125}$$

the distribution is therefore DFR for $0 < \beta < 1$ and IFR for $\beta > 1$. Nakagawa and Osaki noted that when $\beta = 1$ the distribution becomes a geometric distribution (with constant failure rate).

Kalbfleisch and Prentice (1980) gave the following formulas for lifetime probabilities in terms of the hazard function:

$$S_t = \sum_{j \geq t} p_j = \prod_{j=0}^{t-1} (1 - h_j) \qquad \text{and} \qquad p_t = h_t \prod_{j=0}^{t-1} (1 - h_j) \tag{11.126}$$

(c.f. Section 1.2.4). Salvia and Bollinger (1982) used (11.126) to determine three distributions with specified hazard rates:

- **(a)** If $h_t = c$, then $p_t = c(1 - c)^t$, $t = 0, 1, 2, \ldots$ (a geometric distribution).
- **(b)** If $h_t = c/(t + 1)$, then $p_t = c \prod_{j=1}^{t} (j - c)/(t + 1)!$, $t = 0, 1, 2, \ldots$ (a DFR distribution). This is known as the Salvia–Bollinger distribution. When shifted to support $1, 2, \ldots$, it is the special case $w = 0$ of Engen's extended negative binomial distribution.
- **(c)** If $h_t = 1 - c/(t + 1)$, then $p_t = (t - c + 1)c^t/t!$, $t = 0, 1, 2, \ldots$ (an IFR distribution).

Salvia and Bollinger showed that $\lim_{t\to\infty}\{(1 - h_t)/h_t\} < 1$, where $0 \leq h_t \leq 1$, is a sufficient but not necessary condition for h_t to yield a valid pmf. They found that $\lim_{t\to\infty}\{H_t\} \to \infty$, where $H_t = \sum_{j=0}^{t} h_j$, is a necessary and sufficient condition.

Stein and Dattero's (1984) lifetime distribution was created in a similar way by taking $h_t = pt^\beta$, where $0 < p \leq 1$, $\beta \in \mathbb{R}$. It has support $1, 2, \ldots$ when $\beta \leq 1$ and support $1, 2, \ldots, m$ when $\beta > 1$, where m is the integer part of $1 + p^{-1/(\beta-1)}$; see Wimmer and Altmann (1999).

Xekalaki (1983d) proved that, if $h_t = 1/(a + bt)$, where $a > 0$, $b \in \mathbb{R}$, then there are three possibilities. When $b = 0$, the variable T is geometric. For $b > 0$,

T has a Waring distribution with pmf

$$p_t = \frac{\left(\frac{a-1}{b}\right)\cdots\left(\frac{a-1}{b}+t-1\right)}{a\left(\frac{a}{b}+1\right)\cdots\left(\frac{a}{b}+t\right)}, \qquad t = 0, 1, \ldots, \tag{11.127}$$

and pgf

$$G(z) = \frac{1}{a}\,{}_2F_1\left[1, \frac{a-1}{b}; \frac{a+b}{b}; z\right]; \tag{11.128}$$

see Section 6.10.4. This is a decreasing failure rate distribution; as $t \to \infty$, $h_t \to 0$.

The Yule distribution (Xekalaki, 1983c) is a particular case of the Waring distribution with $a - 1 = b$, $\rho = 1/(a-1) > 0$, shifted to support $1, 2, \ldots$. Here $h_0 = 0$, $h_t = \rho/(\rho + t)$, $t = 1, 2, \ldots$; see Section 6.10.3.

Xekalaki's third possibility is $b < 0$. In order for h_t to be positive, we need $t < (1-a)/b$; the support for T is now $0, 1, \ldots, m$, where m is the integer part of $(1-a)/b$. Time T now has the special case of the negative hypergeometric distribution (Section 6.9.2) with pmf

$$p_t = \frac{(-m)(1-m)\cdots(t-1-m)}{(1-mb)(1-m+b^{-1})\cdots(t-m+b^{-1})}, \qquad t = 0, 1, 2, \ldots, m, \tag{11.129}$$

and pgf

$$G(z) = \frac{1}{1-mb}\,{}_2F_1\left[1, -m; -m+1+b^{-1}; z\right]. \tag{11.130}$$

The failure rate is $h_t = 1/(1 - mb + bt)$. We have $t \le m$ and $b < 0$, so this is an IFR distribution; as $t \to m$, $h_t \to 1$.

The mid-1980s saw a surge of interest, mainly in the reliability literature, concerning classes of lifetime distributions, especially those with bathtub and upside-down failure rates; see, for example, Padgett and Spurrier (1985), Ebrahimi (1986), and Guess and Park (1988).

Adams and Watson (1989) proposed a model with the very flexible hazard rate

$$h_t = G(\xi(t)) = \frac{1}{1 + e^{-\xi(t)}}, \qquad t = 0, 1, \ldots,$$

where $\xi(t) = \theta_0 + \theta_1 t + \cdots + \theta_m t^m$. The special case $m = 0$ gives a geometric distribution. If the polynomial $\xi(t)$ is increasing in t, then the failure rate distribution is IFR. A drawback would seem to be lack of simple expressions for the pmf, survivor function, and moments.

Klar (1999) has also studied this model. He pointed out that, as $t \to \infty$, $S(t) = \prod_{s=o}^{t-1} G(-\xi(s))$ tends to zero only if $\theta_m > 0$ but did not seem to find this a handicap. He investigated maximum-likelihood estimation in depth, both theoretically and numerically, and also testing for goodness of fit. He fitted the model to data on the duration in days of four kinds of atmospheric circulation patterns (ACPs) for spring, summer, autumn, and winter in Central Europe during the period 1851–1898 with $m = 1, 2, 3, 4$. Good fits were obtained. In general $m = 3$ was found to be satisfactory.

Mi (1993) gave the following definitions of bathtub shape and upside-down bathtub shape: A sequence $\{a_i, i \geq 0\}$ of real numbers is said to have a bathtub shape if there exist integers $1 \leq n_1 \leq n_2 < \infty$ such that

$$a_0 > a_1 > a_2 > \cdots > a_{n_1} = \cdots = a_{n_2} < a_{n_2+1} < \cdots;$$

it is said to have an upside-down bathtub shape if there exist integers $1 \leq n_1 \leq n_2 < \infty$ such that

$$a_0 < a_1 < a_2 < \cdots < a_{n_1} = \cdots = a_{n_2} > a_{n_2+1} > \cdots.$$

He also related the shape of the hazard function to that of the mean residual life function; however, his statement of the relationship has been queried. Lai, Xie, and Murthy (2001) revised it for lifetime distributions having support $\{0, 1, \ldots\}$. They supposed that h_t has a bathtub shape with change points n_1 and $n_2 < \infty$ and with $h_0 > 1/\mu$. Then the mean residual life function has an upside-down bathtub shape with a unique change point denoted by k_0, and $k_0 \leq n_1$, or two change points $k_0 - 1$ and $k_0 \leq n_1$.

Bathtub distributions are especially useful when there is an initial burn-in period during which the hazard rate falls followed by a (usually much longer) period when the hazard rate rises as the piece of equipment (animal) ages.

Lai and Wang (1995) proposed the parametric model with

$$p_t = \frac{t^\alpha}{\sum_{j=0}^{N} j^\alpha}, \qquad t = 0, 1, \ldots, N \tag{11.131}$$

(truncation of the series circumvents convergence problems). For $\alpha \leq 0$ it is necessary to set $p_0 = 0$. The corresponding hazard rate is

$$h_t = \frac{t^\alpha}{\sum_{j=t}^{N} j^\alpha}, \qquad t = 0, 1, \ldots, N. \tag{11.132}$$

Lai and Wang proved that the distribution is IFR for $\alpha \geq 0$ and that for $\alpha < 0$ it has a bathtub shape. The mean residual life is decreasing for $\alpha \geq 0$ and upside-down bathtub shaped for $\alpha \leq -0.57$. For $-0.57 < \alpha < 0$ the shape of the mean residual life function depends on the mean of the distribution. Lai and Wang discussed estimation of the parameters N and α and fitted the distribution to Halley's human mortality data.

The monotone properties of the failure rate in the discrete case have been studied in depth by Gupta, Gupta, and Tripathi (1997). Consider a distribution with support $0, 1, \ldots$. Let $\eta(t) = 1 - p_{t+1}/p_t$ and $\Delta\eta(t) = \eta(t+1) - \eta(t) = p_{t+1}/p_t - p_{t+2}/p_{t+1}$. Gupta et al. proved that:

(i) If $\Delta\eta(t) > 0$ (logconcavity), then h_t is nondecreasing (IFR).
(ii) If $\Delta\eta(t) < 0$ (logconvexity), then h_t is nonincreasing (DFR).
(iii) If $\Delta\eta(t) = 0$, then $(p_{t+1})/p_t = (p_{t+2})/(p_{t+1})$ for all t. This implies that when the support is $0, 1, \ldots,$ the distribution is geometric (constant FR).

They illustrated the power of this theorem by applying it to members of the Katz family and to linear inverse failure rate distributions. They also investigated the determination of failure rates for the extended Katz and the Kemp families of distributions.

Roy and Gupta (1999) have studied characterizations of the geometric, Waring, and negative hypergeometric distributions via their reliability properties. They discussed and illustrated the application of their results to model selection and validation.

11.2.16 Skellam–Haldane Gene Frequency Distribution

Skellam (1949) has investigated the following branching process: Suppose that for each individual there are a large number of potential offspring whose chances of survival are independent of one another and that generations do not overlap. Assume also that, if a particular gene occurs once in a parental generation, then the number of times that it occurs in the next generation has a Poisson distribution with pgf $e^{c(z-1)}$, $c < 1$. If the gene frequencies are in equilibrium over generations, the probability distribution in successive generations is determined by

$$G(z) = G(e^{c(z-1)}). \tag{11.133}$$

Differentiating with respect to z and setting $z = 1$ give $\mu = G'(1) = cG'(1)$; this is only satisfied for $c < 1$ by taking $\mu = 0$. In the equilibrium state therefore the gene has died out.

Skellam assumed in addition that the gene arises independently by mutation, according to a Poisson process, at an average rate of λ times per generation; the number of representations of the gene in the next generation now has the pgf

$$G(z) = e^{\lambda(z-1)}G(e^{c(z-1)}). \tag{11.134}$$

Skellam gave the formal solution

$$G(z) = \exp\{\lambda[(z-1) + (e^{c(z-1)} - 1) + (e^{c(e^{c(z-1)}-1)} - 1) + \cdots]\}; \tag{11.135}$$

the distribution is therefore infinitely divisible. The cumulant generating function is

$$\ln[G(e^t)] = \lambda[(e^t - 1) + (e^{c(e^t-1)} - 1) + \cdots]. \tag{11.136}$$

This enabled Skellam to show that the first three cumulants are

$$\begin{aligned} \kappa_1 &= \frac{\lambda}{1-c}, \\ \kappa_2 &= \frac{\lambda}{(1-c)(1-c^2)}, \\ \kappa_3 &= \frac{\lambda(1+2c^2)}{(1-c)(1-c^2)(1-c^3)}. \end{aligned} \tag{11.137}$$

Note that the cumulants exist only for $c < 1$.

Skellam found that a negative binomial with pgf

$$G(z) = \left(\frac{1-k}{1-kz}\right)^{2\lambda/c}, \tag{11.138}$$

where $k = c/(2-c)$, is a first approximation to (11.135). He remarked that this result is a discrete analogy to Kendall's (1948) exact derivation of the negative binomial as the outcome of a stochastic process in continuous time. Truncating the zero class and letting $\lambda \to 0$ give a logarithmic distribution (cf. Fisher et al., 1943).

As a second approximation Skellam adopted a negative binomial distribution with $\mu = \lambda/(1-c)$, $\mu_2 = \lambda/[(1-c)(1-c^2)]$, and pgf

$$G(z) = \left(\frac{1-c^2}{1-c^2z}\right)^{\lambda(1+c)/c^2}. \tag{11.139}$$

He demonstrated the closeness of these approximations for the case $\lambda = 1.2$, $c = 0.6$.

Haldane (1949) derived (11.134) independently. By substituting $s = 1 + t$ in (11.134) he derived a recurrence relationship for the factorial cumulants and hence obtained explicit formulas for the first four cumulants, the skewness, and the kurtosis.

Kemp and Kemp (1997) introduced an alternative contagion model leading to (11.134). They compared the distribution numerically with Skellam's approximation (11.139) and with the Poisson, Delaporte, Pólya–Aeppli, and "Short" distributions (the latter three are all convolutions of a Poisson with an infinitely divisible distribution; see Sections 5.12.5, 9.7, and 9.9, respectively). Overall the Delaporte was the best of the approximating distributions. They fitted the Skellam–Haldane distribution using the first two observed moments and also by maximum likelihood to data on car accident insurance claims.

11.2.17 Steyn's Two-Parameter Power Series Distributions

Multiparameter power series distributions were put forward by Steyn Jr. (1980) as extensions of one-parameter power series distributions and modified power series distributions (see Sections 2.2.1 and 2.2.2). Steyn's distributions have pgf's of the form

$$G(z) = \frac{f\left(\sum_{i=0}^{m} \theta_i z^i\right)}{f\left(\sum_{i=0}^{m} \theta_i\right)}, \tag{11.140}$$

where $f(\lambda z)/f(\lambda)$ is the pgf of a classical one-parameter PSD. Reparameterization so that (11.140) has the form

$$G(z) = \omega + \frac{(1-\omega)F\left(\sum_{i=1}^{m} \theta_i^{\dagger} z^i\right)}{F\left(\sum_{i=1}^{m} \theta_i^{\dagger}\right)} \tag{11.141}$$

appears to have certain advantages for estimation purposes; the estimate of ω depends only on the observed proportion of zeros in the data, and then there remain only m, not $m+1$, parameters to estimate.

Steyn obtained the moments of these distributions via their *exponential moments*—these he defined to be the coefficients of $u^i/i!$ in $\exp[G(1+u)-1]$. He found that

$$\mu = \left(\frac{\sum_{i=0}^{m} i\theta_i}{\sum_{i=0}^{m} \theta_i}\right) \mu_{[1]}^{*}\left(\sum_{i=0}^{m} \theta_i\right),$$

$$\mu_2 = \left(\frac{\sum_{i=0}^{m} i(i-1)\theta_i}{\sum_{i=0}^{m} \theta_i}\right) \mu'^{*}_{[2]}\left(\sum_{i=0}^{m} \theta_i\right) + \left(\frac{\sum_{i=0}^{m} i\theta_i}{\sum_{i=0}^{m} \theta_i}\right)^2 \mu'^{*}_{[1]}\left(\sum_{i=0}^{m} \theta_i\right), \tag{11.142}$$

where $\mu'^{*}_{[r]}(\lambda)$ is the rth factorial moment of the distribution with pgf $f(\lambda z)/f(\lambda)$.

Steyn investigated maximum-likelihood, moment, and other estimation procedures. He was particularly interested in modes of genesis; he suggested cluster models, genesis as a univariate multinomial distribution compounded with a PSD, and a derivation from a multivariate PSD. He commented on certain special forms of the distributions. Their relationship to distributions of order k awaits detailed exploration.

Steyn's (1984) paper concentrated on two-parameter versions of these distributions, in particular

1. the two-parameter Poisson distribution (this is the Hermite distribution of Section 9.4);
2. the two-parameter logarithmic distribution with pgf

$$G(z) = \frac{\ln(1-\theta z - c\theta^2 z^2)}{\ln(1-\theta - c\theta^2)} \tag{11.143}$$

and pmf

$$\Pr[X = x] = -\theta^x \sum_{i=0}^{[x/2]} \frac{(x-i-1)!c^i}{(x-2!)!i!} \Big/ \ln(1-\theta-c\theta^2)$$

$$= \frac{-\theta^x \, {}_2F_1[-x/2, (1-x)/2; -x; -4c]}{x \ln(1-\theta-c\theta^2)} \tag{11.144}$$

(note that $[x/2]$ denotes the integer part of $x/2$); and

3. the two-parameter geometric distribution (this is the Gegenbauer distribution; see Section 11.2.7).

Steyn gave recurrence relations for the probabilities and expressions for the factorial moments of these distributions.

In his 1984 paper he showed that such distributions are the outcome of a certain type of equilibrium birth-and-death process, and he interpreted this as a model for free-forming groups. He used maximum-likelihood estimation to fit a two-parameter logarithmic distribution and a truncated two-parameter Poisson (truncated Hermite) distribution to a group size data set.

11.2.18 Univariate Multinomial-Type Distributions

Univariate multinomial-type distributions have been developed by Steyn (1956); see also Patil et al. (1984).

Suppose, first, that there is a sequence of n independent trials, where each trial has $s+1$ possible outcomes $A_0, A_1, A_2, \ldots, A_s$ that are mutually exclusive and have probabilities $p_0, p_1, p_2, \ldots, p_s$, respectively. Let the occurrence of A_i, $i = 1, 2, \ldots, s$, be deemed to be equivalent to i successes and the occurrence of A_0 be deemed to be a failure. The number of successes X achieved in the n trials has then the *univariate multinomial distribution* with parameters $n, s, p_1, \ldots, p_s$, and pmf

$$\Pr[X = x] = \sum \frac{n!}{r_0!r_1!\cdots r_s!} p_0^{r_0} p_1^{r_1} \cdots p_s^{r_s}, \qquad x = 0, 1, 2, \ldots, ns, \tag{11.145}$$

where summation is over all nonnegative integers $r_0, r_1, \ldots, r_s$, such that $\sum_{i=0}^{s} r_i = n$ and $\sum_{i=0}^{s} i r_i = x$. The parameter constraints are as follows: $n = 1, 2, \ldots$; $s = 1, 2, \ldots$; $0 < p_i < 1$ for $i = 0, 1, 2, \ldots$; and $\sum_{i=0}^{s} p_i = 1$. The pgf is

$$G(z) = (p_0 + p_1 z + p_2 z^2 + \cdots + p_s z^s)^n, \tag{11.146}$$

and the mean and variance are

$$\mu = n \sum_{i=1}^{s} i p_i \qquad \text{and} \qquad \mu_2 = n \sum_{i=1}^{s} i^2 p_i - \frac{\mu^2}{n}. \tag{11.147}$$

The special case $s = 1$ is the binomial distribution.

Consider, second, a sequence of trials of the above type that has been continued until k failures have occurred. Then the number of successes X that have accumulated has the *univariate negative multinomial distribution* with pmf

$$\Pr[X = x] = \sum \frac{(k + r - 1)!}{(k - 1)! r_1! r_2! \cdots r_s!} p_0^k p_1^{r_1} \cdots p_s^{r_s}, \qquad x = 0, 1, 2, \ldots, \tag{11.148}$$

where summation is over all nonnegative integers $r_1, r_2 \ldots, r_s$ such that $\sum_{i=1}^{s} r_i = r$ and $\sum_{i=1}^{s} i r_i = x$. The constraints on the parameters are now as follows: $k = 1, 2, \ldots$; $s = 1, 2, \ldots$; $0 < p_i < 1$ for $i = 0, 1, 2, \ldots$; and $\sum_{i=0}^{s} p_i = 1$. The pgf is

$$G(z) = p_0^k (1 - p_1 z - p_2 z^2 - \cdots - p_s z^s)^{-k}, \tag{11.149}$$

and the mean and variance are

$$\mu = k \sum_{i=1}^{s} \frac{i p_i}{p_0} \qquad \text{and} \qquad \mu_2 = k \sum_{i=1}^{s} \frac{i^2 p_i}{p_0} + \frac{\mu^2}{k}. \tag{11.150}$$

The special case $s = 1$ gives the negative binomial distribution.

Third, consider a finite population of N elements consisting of $s + 1$ classes $A_0, A_1, \ldots, A_s$ containing $N_0, N_1, \ldots, N_s$ elements, respectively, with $N = N_0 + \cdots + N_s$. Let the selection of an element from A_i, $i = 1, 2, \ldots, s$, be treated as i successes and selection of an element from A_0 be treated as a failure. Let n elements be selected at random without replacement from the population. Then the number of successes X that have accrued has the *univariate factorial multinomial distribution* with pmf

$$\Pr[X = x] = \sum \binom{N_0}{r_0} \binom{N_1}{r_1} \cdots \binom{N_s}{r_s} \Big/ \binom{N}{n}, \qquad x = 0, 1, 2, \ldots, \tag{11.151}$$

where summation extends over all nonnegative integers $r_0, r_1, \ldots, r_s$ such that $n = \sum_{i=0}^{s} r_i$ and $x = \sum_{i=0}^{s} i r_i$. The parameter constraints are as follows: $s = 1, 2, \ldots$; $n = 1, 2, \ldots$; $N_i = 1, 2, \ldots$ for $i = 0, 1, 2, \ldots, s$; and $\sum_{i=0}^{s} N_i = N$. The pgf is

$$G(z) = \frac{N_0!(N - n)!}{(N_0 - n)! N!} {}_{s+1}F_1[-n, -N_1, -N_2, \ldots, -N_s; N_0 - n + 1; z, z^2, \ldots, z^s]; \tag{11.152}$$

for an explanation of the notation see Steyn (1956). The mean and variance are

$$\mu = n \sum_{i=1}^{s} \frac{i N_i}{N} \qquad \text{and} \qquad \mu_2 = \frac{N - n}{N - 1} \left[n \sum_{i=1}^{s} \left(\frac{i^2 N_i}{N} \right) - \frac{\mu^2}{n} \right]. \tag{11.153}$$

The classical hypergeometric distribution with parameters n, N, N_1 is the special case with $s = 1$. The univariate factorial multinomial distribution tends to a univariate multinomial distribution as $N \to \infty$, $N_i \to \infty$, such that $N_i/N \to p_i$, $0 < p_i < 1$, $i = 0, 1, \ldots, s$, and $\sum_{i=0}^{s} p_i = 1$.

Fourth, consider the previous setup, but suppose now that sampling continues without replacement until exactly k failures have been obtained (i.e., inverse sampling). Then the number of successes that have accrued has the *univariate negative factorial multinomial distribution* with pmf

$$\Pr[X = x] = \frac{N_0 - k + 1}{N - k - r + 1} \sum \binom{N_0}{k-1}\binom{N_1}{r_1}\cdots\binom{N_s}{r_s} \Big/ \binom{N}{k+r-1}, \tag{11.154}$$

where $x = 0, 1, 2, \ldots, \sum_{i=1}^{s} iN_i$ and summation is over all nonnegative integers $r_1, r_2, \ldots, r_s$ such that $r = \sum_{i=1}^{s} r_i$ and $x = \sum_{i=1}^{s} ir_i$. The constraints on the parameters are as follows: $s = 1, 2, \ldots$; $k = 1, 2, \ldots$; $N_i = 1, 2, \ldots$ for $i = 0, 1, 2, \ldots, s$; and $\sum_{i=0}^{s} N_i = N$. The pgf is

$$G(z) = \frac{N_0!(N-k)!}{(N_0-k)!N!}\, {}_{s+1}F_1[k, -N_1, -N_2, \ldots, -N_s; k - N; z, z^2, \ldots, z^s], \tag{11.155}$$

and the mean and variance are

$$\mu = k\sum_{i=1}^{s} \frac{iN_i}{N_0+1} \qquad \text{and} \qquad \mu_2 = \frac{N_0 - k + 1}{N_0 + 2}\left[k\sum_{i=1}^{s}\left(\frac{i^2 N_i}{N_0+1}\right) + \frac{\mu^2}{k}\right]. \tag{11.156}$$

The inverse hypergeometric distribution with parameters k, N, N_1 is the special case with $s = 1$. The univariate negative factorial multinomial distribution tends to a univariate negative multinomial distribution as $N \to \infty$, $N_i \to \infty$, such that $N_i/N \to p_i$, $0 < p_i < 1$, $i = 0, 1, \ldots, s$, and $\sum_{i=0}^{s} p_i = 1$.

11.2.19 Urn Models with Stochastic Replacements

Urn models with stochastic replacements can be constructed in many ways. Usually we are interested in either (1) the distribution of the number of balls of various kinds in the urn after a fixed number of trials, or (2) the outcome of a fixed number of draws, and/or (3) the (discrete) waiting time until a specified set of conditions are fulfilled.

The concept of an urn model has a very long history dating back to biblical times (Rabinovitch, 1973). Sambursky (1956) has discussed ideas concerning random selection in ancient Greece. The theory underlying urn models dates back at least to Bernoulli's (1713) *Ars Conjectandi* (see, e.g., David, 1962). Of interest is not only the wide variety of discrete distributions that can arise from urn models but also, especially, the understanding that is generated when two apparently quite different situations lead to closely related urn models. Here we do no more than indicate very briefly several different types of urn models and various ways of handling them. Interested readers are referred to Berg's (1988b) review article and to the comprehensive monograph devoted exclusively to urn models by Johnson and Kotz (1977).

For the Pólya urn model (Section 6.2.4) initially containing w white balls and b black ones, the replacement strategy is to return each drawn ball together with

c extra balls of similar color, where c may be negative as well as positive. Let $P(n,x)$ denote the probability of x successes in n trials. Then

$$P(n+1,x+1) = \frac{w+cx}{w+b+cn}P(n,x) + \frac{b+c(n-x)}{w+b+cn}P(n,x+1). \quad (11.157)$$

Friedman's (1949) urn model is a generalization of Pólya's. Again initially the urn is assumed to contain w white balls and b black ones. Also each drawn ball is returned together with c of the same color, but *in addition* d balls of the opposite color are also added at the time of replacement. Friedman noted the following special cases: (1) $c = 0$, $d = 0$ gives a binomial model; (2) $c = -1$, $d = 0$ gives sampling without replacement; (3) $d = 0$ gives the Pólya urn model; (4) $c = -1$, $d = 1$ gives the Ehrenfest model of heat exchange; and (5) $c = 0$ gives a safety campaign model in which each draw of a white ball is penalized. Let W_n be the number of white balls in the urn after n draws and let $X_n = 0, 1$ according to whether the nth drawn ball was white or black. Friedman was able to obtain a formula for the mgf by the use of a difference-differential equation. He showed that there is an explicit solution when $\gamma = (c+d)/(c-d)$ takes the values 0, ± 1, and hence in cases (1)–(5) above. (For $\gamma = 1$ we have $d = 0$, $(c \neq 0)$, and for $\gamma = -1$ we have $c = 0$, $(d \neq 0)$.) Freedman (1965) obtained asymptotic results for the behavior of W_n as n becomes large by treating the sampling scheme as a Markov chain.

A quite different model has been studied by Woodbury (1949) and Rutherford (1954). Here the probability of success depends on the number of previous successes but not on the number of previous trials. The following boundary conditions hold: $P(n,x) = 0$ for $x < 0$ or $x > n$ and $P(0,0) = 1$. Let p_x denote the probability of a success given that x successes have already occurred. Then

$$P(n+1,x+1) = p_x P(n,x) + q_{x+1} P(n,x+1), \quad (11.158)$$

where $q_x = 1 - p_x$.

Woodbury was able to investigate this model by setting up the following generalization of (11.158):

$$P(n+1,x+1) = (q - q_x)P(n,x) + q_{x+1}P(n,x+1). \quad (11.159)$$

Changing from $P(n,x)$ to $F(n,x)$, where

$$P(n,x) = (q-q_0)(q-q_1)\cdots(q-q_x)F(n,x), \quad (11.160)$$

he was able to show after clever manipulation that the solution to (11.159) is

$$P(n,x) = (q-q_0)(q-q_1)\cdots(q-q_x) \times \sum_{i=0}^{x} \frac{q_i^x}{(q-q_0)\cdots(q-q_{i-1})(q-q_{i+1})\cdots(q-q_x)}; \quad (11.161)$$

taking $q = 1$ gives the solution to (11.158).

Rutherford (1954) devoted particular attention to the case $p_x = p + cx$, where $0 < p < 1$. If $c > 0$, the restriction $p + cn < 1$ is needed, whereas if $c < 0$, we require $p + cn > 0$. Under this replacement scheme the composition of the urn is altered only when a white ball is drawn. Using Woodbury's result, Rutherford showed that

$$P(n, x) = \frac{(p/c + x - 1)!}{x!(p/c - 1)!} \sum_{r=0}^{x} (-1)^r \binom{x}{r} (1 - p + cr)^n. \tag{11.162}$$

This is the coefficient of $\omega^n/n!$ in $e^{\omega(1-p)}(1 - e^{-c\omega})^x$.

Rutherford derived expressions for the first three factorial moments of the distribution. By equating their theoretical and observed values, he obtained good fits to two data sets. For Greenwood and Yule's (1920) data on women working on high explosive shells, the urn model has a clear physical meaning. Rutherford also obtained negative binomial and Gram–Charlier approximations to the distribution. He regarded this urn model as a possible explanation for the good descriptive fits that these distributions often provide.

Two other special cases of the Woodbury urn model, with (1) $p_x = (p + cx)/(1 + cx)$ and (2) $p_x = p/(1 + cx)$, have been investigated by Chaddha (1965), who related this choice of functions for p_x to the construction of a model representing number of attendances at a sequence of committee meetings.

Naor's (1957) urn model (see Section 11.2.12) was studied in the context of a machine-minding problem.

Wei (1979) has considered a realistic variant of the play-the-winner rule for use in the assignment of patients who present sequentially in a medical trial concerning k types of treatment. Here, if the response to treatment i is a success, then α (>0) balls of color i are added to the urn; if the response to treatment i is a failure, then β (>0) balls of every other color except color i are added. Wei's analysis of the model assumes that the response of a patient is virtually instantaneous.

Another use of urn models (which may involve the simultaneous use of several urns) is to explain modes of genesis for certain Lagrangian distributions; see, for example, Consul (1974), Consul and Mittal (1975), and Janardan and Schaeffer (1977). Applications of urn models in fields as diverse as genetics, capture–recapture sampling of animal populations, learning processes, and filing systems are discussed in Johnson and Kotz (1977).

11.2.20 Zipf-Related Distributions

The papers by Kulasekera and Tonkyn (1992), Kemp (1995), Zörnig and Altmann (1995), and Doray and Luong (1997) have shown renewed interest in the Lerch family of distributions and its subfamilies of Riemann zeta and Hurwitz zeta distributions.

The Lerch function (Gradsteyn and Ryzhik, 1980; Knopp, 1951; Erdélyi et al., 1953) is

$$\phi(q, \eta, v) = \sum_{x=0}^{\infty} \frac{q^x}{(v+x)^\eta}, \qquad v \neq 0, -1, -2, \ldots. \tag{11.163}$$

For

$$G_{\mathrm{Le}}(z) = \frac{\phi(qz, \eta, v)}{\phi(q, \eta, v)} \tag{11.164}$$

to be a valid pgf with nonnegative probabilities and support $x = 0, 1, 2, \ldots$, we require $q > 0$, $v > 0$. For convergence, given $q > 0$, $v > 0$, we also need either

$$\text{(i) } q = 1, \eta > 1 \qquad \text{or} \qquad \text{(ii) } 0 < q < 1, -\infty < \eta < \infty.$$

The special case $q = 1$, $\eta > 1$, $v = 1$ of (11.164) is called the Riemann zeta function $\zeta(\eta)$; the case $q = 1$, $\eta > 1$, $v \neq 0, -1, -2, \ldots$ gives the Hurwitz zeta function $\zeta(\eta, v)$.

Unfortunately, Zörnig and Altmann (1995) and Doray and Luong (1997) have reversed the order of the second and third arguments in (11.164) in their papers. The book by Wimmer and Altmann (1999) has the arguments in the standard order that is used in mathematical texts; however, their convergence condition $\eta > 0$ is unduly restrictive. [When $0 < q < 1$ and $-a = \eta < 0$, the D'Alembert ratio test,

$$r_x = \left(\frac{x+\nu+1}{x+\nu}\right)^a q, \qquad \lim_{x\to\infty} r_x = q < 1,$$

proves that the series converges.] This point is important concerning the validity of the Kulasekera and Tonkyn (1992) distributions (for which $-\infty < \eta < \infty$).

1. The simplest distribution with pgf of the form (11.165) is the *discrete Pareto distribution*, also known as the *Riemann zeta distribution*. Here the pgf and pmf are

$$G_R(z) = \frac{z\phi(z, \eta, 1)}{\phi(1, \eta, 1)}, \qquad \eta > 1, \tag{11.165}$$

$$\Pr[X = x] = \frac{x^{-\eta}}{\zeta(\eta)} = \frac{x^{-(\rho+1)}}{\zeta(\rho+1)}, \qquad x = 1, 2, \ldots, \tag{11.166}$$

where $\rho = \eta - 1 > 0$, using the notation in the previous editions of this book.

In linguistics, this is often called the Zipf–Estoup law (Estoup, 1916; Zipf, 1949); the distribution is then referred to as the *Zipf distribution*. The use of the distribution in this connection has also been studied by Good (1957). Seal (1947) has applied the distribution to the number of insurance policies held by individuals. Some interesting comments on Zipf and Yule distributions were made by Kendall (1961). The distribution has been used to model scientific productivity; see Wimmer and Altmann (1995, 1999) for references. Fox and Lasker (1983) used it to analyze the distribution of surnames.

When $\eta = 0$ the series $\phi(1, 0, 1)$ does not converge; truncation after n terms gives the discrete rectangular distribution with $\Pr[X = x] = 1/n$, $x = 1, 2, \ldots, n$. Similarly, when $\eta = 1$ the harmonic series does not converge; truncation after n terms gives

$$G_E(z) = \frac{z + z^2/2 + z^3/3 + \cdots + z^n/n}{1 + 1/2 + 1/3 + \cdots + 1/n}, \tag{11.167}$$

$$\Pr[X = x] = \frac{1}{x[\psi(1) - \psi(n+1)]}, \qquad x = 1, 2, \ldots, n, \tag{11.168}$$

where $\psi(\cdot)$ is the psi function (digamma function). This is known as the *Estoup distribution* in the linguistics literature; Wimmer and Altmann (1995, 1999) give references.

If $\eta = 2$, then $\phi(1, 2, 1) = \pi^2/6$; the corresponding pgf and pmf are

$$G_L(z) = 6z\,{}_3F_2[1, 1, 1; 2, 2; z]/\pi^2, \tag{11.169}$$

$$\Pr[X = x] = \pi^2/6x^2, \qquad x = 1, 2, 3, \ldots. \tag{11.170}$$

This is the *Lotka distribution* (Lotka, 1926).

Consider now the general case of the Riemann zeta distribution with $\eta > 1$, that is, $\rho > 0$. For $r < \rho$, the rth moment about zero is

$$\mu'_r = \frac{\zeta(\rho - r + 1)}{\zeta(\rho + 1)}. \tag{11.171}$$

If $r \geq \rho$, the moment is infinite.

Very often ρ is found to have a value slightly in excess of unity. For low values of ρ the distribution has a very long positive tail. Table 11.1 compares Riemann zeta, logarithmic, and shifted geometric distributions with the same expected value, $\mu = 2$, and shows that the Riemann zeta distribution decays more rapidly at first, but later more slowly.

Schreider (1967) has attempted to give a theoretical justification for the Zipf distribution based on thermodynamic analogies. His arguments would lead one to expect the Zipf distribution to be approached asymptotically for "sufficiently long" extracts of "stable" text. They also indicate that greater deviations from the Zipf form may be expected at lower, rather than higher, frequencies; this agrees with empirical findings in a considerable number of cases. Nanapoulos (1977) has established a weak law of large numbers for the distribution. Devroye (1986) put forward an interesting method for the computer generation of rv's from the Zipf (Riemann zeta) distribution.

Table 11.1 Comparison of Riemann Zeta, Logarithmic, and Shifted Geometric Distributions, Each with Mean 2

		$\Pr[X = x]$	
x	Riemann Zeta[a]	Logarithmic[b]	Geometric[c]
1	0.7409	0.5688	0.5000
2	0.1333	0.2035	0.2500
3	0.0488	0.0971	0.1250
4	0.0240	0.0521	0.0625
5	0.0138	0.0298	0.0312
6	0.0088	0.0178	0.0156
7	0.0060	0.0109	0.0078
8	0.0043	0.0068	0.0039
9	0.0032	0.0044	0.0020
10	0.0025	0.0028	0.0010
>10	0.0144	0.0060	0.0010

[a] $\Pr[X = x] = 0.74088x^{-2.4749}$ (Riemann zeta),
[b] $\Pr[X = x] = 0.79475\ (0.71563)^x/x$ (logarithmic),
[c] $\Pr[X = x] = 2^{-x}$ (geometric).

Given values of N independent rv's each having the Riemann zeta (Zipf) distribution, the maximum-likelihood estimator $\hat{\rho}$ of ρ satisfies the equation (Seal, 1952)

$$N^{-1}\sum_{i=1}^{N}\ln(x_i) = \frac{-\zeta'(\hat{\rho}+1)}{\zeta(\hat{\rho}+1)}. \tag{11.172}$$

Using appropriate tables such as those given in the second edition of this book, or otherwise, it is not difficult to obtain an acceptable solution of this equation. If $\rho > 4$, then $\zeta'(\rho+1)/\zeta(\rho+1) \approx -(1+2^{\rho+1})^{-1}\ln 2$. The variance of $\hat{\rho}$ is approximately $[Ng(\rho)]^{-1}$, where

$$g(\rho) = \frac{d}{d\rho}\left(\frac{\zeta'(\rho+1)}{\zeta(\rho+1)}\right). \tag{11.173}$$

Fox and Lasker (1983) used maximum-likelihood estimation in their analysis of surname frequencies.

An alternative, though usually considerably poorer, estimation method is based on the ratio f_1/f_2 of the observed frequencies of 1's and 2's among the $x_1, x_2, \ldots, x_N$. Equating f_j/N to $\Pr[X = j]$ for $j = 1, 2$, we obtain

$$\rho^* = \frac{\ln(f_1/f_2)}{\ln 2} - 1. \tag{11.174}$$

The variance of ρ^* is approximately

$$\frac{1+2^{\rho+1}}{(\ln 2)^2\zeta(\rho+1)N}.$$

(Note that this method ignores the possibility that $f_1 = 0$ or $f_2 = 0$.) A better estimator than ρ^* appears to be obtained by equating the population and sample means. This gives $\tilde{\rho}$, where $\tilde{\rho}$ satisfies the equation

$$\overline{x} = \frac{\zeta(\tilde{\rho})}{\zeta(\tilde{\rho}+1)}.$$

Moore (1956) provided tables to facilitate the solution of this equation. When $\rho = 3$, $\mathrm{Var}(\tilde{\rho}) \doteqdot 23.1/N$, whereas $\mathrm{Var}(\hat{\rho}) \doteqdot 17.9/N$, a ratio of about 1.3.

2. Relaxation of the restriction $v = 1$ gives a second subfamily of distributions. These are based on the Hurwitz zeta function and have pgf's and pmf's of the form

$$G_H(z) = \frac{z\phi(z,\eta,v)}{\phi(1,\eta,v)}, \qquad \eta > 1, \qquad v > 0, \tag{11.175}$$

$$\Pr[X = x] = \frac{(x+v)^{-\eta}}{\zeta(\eta, v+1)}, \qquad x = 1, 2, 3, \ldots, \tag{11.176}$$

and are called *Hurwitz distributions*. They have been used for ranking problems in linguistics and in the analysis of publication and citation frequencies; see Zipf (1949), Mandelbrot (1959), and Wen (1980).

The series $\phi(1, 1, v)$ does not converge; in this situation a truncated version with

$$\Pr[X = x] = \frac{(x+v)^{-1}}{\zeta(1, v+1) - \zeta(1, n+v+1)}, \qquad x = 1, 2, \ldots, n, \tag{11.177}$$

has been constructed; it is known as the *Zipf–Mandelbrot distribution*; Wimmer and Altmann give references.

3. Consider now the case $v = 1$, $0 < q < 1$, $-\infty < \eta < \infty$. This gives the *Good distribution* with pgf and pmf

$$G_H(z) = \frac{z\phi(qz,\eta,1)}{\phi(q,\eta,1)}, \qquad v > 0, \tag{11.178}$$

$$\Pr[X = x] = \frac{q^x}{x^\eta \phi(q,\eta,1)}, \qquad x = 1, 2, 3, \ldots, \qquad 0 < q < 1 \tag{11.179}$$

(Good, 1953, 1957). This is a size-biased logarithmic distribution. [The function $F(q,\eta) = q\phi(q,\eta,1)$ was studied in depth by Truesdell (1945)]. Taking $\eta = 0$ gives the geometric distribution with support $1, 2, 3, \ldots$, while $\eta = 1$ gives the logarithmic distribution. The case $\eta = 2$ gives a distribution involving Euler's dilogarithmic function. The shifted negative binomial distribution with pgf $z(1-q)^2/(1-qz)^2$ is the outcome when $\eta = -1$.

Ijiri and Simon (1977) used the distribution to model the sizes of business firms. Kemp (1995), in a reconsideration of the work by Yule (1925) and Williams (1964) on numbers of species per genus, took $\eta > 1$ and called the distribution a *polylogarithmic* distribution. She obtained maximum-likelihood fits for a large number of data sets by searching the likelihood surface; the "polylogarithmic" fits were generally better than the logarithmic fits, especially concerning the first frequency.

Kulasekera and Tonkyn (1992) studied the pmf (11.178) in depth. Their parameterization uses q and $\alpha\,(= -\eta)$. They found that $\eta > 0$ gives monotonically decreasing probabilities and that when $\eta < 0$, $q > 2^{\eta}$ there is a (single) mode away from the origin. They were interested in the distribution for modeling survival processes and proved that its hazard function (failure rate) is monotonically decreasing, constant, or monotonically increasing according as $\eta > 0$, $\eta = 0$, or $\eta < 0$. They discussed the computation of $b(\alpha, q) = \phi(q, -\eta, 1)$ and gave recurrence formulas for the moments and factorial moments in terms of its derivatives. In particular

$$\mu = \frac{\phi(q, 1-\eta, 1)}{\phi(q, -\eta, 1)} \qquad \text{and} \qquad \mu_2 = \frac{\phi(q, 2-\eta, 1)}{\phi(q, -\eta, 1)} - \mu^2. \tag{11.180}$$

The distribution is over- or underdispersed according as

$$\frac{\phi(q, 2-\eta, 1)}{\phi(q, 1-\eta, 1)} - \frac{\phi(q, 1-\eta, 1)}{\phi(q, -\eta, 1)} \gtrless 1.$$

Kulasekera and Tonkyn proposed three ad hoc methods of parameter estimation and fitted several data sets.

Doray and Luong (1997) commented on the great flexibility and useful properties of the Good distribution and suggested that applications of it may have been hampered by difficulties with parameter estimation. They discussed maximum-likelihood estimation and introduced and implemented a new quadratic distance method of estimation.

4. The geometric, Zipf, and Good distributions all belong to the Lerch family and also to the class of *maximum-entropy probability distributions* (MEDs) of Kapur (1989); see also Lisman and van Zuylen (1972). For discrete distributions these are distributions that have maximum Shannon entropy, $-\sum_x p_x \ln p_x$, where $p_x = \Pr[X = x]$, subject to certain restrictions. Kapur has shown that the geometric distribution is the MED with fixed arithmetic mean, the Zipf (Riemann zeta) distribution is the MED with fixed geometric mean, and the Good distribution is the MED with both the arithmetic mean and the geometric mean predetermined. He refers to the Good distribution as the "generalized geometric" distribution. In each case the support of the distribution is assumed to be $1, 2, 3, \ldots$.

5. Good (1957) put forward a second distribution for modeling long-tailed data. It differs from the distributions just mentioned in that the pmf and pgf are

$$\Pr[X = x] = \frac{q^x}{(x+v-1)\phi(q,1,v)}, \qquad x = 1, 2, 3, \ldots, \tag{11.181}$$

$$\begin{aligned} G_L(z) &= \frac{z\phi(qz,1,v)}{\phi(q,1,v)} \\ &= \frac{z\,{}_2F_1[1, v; v+1; qz]}{{}_2F_1[1, v; v+1; q]}, \qquad 0 < q < 1, \qquad v > 0. \end{aligned} \tag{11.182}$$

This resembles an extended Waring distribution; it is not, however, an extended beta–Pascal distribution (for which the hypergeometric series would have the form ${}_2F_1[k, l; m; qz]$, $m > 0$).

6. The *discrete Pearson III distribution* of Haight (1957) appears to be the only distribution in the literature with all three parameters in $\phi(q, \eta, v)$ not equal to unity. Here the support is $0, 1, 2, \ldots$, and the pmf and pgf are

$$G_L(z) = \frac{\phi(qz,\eta,v)}{\phi(q,\eta,v)}, \qquad 0 < q < 1, \qquad v > 0, \tag{11.183}$$

$$\Pr[X = x] = \frac{e^{-\beta x}}{(x+v)^{\eta}\phi(e^{-\beta,\eta,v)}}, \qquad \beta > 0, \quad \text{i.e., } 0 < q = e^{-\beta} < 1, \quad v > 0. \tag{11.184}$$

The distribution exists for $-\infty < \eta < \infty$. There do not seem to be applications of this general form of the Lerch distribution in the literature.

7. We mention also the Riemann zeta distribution of Lin and Hu (2001), who define it to be the infinitely divisible distribution with characteristic function

$$\varphi(t) = \frac{\zeta(\sigma + it)}{\zeta(\sigma)}, \qquad \sigma > 1, \qquad t \text{ real}, \tag{11.185}$$

and support $-\ln n$, $n = 1, 2, \ldots, \infty$. Transforming the variable gives the usual Riemann zeta (discrete Pareto) distribution.

8. The paper by Yeh (2002) is mostly about multivariate Zipf distributions. The univariate version of his Zipf (IV) distribution has the pmf

$$\Pr[X \geq k] = \left[1 + \left(\frac{k}{\sigma}\right)^{1/\gamma}\right]^{-\alpha}, \qquad k = 0, 1, 2, \ldots. \tag{11.186}$$

When $\sigma = \gamma = \alpha = 1$, $\Pr[X = x] = (x+1)^{-1}(x+2)^{-1}$; Arnold and Laguna (1977) call this the *"standard" Zipf* distribution.

11.2.21 Haight's Zeta Distributions

Haight's zeta distributions are unlike the Lerch distributions but are nevertheless related to the Riemann zeta function

$$\zeta(s) = \sum_{j=1}^{\infty} j^{-s}, \qquad s > 1.$$

Haight (1966, 1969) returned to Zipf's conjecture concerning city sizes (cf. Simon, 1955), by considering a theoretical tabulation of categories $x = 1, 2, 3, \ldots,$ with category x containing N_x entities such that

$$x - 0.5 \leq ZN_x^{-\beta} < x + 0.5, \tag{11.187}$$

where $Z \geq 1$, $\beta > 0$ are constants. Rearranging the inequalities gives

$$N_x = \left[\left(\frac{2Z}{2x-1}\right)^{1/\beta}\right] - \left[\left(\frac{2Z}{2x+1}\right)^{1/\beta}\right]; \tag{11.188}$$

note that throughout this section $[\cdot]$ means the integer part. Haight argued that a discrete probability distribution can be constructed by taking

$$\Pr[X = x] = \frac{N_x}{(2Z)^{1/\beta}}, \qquad x = 1, 2, \ldots, \tag{11.189}$$

provided that $(2Z)^{1/\beta}$ is an integer. We note that the restriction on Z is necessary in order that

$$\sum_{x \geq 1} \Pr[X = x] = \left[\left(\frac{2Z}{1}\right)^{1/\beta}\right](2Z)^{-1/\beta} = 1. \tag{11.190}$$

Moreover, for this zeta model to be meaningful, Z must be an integer.

Haight's zeta distribution is obtained by letting $Z \to \infty$ and writing $\sigma = 1/\beta$. This gives

$$\Pr[X = x] = (2x-1)^{-\sigma} - (2x+1)^{-\sigma}, \qquad x = 1, 2, \ldots, \qquad \sigma > 0. \tag{11.191}$$

The name "zeta" was adopted because the first two moments of the distribution can be expressed in terms of the Riemann zeta function:

$$\mu = \sum_{x \geq 1} x \Pr[X = x] = (1 - 2^{-\sigma})\zeta(\sigma), \tag{11.192}$$

$$\mu_2 = \sum_{x \geq 1} x^2 \Pr[X = x] - \mu^2 = (1 - 2^{1-\sigma})\zeta(\sigma - 1) - \mu^2. \tag{11.193}$$

The relationship to the zeta function is here via the mean and variance, not the probabilities. For $\sigma \le 1$ the mean is infinite, and for $\sigma \le 2$ the variance is infinite.

When $\beta = \sigma = 1$ and Z is finite, the outcome is *Haight's harmonic distribution*, with

$$\Pr[X = x] = \frac{1}{2Z}\left\{\left[\frac{2Z}{2x-1}\right] - \left[\frac{2Z}{2x+1}\right]\right\}, \tag{11.194}$$

where $2Z$ is a positive integer and the support of the distribution is $x = 1, 2, \ldots, \left[Z + \frac{1}{2}\right]$. Patil et al. (1984) have given the following formulas for the mean and variance:

$$\mu = \sum_{x=1}^{[Z+1/2]} \left[\frac{2Z}{2x-1}\right](2Z)^{-1}, \tag{11.195}$$

$$\mu_2 = \sum_{x=1}^{[Z+1/2]} \left(\frac{2x-1}{2Z}\right)\left[\frac{2Z}{2x-1}\right] - \mu^2. \tag{11.196}$$

Haight gave tables of the distribution for $Z = 1, 5, 10, 20, 30, 100, 1000$. This is a terminating distribution. It has the unusual feature that $\Pr[X = x]$ is equal to zero for considerable ranges of values of x, interspersed with isolated nonzero probabilities. As Z tends to infinity, $\Pr[X = x]$ tends to $2(2x-1)^{-1}(2x+1)^{-1}$, $x = 1, 2, \ldots$. (Similarity to the Yule distribution with $\rho = 1$ is evident.) The mean is infinite.

Note that Haight's zeta distribution and the harmonic distribution each have a single parameter.

The four data sets on word associations in Haight's (1966) paper were fitted very successfully with Haight's zeta distribution by equating the observed and expected first frequencies. Haight did not try to fit the harmonic distribution to these data sets; with hindsight we can see that equating the observed and expected first frequencies would give the estimator Z^* satisfying

$$\frac{f_1}{N} = 1 - \frac{1}{2Z^*}\left[\frac{2Z^*}{3}\right], \tag{11.197}$$

which does not have a solution if $f_1/N \le \frac{2}{3}$ (as in all four data sets). In Haight (1969) the author attempted to fit both his zeta distribution and the harmonic distribution to two data sets on population sizes by equating observed and expected means. Because moment estimation of the parameter σ for Haight's zeta distribution requires evaluation of the Riemann zeta function, he recommended the use of the approximation

$$\begin{aligned}(1 - 2^{-\sigma})\zeta(\sigma) = {}& 0.5(\sigma-1)^{-1} + 0.635518142 + 0.11634237(\sigma-1)\\ &- 0.01876574(\sigma-1)^2\end{aligned} \tag{11.198}$$

for σ close to unity.

Bibliography

Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution, *Annals of Applied Probability*, **2**, 951–972. [8.3.3]

Abakuks, A. (1979). An optimal hunting policy for a stochastic logistic model, *Journal of Applied Probability*, **16**, 319–331. [6.11.2]

Abate, J., and Whitt, W. (1988). Transient behaviour of the M/M/1 queue via Laplace transforms, *Advances in Applied Probability*, **20**, 145–178. [11.2.14]

Abdel-Aty, S. H. (1954). Ordered variables in discontinuous distributions, *Statistica Neerlandica*, **8**, 61–82. [4.4]

Abdul-Razak, R. S., and Patil, G. P. (1986). Power series distributions and their conjugates in statistical modeling and Bayesian inference, *Communications in Statistics—Theory and Methods*, **15**, 623–641. [2.2.1]

Abouammoh, A. M. (1987). On discrete α-unimodality, *Statistica Neerlandica*, **41**, 239–244. [1.2.7]

Abouammoh, A. M., and Mashour, A. F. (1981). A note on the unimodality of discrete distributions, *Communications in Statistics—Theory and Methods*, **A10**, 1345–1354. [1.2.7]

Abramowitz, M., and Stegun, I. A. (1965). *Handbook of Mathematical Functions*, New York: Dover. [1.1.2, 1.1.3, 1.1.5, 1.1.7, 1.1.9, 1.1.11]

Achuthan, N. R., and Krishnan, T. (1992). EM algorithm for segregation analysis, *Biometrical Journal*, **34**, 971–988. [5.8.5]

Aczél, J. (1972). On a characterization of the Poisson distribution, *Journal of Applied Probability*, **9**, 852–856. [4.8]

Aczél, J. (1975). On two characterizations of Poisson distributions, *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, **44**, 91–100. [4.8]

Adamidis, K. (1999). An EM algorithm for estimating negative binomial parameters, *Australian and New Zealand Journal of Statistics*, **41**, 213–221. [5.8.5]

Adams, G., and Watson, R. (1989). A discrete time parametric model for the analysis of failure time data, *Australian Journal of Statistics*, **31**, 365–384. [11.2.15]

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

Adatia, A. (1991). Computation of variances and covariances of order statistics from the geometric distribution, *Journal of Statistical Computation and Simulation*, **39**, 91–94. [5.2]

Agrawal, N., and Smith, S. (1996). Estimating negative binomial demand for retail inventory management with unobservable lost sales, *Naval Research Logistics*, **43**, 839–861. [5.10]

Agresti, A., and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions, *American Statistician*, **52**, 119–126. [1.2.16]

Ahmed, A. N. (1991). Characterization of beta, binomial and Poisson distributions, *IEEE Transactions on Reliability*, **R-40**, 290–295. [3.9]

Ahmed, M. S. (1961). On a locally most powerful boundary randomized similar test for the independence of two Poisson variables, *Annals of Mathematical Statistics*, **32**, 809–827. [9.4]

Ahmed, S. E. (1991). Combining Poisson means, *Communications in Statistics—Theory and Methods*, **20**, 771–789. [4.7.2]

Ahrens, J. H. (1987). A comparison of hypergeometric distributions with corresponding binomial distributions, *Ökonomie und Mathematik*, O. Opitz and B. Rauhut (editors), 253–265. Berlin: Springer-Verlag. [6.5]

Ahrens, J. H., and Dieter, U. (1974). Computer methods for sampling from gamma, beta, Poisson and binomial distributions, *Computing*, **12**, 223–246. [3.7.2]

Ahrens, J. H., and Dieter, U. (1982). Computer generation of Poisson deviates from modified normal distributions, *ACM Transactions on Mathematical Software*, **8**, 163–179. [4.6.2]

Ahsanullah, M., and Holland, B. (1984). Record values and the geometric distribution, *Statistische Hefte*, **25**, 319–327. [5.9.1]

Ahuja, J. C. (1971a). Certain properties of the Stirling distribution of the second kind, *Australian Journal of Statistics*, **13**, 133–136. [4.12.3]

Ahuja, J. C. (1971b). Distribution of the sum of independent decapitated negative binomial variables, *Annals of Mathematical Statistics*, **42**, 383–384. [5.11]

Ahuja, J. C., and Enneking, E. (1972a). Certain properties of the distribution of the sum of independent zero-one-truncated Poisson variables, *Australian Journal of Statistics*, **14**, 50–53. [4.12.3]

Ahuja, J. C., and Enneking, E. (1972b). Recurrence relation for the minimum variance unbiased estimation of left-truncated logarithmic series distribution, Abstract, *Bulletin of the Institute of Mathematical Statistics*, **1**, 137. [7.1.10]

Ahuja, J. C., and Enneking, E. A. (1974). Convolution of independent left-truncated negative binomial variates and limiting distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **26**, 265–270. [5.11]

Aitchison, J., and Brown, J. A. C. (1957). *The Lognormal Distribution*, Cambridge: Cambridge University Press. [11.1.4]

Aitken, A. C. (1939). *Statistical Mathematics* (first edition), Edinburgh: Oliver and Boyd. [3.12.3]

Aitken, A. C. (1945). *Statistical Mathematics* (fourth edition), Edinburgh: Oliver and Boyd. [3.12.3]

Aki, S. (1985). Discrete distributions of order k on a binary sequence, *Annals of the Institute of Statistical Mathematics, Tokyo*, **37**, 205–224. [10.7.2, 10.7.3, 10.7.6]

Aki, S., Balakrishnan, N., and Mohanty, S. G. (1996). Sooner and later waiting time problems for success and failure runs in higher order Markov dependent trials, *Annals of the Institute of Statistical Mathematics, Tokyo*, **48**, 773–787. [10.7.2]

Aki, S., and Hirano, K. (1988). Some characteristics of the binomial distribution of order k and related distributions, *Statistical Theory and Data Analysis*, Vol. 2, K. Matusita (editor), 211–222. Amsterdam: Elsevier. [10.7.1, 10.7.5]

Aki, S., and Hirano, K. (1989). Estimation of parameters in the discrete distributions of order k, *Annals of the Institute of Statistical Mathematics, Tokyo*, **41**, 47–61. [10.7.5]

Aki, S., Kuboki, H., and Hirano, K. (1984). On discrete distributions of order k, *Annals of the Institute of Statistical Mathematics, Tokyo*, **36**, 431–440. [10.7.1, 10.7.4]

Alanko, T., and Duffy, J. C. (1996). Compound binomial distributions for modelling consumption data, *The Statistician*, **45**, 269–286. [10.4.3]

Al-Awadhi, S. A., and Ghitany, M. E. (2001). Statistical properties of Poisson–Lomax distribution and its statistical applications to repeated accident data, *Journal of Applied Statistical Science*, **10**, 365–372. [8.3.3]

Al-Saleh, M. F., and Al-Batainah, F. K. (2003). Estimation of the proportion of sterile couples using the negative binomial distribution, *Journal of Data Science*, **1**, 261–274. [5.10]

Alt, F. B. (1985). The matching problem, *Encyclopedia of Statistical Sciences*, Vol. 5, S. Kotz, N. L. Johnson, and C. B. Read (editors), 293–296. New York: Wiley. [10.3]

Altham, P. M. E. (1978). Two generalizations of the binomial distribution, *Applied Statistics*, **27**, 162–167. [3.12.6]

Altmann, G. (1993). Phoneme counts, *Glottometrika*, **14**, 54–68. [6.10.4]

Al-Zaid, A. A. (1989). On the unimodality of mixtures, *Pakistan Journal of Statistics*, **5**, 205–209. [8.3.4]

Alzaid, A. A., Rao, C. R., and Shanbhag, D. N. (1986). Characterization of discrete probability distributions by partial independence, *Communications in Statistics—Theory and Methods*, **15**, 643–656. [4.8]

Anderson, T. W. (1943). On card matching, *Annals of Mathematical Statistics*, **14**, 426–435. [2.4.2, 10.3]

Anderson, T. W., and Samuels, S. M. (1967). Some inequalities among binomial and Poisson probabilities, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 1–12. Berkeley: University of California Press. [3.6.1, 4.4]

Andrews, G. E. (1986). *q Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra*, Providence, RI: American Mathematical Society. [1.1.12]

Andrews, G. E. (1998). *The Theory of Partitions*, Cambridge: Cambridge University Press. [10.8.1]

Angus, J. E., and Schafer, R. E. (1984). Improved confidence statements for the binomial parameter, *American Statistician*, **38**, 189–191. [3.8.3]

Anraku, K., and Yanagimoto, T. (1990). Estimation for the negative binomial distribution based on the conditional likelihood, *Communications in Statistics—Simulation and Computation*, **19**, 771–806. [5.8.3, 5.8.5]

Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative binomial data, *Biometrika*, **35**, 246–254. [3.6.3, 4.5, 5.6]

Anscombe, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions, *Biometrika*, **37**, 358–382. [5.1, 5.8.4, 5.8.5, 7.1.2, 7.1.7, 7.1.12, 9.13, 11.1.4]

Antzoulakos, D. L. (1999). On waiting time problems associated with runs in Markov dependent trials, *Annals of the Institute of Statistical Mathematics, Tokyo*, **51**, 323–330. [10.7.3]

Antzoulakos, D. L., and Chadjiconstantinidis, S. (2001). Distributions of numbers of success runs of fixed length in Markov dependent trials, *Annals of the Institute of Statistical Mathematics, Tokyo*, **53**, 599–619. [10.7.5]

Antzoulakos, D. L., and Philippou, A. N. (1997). Probability distribution functions of succession quotas in the case of Markov dependent trials, *Annals of the Institute of Statistical Mathematics, Tokyo*, **49**, 531–539. [10.7.2]

Appell, P., and Kampé de Fériet, J. (1926). *Fonctions Hypergéométriques et Hypersphériques*, Paris: Gauthier-Villars. [8.2.4]

Aragon, J., Eberly, D., and Eberly, S. (1992). Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution, *Statistics and Probability Letters*, **15**, 375–379. [5.8.5]

Arbous, A. G., and Kerrich, J. E. (1951). Accident statistics and the concept of accident proneness, *Biometrics*, **7**, 340–432. [4.12.5, 5.10]

Archibald, E. E. A. (1948). Plant populations I: A new application of Neyman's contagious distribution, *Annals of Botany, New Series*, **12**, 221–235. [9.6.5]

Arfwedson, G. (1951). A probability distribution connected with Stirling's second class numbers, *Skandinavisk Aktuarietidskrift*, **34**, 121–132. [10.4.1]

Arley, N. (1943). *On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation*, Copenhagen: Gads Forlag. [11.2.7]

Arley, N., and Buch, K. R. (1950). *Introduction to the Theory of Probability and Statistics*, New York: Wiley. [5.3]

Armitage, P. (1952). The statistical theory of bacterial populations subject to mutation, *Journal of the Royal Statistical Society, Series B*, **14**, 1–44. [11.2.11]

Arnold, B. C. (1980). Two characterizations of the geometric distribution, *Journal of Applied Probability*, **17**, 570–573. [5.9.1]

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics*, New York: Wiley. [1.2.12]

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998). *Records*, New York: Wiley. [10.5]

Arnold, B. C., Becker, A., Gather, U., and Zahedi, H. (1984). On the Markov property of order statistics, *Journal of Statistical Planning and Inference*, **9**, 147–154. [5.2]

Arnold, B. C., and Laguna, L. (1977). The generalized Pareto distributions, with applications to income data, *International Studies in Economics: Monograph No. 10*, Ames, IA: Iowa State University Department of Economics. [11.2.20]

Arnold, B. C., and Meeden, G. (1975). Characterization of distributions by sets of moments of order statistics, *Annals of Statistics*, **3**, 754–758. [1.2.12]

Aroian, L. A. (1937). The Type B Gram-Charlier series, *Annals of Mathematical Statistics*, **8**, 183. [11.2.8]

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method, *Annals of Probability*, **17**, 9–25. [4.2.2]

Arratia, R., Goldstein, L., and Gordon, L. (1990). Poisson approximation and the Chen-Stein method, *Statistical Science*, **5**, 403–434. [4.2.2]

Atkinson, A. C., and Yeh, L. (1982). Inference for Sichel's compound Poisson distribution, *Journal of the American Statistical Association*, **77**, 153–158. [11.1.5]

Austin, T., Fagan, R., Lehrer, T., and Penney, W. (1957). The distribution of the number of locally maximal elements in a random sample from a continuous distribution, *Annals of Mathematical Statistics*, **28**, 786–790. [10.5]

Bahadur, R. R. (1960). Some approximations to the binomial distribution function, *Annals of Mathematical Statistics*, **31**, 43–54. [3.6.2]

Bailey, N. T. J. (1951). On estimating the size of mobile populations from recapture data, *Biometrika*, **38**, 293–306. [6.9.2]

Bailey, N. T. J. (1954). A continuous time treatment of a simple queue, using generating functions, *Journal of the Royal Statistical Society, Series B*, **16**, 288–291. [11.2.14]

Bailey, N. T. J. (1964). *The Elements of Stochastic Processes with Applications to the Natural Sciences*, New York: Wiley. [3.12.2, 11.2.7]

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases and Its Applications* (second edition), London: Griffin. [3.12.5]

Bailey, W. N. (1935). *Generalized Hypergeometric Series*, London: Cambridge University Press. [1.1.6, 1.1.12, 2.9]

Bain, L. J., Engelhardt, M., and Williams, D. H. (1990). Confidence bounds for the binomial N parameter, *Communications in Statistics—Simulation and Computation*, **19**, 335–348. [3.8.3]

Balakrishnan, N. (1986). Order statistics from discrete distributions, *Communications in Statistics—Theory and Methods*, **15**, 657–675. [1.2.12, 3.5]

Balakrishnan, N. (1997). Joint distributions of numbers of success-runs and failures until the first consecutive k successes in a binary sequence, *Annals of the Institute of Statistical Mathematics, Tokyo*, **49**, 519–529. [10.7.2]

Balakrishnan, N., Balasubramanian, K., and Viveros, R. (1995). Start-up demonstration tests under correlation and corrective action, *Naval Research Logistics*, **42**, 1271–1276. [10.7.2]

Balakrishnan, N., and Cohen, A. C. (1991). *Order Statistics and Inference: Estimation Methods*, San Diego, CA: Academic Press. [1.2.12]

Balakrishnan, N., Johnson, N. L., and Kotz, S. (1998). A note on relationships between moments, central moments and cumulants from multivariate distributions, *Statistics and Probability Letters*, **39**, 49–54. [1.2.8]

Balakrishnan, N., and Koutras, M. V. (2002). *Runs and Scans with Applications*, New York: Wiley. [2.8, 10.7.2–10.7.6]

Balakrishnan, N., and Nevzorov, V. B. (2003) *A Primer on Statistical Distributions*, New York: Wiley. [5.4]

Baldessari, B., and Weber, J. (1987). Intrinsic hypergeometric random variable, *Statistica*, **47**, 275–286. [6.11.4]

Barbour, A. D., and Hall, P. (1984). On the rate of Poisson convergence, *Mathematical Proceedings of the Cambridge Philosophical Society*, **95**, 473–480. [3.12.3]

Barbour, A. D., Holst, L., and Janson, S. (1992). *Poisson Approximation*, Oxford: Oxford University Press. [4.2.2, 10.7.2]

Bardwell, G. E. (1960). On certain characteristics of some discrete distributions, *Biometrika*, **47**, 473–475. [2.3.1]

Bardwell, G. E., and Crow, E. L. (1964). A two-parameter family of hyper-Poisson distributions, *Journal of the American Statistical Association*, **59**, 133–141. [2.3.1, 4.12.4]

Bar Lev, S. K., and Enis, P. (1986). Reproducibility and natural exponential families with power variance functions, *Annals of Statistics*, **14**, 1507–1522. [11.1.2]

Barlow, R. E., and Proschan, F. (1965). *Mathematical Theory of Reliability*, New York: Wiley. [3.4, 11.2.15]

Barlow, R. E., and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*, New York: Holt, Rinehart & Winston. [11.2.15]

Barnard, G. A. (1954). Sampling inspection and statistical decisions, *Journal of the Royal Statistical Society, Series B*, **16**, 151–174. [6.9.2]

Barndorff-Neilsen, O. E. (1965). Identifiability of mixtures of exponential families, *Journal of Mathematical Analysis and Applications*, **12**, 115–121. [8.2.2]

Barndorff-Neilsen, O. E. (1991). Likelihood theory, *Statistical Theory and Modelling: In Honour of Sir David Cox, FRS*, D. V. Hinkley, N. Reid, and E. J. Snell (editors), 232–264. London: Chapman & Hall. [1.2.16]

Barndorff-Neilsen, O. E., and Jørgensen, B. (1991). Some parametric models on the simplex, *Journal of Multivariate Analysis*, **39**, 106–116. [1.2.16, 11.1.9]

Barnett, V. D. (1999). *Comparative Statistical Inference* (third edition), New York: Wiley. [1.2.16]

Barry, D. (1990). Empirical Bayes estimation of binomial probabilities in one-way and two-way layouts, *The Statistician*, **39**, 437–453. [3.8.2]

Bartko, J. J. (1966). Approximating the negative binomial, *Technometrics*, **8**, 345–350. [5.6]

Bartlett, M. S. (1947). The use of transformations, *Biometrics*, **3**, 39–52. [3.6.3]

Bartlett, M. S. (1952). Contribution to discussion on "The statistical theory of bacterial populations subject to mutation," *Journal of the Royal Statistical Society, Series B*, **14**, 37–38. [11.2.11]

Bartlett, M. S. (1960). *Stochastic Population Models*, London: Methuen. [7.1.10]

Barton, D. E. (1957). The modality of Neyman's contagious distribution of type A, *Trabajos de Estadística*, **8**, 13–22. [9.6.1]

Barton, D. E. (1958). The matching distributions: Poisson limiting forms and derived methods of approximation, *Journal of the Royal Statistical Society, Series B*, **20**, 73–92. [10.3]

Barton, D. E. (1961). Unbiased estimation of a set of probabilities, *Biometrika*, **48**, 227–229. [4.7.2]

Barton, D. E. (1966). Review of P. J. Staff: The displaced Poisson distribution, *Mathematical Reviews*, **33**, 6583. [4.12.4, 8.3.3, 8.4]

Barton, D. E., and David, F. N. (1959a). Contagious occupancy, *Journal of the Royal Statistical Society, Series B*, **21**, 120–133. [10.4.1]

Barton, D. E., and David, F. N. (1959b). The dispersion of a number of species, *Journal of the Royal Statistical Society, Series B*, **21**, 190–194. [10.4.1]

Barton, D. E., David, F. N., and Merrington, M. (1963). Tables for the solution of the exponential equation $\exp(b) - b/(1-p) = 1$, *Biometrika*, **50**, 169–172. [7.1.6, 7.1.7]

Basu, D. (1955). A note on the structure of a stochastic model considered by V. M. Dandekar, *Sankhyā*, **15**, 251–252. [11.2.1]

Bateman, H. (1910). On the probability distribution of particles, *Philosophical Magazine, Series 6*, **20**, 704–707. [4.2.1]

Beall, G. (1940). The fit and significance of contagious distributions when applied to observations on larval insects, *Ecology*, **21**, 460–474. [9.6.1, 9.9]

Beall, G., and Rescia, R. R. (1953). A generalization of Neyman's contagious distributions, *Biometrics*, **9**, 354–386. [5.12.6, 9.9]

Bebbington, M. S., and Lai, C. D. (1998). A generalized negative binomial and applications, *Communications in Statistics—Theory and Methods*, **27**, 2515–2533. [5.12.9]

Behboodian, J. (1975). Structural properties and statistics of finite mixtures, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 103–112. Dordrecht: Reidel. [8.2.2]

Benkherouf, L. (1996). On the generalized Heine distribution, *Journal of Applied Statistical Science*, **3**, 267–272. [10.8.2]

Benkherouf, L., and Alzaid. A. A. (1993). On the generalized Euler distribution, *Statistics and Probability Letters*, **18**, 323–326. [10.8.2]

Benkherouf, L., and Bather, J. A. (1988). Oil exploration: Sequential decisions in the face of uncertainty, *Journal of Applied Probability*, **25**, 529–543. [2.9, 4.12.6, 7.1.12, 10.8.2]

Bennett, B. M., and Birch, B. (1964). Sampling inspection tables for comparison of two groups using the negative binomial model, *Trabajos de Estadística*, **15**, 1–12. [5.10]

Bennett, B. M., and Nakamura, E. (1970). Note on an approximation to the distribution of the range from Poisson samples, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 119–121. University Park: Pennsylvania State University Press. [4.5]

Bennett, W. S. (1965). *A New Binomial Approximation for Cumulative Hypergeometric Probabilities*, PhD Thesis, Washington, DC: American University. [6.5]

Berg, S. (1974). Factorial series distributions, with applications to capture-recapture problems, *Scandinavian Journal of Statistics*, **1**, 145–152. [2.7, 3.4]

Berg, S. (1975). A note on the connection between factorial series distributions and zero-truncated power series distributions, *Scandinavian Actuarial Journal*, **1975**, 233–237. [2.7]

Berg, S. (1978). Characterization of a class of discrete distributions by properties of their moment distributions, *Communications in Statistics—Theory and Methods*, **A7**, 785–789. [2.7]

Berg, S. (1983a). Factorial series distributions, *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz, N. L. Johnson, and C. B. Read (editors), 17–22. New York: Wiley. [2.7, 3.4]

Berg, S. (1983b). Random contact processes, snowball sampling and factorial series distributions, *Journal of Applied Probability*, **20**, 31–46. [2.7]

Berg, S. (1985). Generating discrete distributions from modified Charlier type B expansions, *Contributions to Probability and Statistics in Honour of Gunnar Blom*, J. Lanke and G. Lindgren (editors), 39–48. Lund: University of Lund. [11.2.8]

Berg, S. (1987). Associated LAH numbers, factorial series distributions and unbiased estimation of a size parameter, *Biometrical Journal*, **29**, 163–171. [2.7]

Berg, S. (1988a). Stirling distributions, *Encyclopedia of Statistical Sciences*, Vol. 8, S. Kotz, N. L. Johnson, and C. B. Read (editors), 773–776. New York: Wiley. [7.1.12]

Berg, S. (1988b). Urn models, *Encyclopedia of Statistical Sciences*, Vol. 9, S. Kotz, N. L. Johnson, and C. B. Read (editors), 424–436. New York: Wiley. [5.1, 11.2.19]

Berg, S., and Mutafchiev, L. (1990). Random mappings with an attracting centre: Lagrangian distributions and a regression function, *Journal of Applied Probability*, **27**, 622–636. [7.2.5]

Berg, S., and Nowicki, K. (1991). Statistical inference for a class of modified power series distributions with applications to random mapping theory, *Journal of Statistical Planning and Inference*, **28**, 247–261. [7.2.5]

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (second edition), New York: Springer-Verlag. [1.2.16]

Berger, R. L. (1991). A remark on algorithm AS 152: Cumulative hypergeometric probabilities, *Applied Statistics*, **40**, 374–375. [6.6]

Berljand, O. S., Nazarov, I. M., and Pressman, A. J. (1962). i^n erfc distribution—or composite Poisson distribution, *Soviet Mathematics*, **3**, 1744–1746. [9.4]

Bernardo, J. M. (1976). Algorithm AS103: Psi (digamma) function, *Applied Statistics*, **25**, 315–317. [11.2.2]

Bernado, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (editors) (1988). *Bayesian Statistics*, Vol. 3: *Proceedings of the Third Valencia International Meeting*. Amsterdam: North-Holland. [1.2.16]

Bernado, J. M., et al. (editors) (1992, 1996, 1999). *Bayesian Statistics*, Vols. 4–6. Oxford: Oxford University Press. [1.2.2, 1.2.16]

Bernoulli, J. (1713). *Ars Conjectandi*, Basilea: Thurnisius. [11.2.19]

Bertin, E., and Theodorescu, R. (1995). Preserving unimodality by mixing, *Statistics and Probability Letters*, **25**, 281–288. [8.3.2]

Bertrand, J. L. F. (1889). *Calcul des Probabilités*, Paris: Gauthier-Villars. [3.3]

Best, D. J. (1974). The variance of the inverse binomial estimator, *Biometrika*, **61**, 385–386. [5.8.2]

Best, D. J., and Gipps, P. G. (1974). An improved gamma approximation to the negative binomial, *Technometrics*, **16**, 621–624. [5.6]

Best, D. J., and Rayner, J. C. W. (1989). Goodness of fit for the geometric distribution, *Biometrical Journal*, **31**, 307–311. [5.8.5]

Best, D. J., and Rayner, J. W. C. (1999). Goodness of fit for the Poisson distribution, *Statistics and Probability Letters*, **44**, 259–265. [4.7.4]

Best, D. J., and Rayner, J. C. W. (2003). Tests of fit for the geometric distribution, *Communications in Statistics—Simulation and Computation*, **32**, 1065–1078. [5.8.5]

Bhalerao, N. R., and Gurland, J. (1977). *A Unified Approach to Estimating Parameters in Some Generalized Poisson Distributions*, Technical Report No. 494, Madison, WI: University of Wisconsin. [9.9]

Bhalerao, N. R., Gurland, J., and Tripathi, R. C. (1980). A method of increasing power of a test for the negative binomial and Neyman type A distributions, *Journal of the American Statistical Association*, **75**, 934–938. [9.6.4]

Bhat, U. N. (1978). Theory of queues, *Handbook of Operations Research*, Vol. 1, J. J. Moder and S. E. Elmaghraby (editors), 352–397. New York: Van Nostrand-Reinhold. [11.2.14]

Bhat, U. N. (2002). *Elements of Applied Stochastic Processes* (third edition), New York: Wiley. [5.2, 5.3, 5.10, 11.2.14]

Bhattacharya, S. K. (1966). Confluent hypergeometric distributions of discrete and continuous type with applications to accident proneness, *Bulletin of the Calcutta Statistical Association*, **15**, 20–31. [4.12.4, 5.3, 8.3.3, 8.4]

Bhattacharya, S. K., and Holla, M. S. (1965). On a discrete distribution with special reference to the theory of accident proneness, *Journal of the American Statistical Association*, **60**, 1060–1066. [8.3.3]

Biedenhahn, L. C. (1989). The quantum group $SU_q(2)$ and a q analogue of the boson operators, *Journal of Physics A: Mathematics and General*, **22**, L873–878. [10.8.2]

Bienaymé, I. J. (1874). Sur une question de probabilités, *Bulletin de la Socièté Mathématique de France*, **2**, 153–154. [10.6.2]

Binet, F. (1953). The fitting of the positive binomial distribution when both parameters are estimated from the sample, *Annals of Eugenics, London*, **18**, 117–119. [3.8.2]

Bingham, N. H. (1991). Fluctuation theory for the Ehrenfest urn, *Advances in Applied Probability*, **23**, 598–611. [6.11.2]

Binns, M. R. (1975). Sequential estimation of the mean of a negative binomial distribution, *Biometrika*, **62**, 433–440. [5.8.2]

Binns, M. R. (1986). Behavioral dynamics and the negative binomial distribution, *Oikos*, **47**, 315–318. [5.10]

Birch, M. W. (1963). An algorithm for the logarithmic series distribution, *Biometrics*, **19**, 651–652. [7.1.7]

Birnbaum, A. (1964). Median-unbiased estimators, *Bulletin of Mathematical Statistics*, **11**, 25–34. [3.8.2]

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MA: MIT Press. [4.9]

Bissell, A. F. (1972a). A negative binomial model with varying element sizes, *Biometrika*, **59**, 435–441. [5.12.4]

Bissell, A. F. (1972b). Another negative binomial model with varying element sizes, *Biometrika*, **59**, 691–693. [5.12.4]

Bissinger, B. H. (1965). A type-resisting distribution generated from considerations of an inventory decision problem, *Classical and Contagious Discrete Distributions*,

G. P. Patil (editor), 15–17. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [6.2.5, 11.2.13]

Bizley, M. T. L. (1951). Some notes on probability, *Journal of the Institute of Actuaries Students' Society*, **10**, 161–203. [3.6.1]

Blackman, G. E. (1935). A study by statistical methods of the distribution of species in a grassland association, *Annals of Botany (New Series)*, **49**, 749–778. [7.1.9]

Blischke, W. R. (1962). Moment estimation for the parameters of a mixture of two binomial distributions, *Annals of Mathematical Statistics*, **33**, 444–454. [8.2.6]

Blischke, W. R. (1964). Estimating the parameters of mixtures of binomial distributions, *Journal of the American Statistical Association*, **59**, 510–528. [8.2.5, 8.2.6]

Blischke, W. R. (1965). Mixtures of discrete distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 351–372. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [8.2.2, 8.2.6, 8.3.1]

Bliss, C. I. (1965). An analysis of some insect trap records, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 385–397. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [7.1.12]

Bliss, C. I., and Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial, *Biometrics*, **9**, 176–200. [5.8.3, 5.8.4, 5.10]

Bliss, C. I., and Owen, A. R. G. (1958). Negative binomial distributions with a common k, *Biometrika*, **45**, 37–58. [5.8.4]

Blom, G., and Holst, L. (1986). Random walks of ordered elements with applications, *American Statistician*, **40**, 271–274. [10.5]

Blom, G., Thorburn, D., and Vessey, T. A. (1990). The distribution of the record position and its applications, *American Statistician*, **44**, 151–153. [10.5]

Blomqvist, N. (1952). On an exhaustion process, *Skandinavisk Aktuarietidskrift*, **35**, 201–210. [10.8.1]

Blum, J. R., and Susarla, V. (1977). Estimation of a mixing distribution function, *Annals of Probability*, **5**, 200–209. [8.3.1]

Blumenthal, S., and Dahiya, R. C. (1981). Estimating the binomial parameter n, *Journal of the American Statistical Association*, **76**, 903–909. [3.8.2]

Blumenthal, S., Dahiya, R. C., and Gross, A. J. (1978). Estimating the complete sample size from an incomplete Poisson sample, *Journal of the American Statistical Association*, **73**, 182–187. [4.10.1]

Blyth, C. R. (1980). Expected absolute error of the usual estimator of the binomial parameter, *American Statistician*, **34**, 155–157. [3.8.2]

Blyth, C. R. (1986). Approximate binomial confidence limits, *Journal of the American Statistical Association*, **81**, 843–855. [3.8.3]

Blyth, C. R., and Hutchinson, D. W. (1960). Tables of Neyman-shortest confidence intervals for the binomial parameter, *Biometrika*, **47**, 381–391. [3.8.3]

Blyth, C. R., and Hutchinson, D. W. (1961). Tables of Neyman-shortest-unbiased confidence intervals for the Poisson parameter, *Biometrika*, **48**, 191–194. [4.7.3]

Blyth, C. R., and Still, H. A. (1983). Binomial confidence intervals, *Journal of the American Statistical Association*, **78**, 108–116. [3.8.3]

Bohman, H. (1963). Two inequalities for Poisson distributions, *Skandinavisk Aktuarietidskrift*, **46**, 47–52. [4.5]

Böhning, D. (1983a). Refined estimation of the logarithmic series distribution, *EDV in Medizin und Biologie*, **14**, 58–61. [7.1.7]

Böhning, D. (1983b). Maximum likelihood estimation of the logarithmic series distribution, *Statistische Hefte*, **24**, 121–140. [7.1.7]

Böhning, D. (1994). A note on a test for Poisson overdispersion, *Biometrika*, **81**, 418–419. [8.2.5]

Böhning, D. (1998). Zero-inflated Poisson models and C.A.MAN: A tutorial of evidence, *Biometrical Journal*, **40**, 833–843. [8.2.4, 11.1.1]

Böhning, D. (1999). *Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others*, New York: Chapman & Hall/CRC. [8.1.2]

Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L., and Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology, *Journal of the Royal Statistical Society, Series A*, **162**, 195–209. [8.1.2, 8.2.4]

Bollinger, R. C., and Salvia, A. A. (1982). Consecutive-k-out-of-n: F networks, *IEEE Transactions on Reliability*, **R-31**, 53–55. [10.7.5]

Bol'shev, L. N. (1964). Distributions related to the hypergeometric, *Teoriya Veroyatnostei i ee Primeneniya*, **9**, 687–692. [6.2.2, 6.5]

Bol'shev, L. N. (1965). On a characterization of the Poisson distribution, *Teoriya Veroyatnostei i ee Primeneniya*, **10**, 446–456. [4.7.3, 4.8]

Bondesson, L. (1988). On the gain by spreading seeds: A statistical analysis of sowing experiments, *Scandinavian Journal of Forest Research*, 305–314. [8.2.6]

Boole, G. (1854). *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probability* (reprinted 1951), New York: Dover. [10.2]

Borah, M. (1984). The Gegenbauer distribution revisited: Some recurrence relations for moments, cumulants, etc, estimation of parameters and its goodness of fit, *Journal of the Indian Society for Agricultural Statistics*, **36**, 72–78. [11.2.7]

Borel, E. (1942). Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'un infinité de coefficients. Application au problème de l'attente à un guichet, *Comptes Rendus, Académie des Sciences, Paris, Series A*, **214**, 452–456. [7.2.2, 9.11, 11.2.14]

Borenius, G. (1953). On the statistical distribution of mine explosions, *Skandinavisk Aktuarietidskrift*, **36**, 151–157. [10.8.1]

Borges, R. (1970). Eine Approximation der Binomialverteilung durch die Normalverteilung der Ordnung $1/n$, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebeite*, **14**, 189–199. [3.6.1]

Bortkiewicz, L. von (1898). *Das Gesetz der Kleinen Zahlen*, Leipzig: Teubner. [4.2.1]

Bortkiewicz, L. von (1915). Über die Zeitfolge zufälliger Ereignisse, *Bulletin de l'Institut International de Statistique*, **20**, 30–111. [4.2.1]

Bortkiewicz, L. von (1917). *Die Iterationen*, Berlin: Springer. [10.6.1]

Bosch, A. J. (1963). The Pólya distribution, *Statistica Neerlandica*, **17**, 201–213. [5.3, 6.2.4]

Boswell, M. T., Gore, S. D., Patil, G. P., and Taillie, C. (1993). The art of computer generation of random variables, *Handbook of Statistics*, Vol. 9, C. R. Rao (editor). Amsterdam: North Holland. [1.2.17]

Boswell, M. T., Ord, J. K., and Patil, G. P. (1979). Chance mechanisms underlying univariate distributions, *Statistical Ecology*, Vol. 4: *Statistical Distributions in Ecological Work*, J. K. Ord, G. P. Patil, and C. Taillie (editors), 1–156. Fairland, MD: International Co-operative Publishing House. [3.10, 9.2]

Boswell, M. T., and Patil, G. P. (1970). Chance mechanisms generating the negative binomial distributions, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 3–22. University Park: Pennsylvania State University Press. [5.3, 5.10, 5.11]

Boswell, M. T., and Patil, G. P. (1971). Chance mechanisms generating the logarithmic series distribution used in the analysis of number of species and individuals, *Statistical Ecology*, Vol. 1: *Spatial Patterns and Statistical Distributions*, G. P. Patil, E. C. Pielou, and W. E. Waters (editors), 99–130. University Park: Pennsylvania State University Press. [7.1.1, 7.1.2]

Boswell, M. T., and Patil, G. P. (1973). Characterization of certain discrete distributions by differential equations with respect to their parameters, *Australian Journal of Statistics*, **15**, 128–131. [4.8]

Bowerman, P. N., and Scheuer, E. M. (1990). Calculation of the binomial survivor function, *IEEE Transactions on Reliability*, **39**, 162–166. [3.7.1]

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., and Nesbitt, C. J. (1986). *Actuarial Mathematics*, Chicago: Society of Actuaries. [9.13]

Bowman, D., and George, E. O. (1995). A saturated model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies, *Journal of the American Statistical Association*, **90**, 871–879. [3.12.6]

Bowman, K. O. (1984). Extended moment series and the parameters of the negative binomial distribution, *Biometrics*, **40**, 249–252. [5.8.3]

Bowman, K. O., Kastenbaum, M. A., and Shenton, L. R. (1992). The negative hypergeometric distribution and estimation by moments, *Communications in Statistics—Simulation and Computation*, **21**, 301–332. [6.7.2, 8.3.5]

Bowman, K. O., and Shenton, L. R. (1965). *Asymptotic Covariance for the Maximum Likelihood Estimators of the Parameters of a Negative Binomial Distribution*, Union Carbide Corporation Report K-1643, Oak Ridge, TN: Union Carbide. [5.8.3]

Bowman, K. O., and Shenton, L. R. (1966). *Biases of Estimators for the Negative Binomial Distribution*, Union Carbide Corporation Report ORNL-4005, Oak Ridge, TN: Union Carbide. [5.8.3]

Bowman, K. O., and Shenton, L. R. (1967). *Remarks on Estimation Problems for the Parameters of the Neyman Type A Distribution*, Oak Ridge National Laboratory Report ORNL-4102, Oak Ridge, TN: ORNL. [9.6.3, 9.6.4]

Bowman, K. O., and Shenton, L. R. (1970). Properties of the maximum likelihood estimator for the parameter of the logarithmic series distribution, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 127–150. University Park: Pennsylvania State University Press. [7.1.7]

Bowman, K. O., Shenton, L. R., and Kastenbaum, M. A. (1991). *Discrete Pearson Distributions*, Oak Ridge National Laboratory Technical Report TM-11899, Oak Ridge, TN: ORNL. [2.3.3, 10.8.3]

Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, **26**, 211–252. [5.10]

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistics*, Reading, MA: Addison-Wesley. [1.2.2]

Boyer, C. B. (1950). Cardan and the Pascal triangle, *American Mathematical Monthly*, **57**, 387–390. [3.2]

Brainerd, B. (1972). On the relation between types and tokens in literary text, *Journal of Applied Probability*, **9**, 507–518. [3.12.3]

Branson, D. (2000). Stirling numbers and Bell numbers: Their role in combinatorics and probability theory, *Mathematical Scientist*, **25**, 1–31. [7.1.12]

Brass, W. (1958). Simplified methods of fitting the truncated negative binomial distribution, *Biometrika*, **45**, 59–68. [5.11]

Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A Guide to Simulation* (second edition), New York: Springer-Verlag. [1.2.17]

Brenner, D. J., and Quan, H. (1990). Exact confidence limits for binomial proportions—Pearson and Hartley revisited, *The Statistician*, **39**, 391–397. [3.8.3]

Broderick, T. S. (1937). On some symbolic formulae in probability theory, *Proceedings of the Royal Irish Academy, Series A*, **44**, 19–28. [10.2]

Brooks, R. J., James, W. H., and Gray, E. (1991). Modelling sub-binomial variation in the frequency of sex combinations in litters of pigs, *Biometrics*, **47**, 403–417. [3.12.6]

Brooks, S. P., Morgan, B. J. T., Ridout, M. S., and Pack, S. E. (1997). Finite mixture models for proportions, *Biometrics*, **53**, 1097–1115. [6.9.2, 11.1.1]

Brown, B. (1965). *Some Tables of the Negative Binomial Distribution and Their Use*, RAND Corporation Memorandum RM-4577-PR, Santa Monica, CA: RAND. [5.7]

Brown, M. B., and Bromberg, J. (1984). An efficient two-stage procedure for generating variates from the multinomial distribution, *American Statistician*, **38**, 216–219. [4.8]

Brown, T. C., and Pallant, D. L. (1992). A note on antithetic variates from discrete distributions, *Journal of Statistical Computation and Simulation*, **40**, 193–199. [4.6.2]

Brown, T. C., and Xia, A. (1995). On Stein-Chen factors for Poisson approximation, *Statistics and Probability Letters*, **23**, 327–332. [4.2.2]

Brunk, H. D., Holstein, J. E., and Williams, F. (1968). A comparison of binomial approximations to the hypergeometric distribution, *American Statistician*, **22**, 24–26. [6.5]

Bruns, H. (1906). *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Leipzig and Berlin: Teubner. [10.6.1]

Bryc, W. (1987). A characterization of the Poisson process by conditional moments, *Stochastics*, **20**, 17–26. [4.8]

Burr, I. W. (1973). Some approximate relations between terms of the hypergeometric, binomial and Poisson distributions, *Communications in Statistics*, **1**, 297–301. [6.5]

Burrell, Q. L., and Cane, V. R. (1982). The analysis of library data, *Journal of the Royal Statistical Society, Series A*, **145**, 439–471. [5.10]

Bustoz, J., Feldstein, A., Goodman, R., and Linnainmaa, S. (1979). Improved trailing digits estimates applied to optimal computer arithmetic, *Journal of the ACM*, **26**, 716–730. [7.1.1]

Buxton, P. A. (1940). Studies on populations of head lice, III: Material from South India, *Parasitology*, **32**, 296–302. [7.1.9]

Byrne, J., and Kabaila, P. (2001). Short exact confidence intervals for the Poisson mean, *Communications in Statistics—Theory and Methods*, **30**, 257–261. [4.7.3]

Cacoullos, T. (1975). Multiparameter Stirling and C-type distributions, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 19–30. Dordrecht: Reidel. [4.12.3, 7.1.12]

Cacoullos, T., and Charalambides, C. A. (1975). On MVUE for truncated binomial and negative binomial distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **27**, 235–244. [5.11]

Cairns, J., Overbaugh, J., and Miller, S. (1988). The origin of mutants, *Nature*, **335**, 142–145. [11.2.11]

Cameron, A. C., and Johansson, P. (1997). Count data regressions using series expansions with applications, *Journal of Applied Econometrics*, **12**, 203–224. [11.1.1, 11.1.6]

Cameron, A. C., and Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators, *Journal of Applied Econometrics*, **1**, 29–53. [11.1.3]

Cameron, A. C., and Trivedi, P. K. (1996). Count data models for financial data, *Handbook of Statistics*, Vol. 14, G. S. Maddala and C. R. Rao (editors), 363–391. Amsterdam: North-Holland. [11.1.1]

Cameron, A. C., and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, Cambridge: Cambridge University Press. [5.11, 11.1.1, 11.1.6]

Campbell, M. J. (1994). Time series regression for counts: An investigation into the relationship between Sudden Infant Death Syndrome and environmental temperature, *Journal of the Royal Statistical Society, Series A*, **157**, 191–208. [11.1.1]

Cane, V. (1967). Mathematical models for neural networks, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probablility*, **4**, 21–26. [3.12.2]

Caraco, T. (1979). Ecological response of animal group size frequencies, *Statistical Ecology*, Vol. 4: *Statistical Distributions in Ecological Work*, J. K. Ord, G. P. Patil, and C. Taillie (editors), 371–386. Fairland, MD: International Co-operative Publishing House. [7.1.2]

Carroll, R. J., and Lombard, F. (1985). A note on *N* estimators for the binomial distribution, *Journal of the American Statistical Association*, **80**, 423–426. [3.8.2]

Carver, H. C. (1919). On the graduation of frequency distributions, *Proceedings of the Casualty Actuarial Society of America*, **6**, 52–72. [2.3.1]

Carver, H. C. (1923). Frequency curves, *Handbook of Mathematical Statistics*, H. L. Rietz (editor), 92–119. Cambridge MA: Riverside. [2.3.1, 2.3.3]

Casella, G. (1986). Refining binomial confidence intervals, *Canadian Journal of Statistics*, **14**, 113–129. [3.8.2]

Casella, G., and Robert, C. (1989). Refining Poisson confidence intervals, *Canadian Journal of Statistics*, **17**, 45–57. [4.7.3]

Casella, G., and Strawderman, W. E. (1994). On estimating several binomial n's, *Sankhyā: The Indian Journal of Statistics, Series B*, **56**, 115–120. [3.8.2]

Cassie, R. M. (1962). Frequency distribution models in the ecology of plankton and other organisms, *Journal of Animal Ecology*, **31**, 65–92. [7.1.12]

Castoldi, L. (1963). Poisson processes with events in clusters, *Rendiconti del Seminaro della Facoltà di Scienze della Università di Cagliari*, **33**, 433–437. [9.3]

Catcheside, D. G. (1948). Genetic effects of radiation, *Advances in Genetics*, Vol. 2, M. Demerec (editor), 271–358. New York: Academic Press. [9.2]

Čekanavičius, V. (1998). Poisson approximations for sequences of random variables, *Statistics and Probability Letters*, **39**, 101–107. [4.2.2]

Čekanavičius, V. (2002). On the convergence of Markov binomial to Poisson distribution, *Statistics and Probability Letters*, **58**, 83–91. [4.2.2]

Cernuschi, F., and Castagnetto, L. (1946). Chains of rare events, *Annals of Mathematical Statistics*, **17**, 53–61. [9.6.1]

Chaddha, R. L. (1965). A case of contagion in binomial distribution, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 273–290. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [11.2.19]

Chahine, J. (1965). Une généralization de la loi binomiale négative, *Revue Statistique Appliquée*, **13**, 33–43. [6.3]

Champernowne, D. G. (1956). An elementary method of solution of the queueing problem with a single server and constant parameters, *Journal of the Royal Statistical Society, Series B*, **18**, 125–128. [11.2.14]

Chan, B. (1982). Derivation of moment formulas by operator valued probability generating functions, *American Statistician*, **36**, 179–181. [1.2.7]

Chandler, K. N. (1952). The distribution and frequency of record values, *Journal of the Royal Statistical Society, Series B*, **14**, 220–228. [10.5]

Chandra, S. (1977). On the mixtures of probability distributions, *Scandinavian Journal of Statistics*, **4**, 105–112. [8.2.2]

Chang, D. K. (1989). On infinitely divisible discrete distributions, *Utilitas Mathematica*, **36**, 215–217. [9.3]

Chao, M. T., and Strawderman, W. E. (1972). Negative moments of positive random variables, *Journal of the American Statistical Association*, **67**, 429–431. [3.3]

Chapman, D. G. (1948). A mathematical study of confidence limits of salmon populations calculated from sample tag ratios, *International Pacific Salmon Fisheries Commission Bulletin*, **2**, 69–85. [6.7.1]

Chapman, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological sample censuses, *University of California Publications in Statistics*, **1**, 131–159. [6.7.1]

Chapman, D. G. (1952a). Inverse, multiple and sequential sample censuses, *Biometrics*, **8**, 286–306. [6.9.1, 6.9.2]

Chapman, D. G. (1952b). On tests and estimates for the ratio of Poisson means, *Annals of the Institute of Statistical Mathematics, Tokyo*, **4**, 45–49. [4.7.2, 4.7.3]

Chapman, D. W. (1934). The statistics of the method of correct matchings, *American Journal of Psychology*, **46**, 287–298. [10.3]

Charalambides, C. A. (1977a). On the generalized discrete distributions and the Bell polynomials, *Sankhyā, Series B*, **39**, 36–44. [9.1]

Charalambides, C. A. (1977b). A new kind of numbers appearing in the n-fold convolution of truncated binomial and negative binomial distributions, *SIAM Journal of Applied Mathematics*, **33**, 279–288. [5.11]

Charalambides, C. A. (1986a). Gould series distributions with applications to fluctuations of sums of random variables, *Journal of Statistical Planning and Inference*, **14**, 15–28. [2.6, 5.12.3, 7.2.7, 9.1, 11.2.10]

Charalambides, C. A. (1986b). On discrete distributions of order k, *Annals of the Institute of Statistical Mathematics, Tokyo*, **38**, 557–568. [10.7.1, 10.7.4]

Charalambides, C. A. (1986c). Derivation of probabilities and moments of certain generalized discrete distributions via urn models, *Communications in Statistics—Theory and Methods*, **15**, 677–696. [9.1]

Charalambides, C. A. (1990). Abel series distributions with applications to fluctuations of sample functions of stochastic processes, *Communications in Statistics—Theory and Methods*, **19**, 317–335. [2.6, 7.2.5]

Charalambides, C. A. (1991). On a generalized Eulerian distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **43**, 197–206. [2.6]

Charalambides, Ch. A. (2002). *Enumerative Combinatorics*, Boca Raton, FL: Chapman & Hall. [1.1.3, 2.6, 7.1.12]

Charalambides, C. A., and Singh, J. (1988). A review of the Stirling numbers, their generalizations and statistical applications, *Communications in Statistics—Theory and Methods*, **17**, 2533–2595. [1.1.3, 7.1.12]

Charlier, C. V. L. (1905a). Die zweite Form des Fehlergesetzes, *Arkiv für Matematik, Astronomi och Fysik No. 15*, **2**, 1–8. [4.2.1]

Charlier, C. V. L. (1905b). Über die Darstellung willkürlicher Funktionen, *Arkiv für Matematik, Astronomi och Fysik No. 20*, **2**, 1–35. [4.2.1]

Charlier, C. V. L. (1920). *Vorlesungen über die Grundzüge der Mathematischen Statistik*, Hamburg. [3.12.3]

Chatfield, C. (1969). On estimating the parameters of the logarithmic series and negative binomial distributions, *Biometrika*, **56**, 411–414. [7.1.7]

Chatfield, C. (1970). Discrete distributions in market research, *Random Counts in Scientific Work*, Vol. 3: *Random Counts in Physical Science, Geo Science, and Business*, G. P. Patil (editor), 163–181. University Park: Pennsylvania State University Press. [7.1.9, 10.4.3]

Chatfield, C. (1975). A marketing application of a characterization theorem, *Statistical Distributions in Scientific Work*, Vol. 2: *Model Building and Model Selection*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 175–185. Dordrecht: Reidel. [5.10]

Chatfield, C. (1983). *Statistics for Technology* (third edition), London: Chapman & Hall. [4.9]

Chatfield, C. (1986). Discrete distributions and purchasing models, *Communications in Statistics—Theory and Methods*, **15**, 697–708. [7.1.9]

Chatfield, C., Ehrenberg, A. S. C., and Goodhardt, G. J. (1966). Progress on a simplified model of stationary purchasing behaviour (with discussion), *Journal of the Royal Statistical Society, Series A*, **129**, 317–367. [5.10, 7.1.9, 7.1.10]

Chatfield, C., and Goodhardt, G. J. (1970). The beta-binomial model for consumer purchasing behaviour, *Applied Statistics*, **19**, 240–250. [6.7.2, 6.9.2, 10.4.3]

Chatfield, C., and Goodhardt, G. J. (1973). A consumer purchasing model with Erlang inter-purchase times, *Journal of the American Statistical Association*, **68**, 828–835. [5.12.8]

Chatterji, S. D. (1963). Some elementary characterizations of the Poisson distribution, *American Mathematical Monthly*, **70**, 958–964. [4.8]

Chaundy, T. W. (1962). F. H. Jackson, (obituary), *Journal of the London Mathematical Society*, **37**, 126–128. [1.1.12]

Chen, H.-C., and Asau, Y. (1974). On generating random variates from an empirical distribution, *AIIE Transactions*, **6**, 163–166. [1.2.17]

Chen, L. H. Y. (1975). Poisson approximation for dependent trials, *Annals of Probability*, **3**, 534–545. [4.2.2, 10.7.2]

Chen, R. (1978). A surveillance system for congenital malformations, *Journal of the American Statistical Association*, **73**, 323–327. [5.2]

Cheng, T.-T. (1949). The normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan, *Bulletin of the American Mathematical Society*, **55**, 396–401. [4.4]

Chernoff, H., and Lander, E. (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial, *Journal of Statistical Planning and Inference*, **43**, 19–40. [3.8.4]

Chew, V. (1964). Applications of the negative binomial distribution with probability of misclassification, *Virginia Journal of Science (New Series)*, **15**, 34–40. [5.10]

Chew, V. (1971). Point estimation of the parameter of the binomial distribution, *American Statistician*, **25**, 47–50. [3.8.2]

Chiang, D., and Niu, S. C. (1981). Reliability of consecutive-k-out-of-n:F systems, *IEEE Transactions on Reliability*, **R-30**, 87–89. [10.7.5]

Chong, K.-M. (1977). On characterizations of the exponential and geometric distributions by expectations, *Journal of the American Statistical Association*, **72**, 160–161. [5.9.1]

Christoph, G., and Schreiber, K. (1998). Discrete random variables, *Statistics and Probability Letters*, **37**, 243–247. [5.12.2, 11.2.5]

Christoph, G., and Schreiber, K. (2000). Scaled Sibuya distribution and discrete self-decomposability, *Statistics and Probability Letters*, **48**, 181–187. [5.12.2]

Chryssaphinou, O., and Papastavridis, S. (1990). Limit distribution for a consecutive-k-out-of-n:F system, *Advances in Applied Probability*, **22**, 491–493. [10.7.2]

Chukwu, W. I. E., and Gupta, D. (1989). On mixing generalized Poisson with generalized gamma distribution, *Metron*, **47**, 314–318. [8.3.3]

Chung, J. H., and DeLury, D. B. (1950). *Confidence Limits for the Hypergeometric Distribution*, Toronto: University of Toronto Press. [6.6, 6.7.1, 6.9.1]

Chung, K. L., and Feller, W. (1949). Fluctuations in coin tossing, *Proceedings of the National Academy of Sciences USA*, **35**, 605–608. [6.10.2]

Clark, C. R., and Koopmans, L. H. (1959). *Graphs of the Hypergeometric O.C., and A.O.Q. Functions for Lot Sizes 10 to 225*, Monograph SCR-121, Sandia, TX: Sandia Corporation. [6.6]

Clark, R. E. (1953). Percentage points of the incomplete beta function, *Journal of the American Statistical Association*, **48**, 831–843. [3.8.3]

Clark, S. J., and Perry, J. N. (1989). Estimation of the negative binomial parameter k by maximum quasi-likelihood, *Biometrics*, **45**, 309–316. [5.8.3, 5.8.5, 5.10]

Clemans, K. G. (1959). Confidence limits in the case of the geometric distribution, *Biometrika*, **46**, 260–264. [5.2]

Clevenson, M. L., and Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws, *Journal of the American Statistical Association*, **70**, 698–705. [4.7.2]

Cliff, A. D., and Ord, J. K. (1981). *Spatial Processes: Models and Applications*, London: Pion. [4.9]

Clopper, C. J., and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, **26**, 404–413. [3.8.3]

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests, *Biometrics*, **10**, 417–451. [6.9.1]

Cohen, A. C. (1954). Estimation of the Poisson parameter from truncated samples and from censored samples, *Journal of the American Statistical Association*, **49**, 158–168. [4.10.2]

Cohen, A. C. (1959). Estimation in the Poisson distribution when sample values of $c + 1$ are sometimes erroneously reported as c, *Annals of the Institute of Statistical Mathematics, Tokyo*, **9**, 189–193. [4.10.3]

Cohen, A. C. (1960a). Estimating the parameter in a conditional Poisson distribution, *Biometrics*, **16**, 203–211. [4.10.1]

Cohen, A. C. (1960b). An extension of a truncated Poisson distribution, *Biometrics*, **16**, 446–450. [4.10.3]

Cohen, A. C. (1960c). Estimating the parameters of a modified Poisson distribution, *Journal of the American Statistical Association*, **55**, 139–143. [4.10.3]

Cohen, A. C. (1960d). Estimation in a truncated Poisson distribution when zeros and some ones are missing, *Journal of the American Statistical Association*, **55**, 342–348. [4.10.3]

Cohen, A. C. (1961). Estimating the Poisson parameter from samples that are truncated on the right, *Technometrics*, **3**, 433–438. [4.10.2]

Cohen, A. C. (1965). Estimation in mixtures of discrete distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 373–378. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [8.2.7]

Cohen, J. W. (1982). *The Single Server Queue*, Amsterdam: North Holland. [11.2.14]

Collett, D. (2003). *Modelling Binary Data*, Boca Raton, FL: Chapman & Hall. [6.9.2]

Congdon, P. (2003). *Applied Bayesian Modelling*, Chichester: Wiley. [1.2.2, 1.2.16]

Consael, R. (1952). Sur les processus de Poisson du type composé, *Académie Royale de Belgique Bulletin, Classe de Sciences 5^e Série*, **38**, 442–461. [8.3.2]

Consul, P. C. (1974). A simple urn model dependent upon predetermined strategy, *Sankhyā, Series B*, **36**, 391–399. [2.6, 5.12.3, 7.2.7, 11.2.19]

Consul, P. C. (1975). Some new characterizations of discrete Lagrangian distributions, *Statistical Distributions in Scientific Work*, Vol. 3: *Characterizations and Applications*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 279–290. Dordrecht: Reidel. [7.2.6]

Consul, P. C. (1981). Relation of modified power series distributions to Lagrangian probability distributions, *Communications in Statistics—Theory and Methods*, **A10**, 2039–2046. [2.2.2, 7.2.5]

Consul, P. C. (1983). Lagrange and related probability distributions, *Encyclopedia of Statistical Sciences*, Vol. 4, S. Kotz, N. L. Johnson, and C. B. Read (editors), 448–454. New York: Wiley. [7.2.3, 7.2.5]

Consul, P. C. (1984). On the distributions of order statistics for a random sample size, *Statistica Neerlandica*, **38**, 249–256. [7.2.6]

Consul, P. C. (1989). *Generalized Poisson Distributions*, New York: Dekker. [4.11, 5.12.3, 7.2, 7.2.6, 7.2.7, 8.2.4, 9.11]

Consul, P. C. (1990a). Geeta distribution and its properties, *Communications in Statistics—Theory and Methods*, **19**, 3051–3068. [7.2.4]

Consul, P. C. (1990b). New class of location-parameter discrete probability distributions and their characterizations, *Communications in Statistics—Theory and Methods*, **19**, 4653–4666. [2.2.2, 7.2.4]

Consul, P. C. (1990c). Two stochastic models for the Geeta distribution, *Communications in Statistics—Theory and Methods*, **19**, 3699–3706. [7.2.4]

Consul, P. C. (1990d). On some properties and applications of quasi-binomial distributions, *Communications in Statistics—Theory and Methods*, **19**, 477–504. [2.6]

Consul, P. C. (1991). Evolution of surnames, *International Statistical Review*, **59**, 271–278. [7.2.4]

Consul, P. C., and Famoye, F. (1989). Minimum variance unbiased estimation for the Lagrange power series distributions, *Statistics*, **20**, 407–415. [5.12.3, 7.2.5, 7.2.7]

Consul, P. C., and Famoye, F. (1992). Generalized Poisson regression model, *Communications in Statistics—Theory and Methods*, **21**, 89–109. [7.2.6, 11.1.1]

Consul, P. C., and Famoye, F. (1995). On the generalized negative binomial distribution, *Communications in Statistics—Theory and Methods*, **24**, 459–472. [7.2.7, 11.1.1]

Consul, P. C., and Famoye, F. (1996). Lagrangian Katz family of distributions, *Communications in Statistics—Theory and Methods*, **25**, 415–434. [7.2.4]

Consul, P. C., and Famoye, F. (2001). On Lagrangian distributions of the second kind, *Communications in Statistics—Theory and Methods*, **30**, 165–178. [2.5, 7.2, 7.2.9]

Consul, P. C., and Gupta, H. C. (1980). The generalized negative binomial distribution and its characterization by zero regression, *SIAM Journal of Applied Mathematics*, **39**, 231–237. [5.12.3, 7.2.7]

Consul, P. C., and Jain, G. C. (1971). On the log-gamma distribution and its properties, *Statistische Hefte*, **12**, 100–106. [10.4.3]

Consul, P. C., and Jain, G. C. (1973a). A generalization of the Poisson distribution, *Technometrics*, **15**, 791–799. [2.6, 7.2]

Consul, P. C., and Jain, G. C. (1973b). On some interesting properties of the generalized Poisson distribution, *Biometrische Zeitschrift*, **15**, 495–500. [2.6, 7.2.6]

Consul, P. C., and Mittal, S. P. (1975). A new urn model with predetermined strategy, *Biometrische Zeitschrift*, **17**, 67–75. [2.6, 7.2.5, 11.2.19]

Consul, P. C., and Shenton, L. R. (1972). Use of Lagrange expansion for generating generalized probability distributions, *SIAM Journal of Applied Mathematics*, **23**, 239–248. [2.6, 5.12.3, 7.2.5, 7.2.7]

Consul, P. C., and Shenton, L. R. (1973). Some interesting properties of Lagrangian distributions, *Communications in Statistics*, **2**, 263–272. [5.12.3, 7.2.5, 7.2.7]

Consul, P. C., and Shenton, L. R. (1975). On the probabilistic structure and properties of discrete Lagrangian distributions, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 41–48. Dordrecht: Reidel. [7.2.3, 7.2.5, 7.2.6]

Consul, P. C., and Shoukri, M. M. (1984). Maximum likelihood estimation for the generalized Poisson distribution, *Communications in Statistics—Theory and Methods*, **13**, 1533–1547. [7.2.6]

Consul, P. C., and Shoukri, M. M. (1985). The generalized Poisson distribution when the sample mean is larger than the sample variance, *Communications in Statistics—Theory and Methods*, **14**, 667–681. [7.2.6]

Consul, P. C., and Shoukri, M. M. (1986). The negative integer moments of the generalized Poisson distribution, *Communications in Statistics—Theory and Methods*, **15**, 1053–1064. [7.2.6]

Cook, G. W., Kerridge, D. F., and Pryce, J. D. (1974). Estimation of functions of a binomial parameter, *Sankhyā, Series A*, **36**, 443–448. [3.8.2]

Coolidge, J. L. (1921). The dispersion of observations, *Bulletin of the American Mathematical Society*, **27**, 439–442. [3.12.3]

Cooper, W. L. (1998). Negative binomial sums of random variables and discounted reward processes, *Journal of Applied Probability*, **35**, 589–599. [5.10]

Cormack, R. M. (1979). Models for capture-recapture, *Statistical Ecology*, Vol. 5, R. M. Cormack, G. P. Patil, and D. S. Robson (editors), 217–255. Burtonsville, MD: International Co-operative Publishing House. [6.9.1]

Cormack, R. M. (1989). Log-linear models for capture-recapture, *Biometrics*, **45**, 395–413. [4.9]

Cormack, R. M., and Jupp, P. E. (1991). Inference for Poisson and multinomial models for capture-recapture experiments, *Biometrika*, **78**, 911–916. [4.9]

Cournot, M. A. A. (1843). *Exposition de la Théorie des Chances et des Probabilités*, Paris: Librairie de L. Machette. [6.2.1]

Cox, D. R. (1975). Partial likelihood, *Biometrika*, **62**, 269–276. [1.2.16]

Cox, D. R. (1983). Some remarks on overdispersion, *Biometrika*, **70**, 269–274. [8.3.3]

Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics*, London: Chapman & Hall. [1.2.16]

Cox, D. R., and Isham, V. (1980). *Point Processes*, London: Chapman & Hall. [3.12.2, 8.3.3]

Cox, D. R., and Miller, H. D. (1965). *The Theory of Stochastic Processes*, London: Methuen. [5.12.5]

Cox, D. R., and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *Journal of the Royal Statistical Society, Series B*, **49**, 1–39. [1.2.16]

Cox, D. R., and Smith, W. L. (1961). *Queues*, London: Methuen. [11.2.14]

Cox, D. R., and Snell, E. J. (1981). *Applied Statistics: Principles and Examples*, London: Chapman & Hall. [3.10]

Cox, D. R., and Snell, J. E. (1989). *Analysis of Binary Data* (second edition), London: Chapman & Hall. [4.9, 6.9.1]

Craig, A. T. (1934). Note on the moments of a Bernoulli distribution, *Bulletin of the American Mathematical Society*, **40**, 262–264. [2.2.1]

Craig, C. C. (1953). On the utilization of marked specimens in estimating populations of flying insects, *Biometrika*, **40**, 170–176. [10.1, 10.4.1]

Crawford, G. B. (1966). Characterization of geometric and exponential distributions, *Annals of Mathematical Statistics*, **37**, 1790–1795. [5.9.1]

Cressie, N. (1978). A finely tuned continuity correction, *Annals of the Institute of Statistical Mathematics, Tokyo*, **30**, 436–442. [3.6.1]

Cressie, N., Davis, A. S., Folks, J. L., and Policello, G. E. (1981). The moment generating function and negative integer moments, *American Statistician*, **35**, 148–150. [3.3]

Cresswell, W. L., and Froggatt, P. (1963). *The Causation of Bus Driver Accidents*, London: Oxford University Press. [9.6.5, 9.9]

Crick, F. H. C., and Lawrence, P. A. (1975). Compartments and polychones in insect development, *Science*, **189**, 340–347. [5.10]

Crippa, D., and Simon, K. (1997). q-Distributions and Markov processes, *Discrete Mathematics*, **170**, 81–98. [10.8.1]

Crow, E. L. (1956). Confidence limits for a proportion, *Biometrika*, **43**, 423–435 (erratum **45**, 291). [3.8.3]

Crow, E. L. (1958). The mean deviation of the Poisson distribution, *Biometrika*, **45**, 556–559. [4.3]

Crow, E. L. (1975). Confidence limits for small probabilities from Bernoulli trials, *Communications in Statistics*, **4**, 397–414. [3.8.2, 3.8.3]

Crow, E. L., and Bardwell, G. E. (1965). Estimation of the parameters of the hyper-Poisson distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 127–140. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [2.3.1, 4.12.4]

Crow, E. L., and Gardner, R. S. (1959). Confidence intervals for the expectation of a Poisson variable, *Biometrika*, **46**, 444–453. [4.4, 4.7.3, 4.12.3]

Crow, E. L., and Shimizu, K. (editors) (1988). *Lognormal Distributions: Theory and Applications*, New York: Dekker. [11.1.4]

Crowder, M. J. (1978). Beta-binomial anova for proportions, *Applied Statistics*, **27**, 34–37. [6.9.2]

Daboni, L. (1959). A property of the Poisson distribution, *Bolletino della Unione Mathematica Italiana*, **14**, 318–320. [4.8]

Dacey, M. F. (1972). A family of discrete probability distributions defined by the generalized hypergeometric series, *Sankhyā, Series B*, **34**, 243–250. [2.4.1, 6.11.2]

Dacey, M. F. (1975). Probability laws for topological properties of drainage basins, *Statistical Distributions in Scientific Work*, Vol. 2: *Model Building and Model Selection*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 327–341. Dordrecht: Reidel. [2.4.1]

D'Agostino, R. B., Chase, W., and Belanger, A. (1988). The appropriateness of some common procedures for testing the equality of two independent binomial populations, *American Statistician*, **42**, 198–202. [3.8.4]

D'Agostino, R. B., and Stephens, M. A. (editors) (1986). *Goodness-of-Fit Techniques*, New York: Dekker. [1.2.16]

Dagpunar, J. (1988). *Principles of Random Variate Generation*, Oxford: Oxford University Press. [1.2.17]

Dahiya, R. C. (1981). An improved method of estimating an integer-parameter by maximum likelihood, *American Statistician*, **35**, 34–37. [3.8.2]

Dahiya, R. C. (1986). Integer-parameter estimation in discrete distributions, *Communications in Statistics—Theory and Methods*, **15**, 709–725. [3.8.2]

Dahiya, R. C., and Gross, A. J. (1973). Estimating the zero class from a truncated Poisson sample, *Journal of the American Statistical Association*, **68**, 731–733. [4.10.1]

Daley, D. J., and Gani, J. (1999). *Epidemic Modelling: An Introduction*, Cambridge: Cambridge University Press. [3.12.5, 10.4.3]

Dallas, A. C. (1974). A characterization of the geometric distribution, *Journal of Applied Probability*, **11**, 609–611. [5.9.1]

Dandekar, V. M. (1955). Certain modified forms of binomial and Poisson distributions, *Sankhyā*, **15**, 237–250. [11.2.1]

Daniels, H. E. (1961). Mixtures of geometric distributions, *Journal of the Royal Statistical Society, Series B*, **23**, 409–413. [5.2, 8.2.7]

Danish, M. B., and Hundley, J. T. (1979). *A Generalized Method for Computing the Probability of Killing X Targets Out of N Targets*, Memorandum Report ARBL-MR-02890, Aberdeen Proving Ground, MD: Ballistic Research Laboratory. [3.12.3]

Darroch, J. N. (1974). Multiplicative and additive interaction in contingency tables, *Biometrika*, **61**, 207–214. [3.12.6]

Darwin, J. H. (1960). An ecological distribution akin to Fisher's logarithmic distribution, *Biometrics*, **16**, 51–60. [6.11.2, 7.1.12]

Dasgupta, R. (1993). Cauchy equation on discrete domain and some characterizations, *Theory of Probability and Its Applications*, **38**, 318–328. [10.8.3]

David, F. N. (1950). Two combinatorial tests of whether a sample has come from a given population, *Biometrika*, **37**, 97–110. [10.4.1]

David, F. N. (1962). *Games, Gods, and Gambling*, London: Griffin. [11.2.19]

David, F. N., and Barton, D. E. (1962). *Combinatorial Chance*, London: Griffin. [1.1.3, 10.2, 10.3, 10.4.1, 10.5, 10.6.1]

David, F. N., and Johnson, N. L. (1952). The truncated Poisson distribution, *Biometrics*, **8**, 275–285. [4.10.1, 5.11]

David, F. N., Kendall, M. G., and Barton, D. E. (1966). *Symmetric Function and Allied Tables*, London: Cambridge University Press. [1.1.9, 10.6.1]

David, F. N., and Moore, P. G. (1954). Notes on contagious distributions in plant populations, *Annals of Botany, New Series*, **18**, 47–53. [9.6.1, 9.6.5]

David, H. A. (1981). *Order Statistics* (second edition), New York: Wiley. [1.2.12, 3.5]

Davies, O. L. (1933). On asymptotic formulae for the hypergeometric series I, *Biometrika*, **25**, 295–322. [6.2.5, 6.5]

Davies, O. L. (1934). On asymptotic formulae for the hypergeometric series II, *Biometrika*, **26**, 59–107. [6.2.5, 6.5]

Davis, H. T. (1933, 1935). *Tables of the Higher Mathematical Functions*, 2 volumes, Bloomington, IN: Principia. [1.1.2]

Davutyan, N. (1989). Bank failures as Poisson variates, *Economics Letters*, **29**, 333–338. [11.1.1]

Dean, C., Lawless, J. F., and Willmot, G. E. (1989). A mixed Poisson–inverse Gaussian regression model, *Canadian Journal of Statistics*, **17**, 171–182. [11.1.5]

De Castro, G., (1952). Note on differences of Bernoulli and Poisson variables, *Portugaliae Mathematica*, **11**, 173–175. [4.12.3]

Decker, R. D., and Fitzgibbon, D. J. (1991). *The Normal and Poisson Approximations to the Binomial: A Closer Look*, Department of Mathematics Technical Report No. 82.3, Hartford, CT: University of Hartford. [3.6.1]

Deely, J. J., and Lindley, D. V. (1981). Bayes empirical Bayes, *Journal of the American Statistical Association*, **76**, 833–841. [3.8.2]

De Finetti, B. (1931). Le funzioni caratteristiche di legge instantanea dotate di valori eccezionali, *Atti della Reale Academia Nazionale dei Lincei, Serie VI*, **14**, 259–265. [9.3]

DeGroot, M. H. (1970). *Optimal Statistical Decisions*, New York: McGraw-Hill. [1.2.16]

Delaporte, P. (1959). Quelques problémes de statistique mathématique posés par l'assurance automobile et le bonus pour non sinistre, *Bulletin Trimestriel de l'Institut des Actuaires Français*, **227**, 87–102. [5.12.5, 9.12]

Del Castillo, J., and Pérez-Casany, M. (1998). Weighted Poisson distributions for overdispersion and underdispersion situations, *Annals of the Institute of Statistical Mathematics, Tokyo*, **50**, 567–585. [11.1.1, 11.1.7]

De Moivre, A. (1711). De Mensura Sortis, *Philosophical Transactions of the Royal Society,*, **27**(329), 213–264. [4.2.1, 6.2.1]

De Moivre, A. (1718). *The Doctrine of Chances: Or, A Method of Calculating the Probability of Events in Play*, London: Pearson. [10.2]

De Moivre, A. (1725). *Annuities upon Lives: Or, The Valuation of Annuities upon Any Number of Lives: As also, of Reversions. To which is added, An Appendix Concerning the Expectations of Life, and Probabilities of Survivorship*, London: Fayram, Motte and Pearson. [10.2]

De Moivre, A. (1738). *The Doctrine of Chances: Or, A Method of Calculating the Probability of Events in Play* (second edition), London: Woodfall (reprinted by Cass, London, 1967). [10.7.1]

De Montmort, P. R. (1713). *Essai d'Analyse sur les Jeux de Hasard* (second edition), Paris: Quillau (reprinted by Chelsea, New York, 1980). [5.3, 10.2]

Denning, P. J., and Schwartz, S. C. (1972). Properties of the working set model, *Communications of the ACM*, **15**, 191–198. [10.4.1]

DeRiggi, D. F. (1983). Unimodality of likelihood functions for the binomial distribution, *Journal of the American Statistical Association*, **78**, 181–183. [3.8.2]

DeRouen, T. A., and Mitchell, T. J. (1974). A G_1 minimax estimator for a linear combination of binomial probabilities, *Journal of the American Statistical Association*, **69**, 231–233. [3.8.2]

Desloge, E. A. (1966). *Statistical Physics*, New York: Holt, Reinhart and Winston. [10.1]

Desmond, A. F., and Godambe, V. P. (1998). Estimating functions, *Encyclopedia of Biostatistics*, P. Armitage and T. Colton (editors), Vol. 2, 1375–1386. Chichester: Wiley. [1.2.16]

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, New York: Springer-Verlag. [1.2.17, 3.7.2, 4.5, 4.6.2, 5.2, 5.7, 6.6, 7.1.6, 8.1.2, 11.2.20]

Devroye, L. (1990). A note on Linnik's distribution, *Statistics and Probability Letters*, **9**, 305–306. [11.2.5]

Devroye, L. (1992). The branching process method in Lagrangian random variate generation, *Communications in Statistics-Simulation and Computation*, **21**, 1–14. [7.2.5]

Devroye, L. (1993). A triptych of discrete distributions related to the stable law, *Statistics and Probability Letters*, **18**, 349–351. [5.12.2, 11.2.5]

Dhanavanthan, P. (1998). Compound intervened Poisson distribution, *Biometrical Journal*, **40**, 641–646. [4.12.7]

Dhanavanthan, P. (2000). Estimation of the parameters of compound intervened Poisson distribution, *Biometrical Journal*, **42**, 315–320. [4.12.7]

Diaconis, P., and Zarbell, S. (1991). Closed form summation for classical distributions: Variations on a theme of De Moivre, *Statistical Science*, **6**, 284–302. [3.3, 3.4, 4.8]

Di Bucchianico, A. (1999). Combinatorics, computer algebra, and the Wilcoxon-Mann-Whitney test, *Journal of Statistical Planning and Inference*, **79**, 349–364. [10.8.4]

Dietz, K. (1966). On the model of Weiss for the spread of epidemics by carriers, *Journal of Applied Probability*, **3**, 375–382. [10.4.3]

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*, London: Academic. [5.10]

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models* (second edition), Boca Raton, FL: Chapman & Hall/CRC. [11.1.1]

Doksum, K. A. (1998). Estimation, *Encyclopedia of Biostatistics*, P. Armitage and T. Colton (editors), Vol. 2, 1386–1391. Chichester: Wiley. [1.2.16]

Domb, C. (1952). On the use of a random parameter in combinatorial problems, *Proceedings of the Physical Society, Series A*, **65**, 305–309. [10.4.1]

Doob, J. L. (1953). *Stochastic Processes*, New York: Wiley. [4.9]

Doray, L. G., and Luong, A. (1997). Efficient estimators for the Good family, *Communications in Statistics—Simulation and Computation*, **26**, 1075–1088. [11.2.20]

Doss, S. A. D. C. (1963). On the efficiency of best asymptotically normal estimates of the Poisson parameter based on singly and doubly truncated or censored samples, *Biometrics*, **19**, 588–594. [4.10.2]

Douglas, J. B. (1955). Fitting the Neyman type A (two parameter) contagious distribution, *Biometrics*, **11**, 149–173. [9.6.1, 9.6.3, 9.6.4]

Douglas, J. B. (1965). Asymptotic expansions for some contagious distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 291–302. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [9.6.3, 9.7]

Douglas, J. B. (1971). Stirling numbers in discrete distributions, *Statistical Ecology*, Vol. 1: *Spatial Patterns and Statistical Distributions*, G. P. Patil, E. C. Pielou, and W. E. Waters (editors), 69–98. University Park: Pennsylvania State University Press. [7.1.12]

Douglas, J. B. (1980). *Analysis with Standard Contagious Distributions*, Burtonsville, MD: International Co-operative Publishing House. [1.2.8, 2.2.1, 3.8.1, 3.10, 4.9, 4.11, 8.3.2, 9.0, 9.1, 9.5, 9.6.4, 9.7, 9.8, 9.10]

Douglas, J. B. (1986). Pólya-Aeppli distribution, *Encyclopedia of Statistical Sciences*, Vol. 7, S. Kotz, N. L. Johnson, and C. B. Read (editors), 56–59. New York: Wiley. [9.7]

Dowling, M. M., and Nakamura, M. (1997). Estimating parameters for discrete distributions via the empirical probability generating function, *Communications in Statistics—Simulation and Computation*, **26**, 301–313. [8.2.4]

Downton, F. (1967). Epidemics with carriers: A note on a paper by Dietz, *Journal of Applied Probability*, **4**, 264–270. [10.4.3]

Draper, N., and Guttman, I. (1971). Bayesian estimation of the binomial parameter, *Technometrics*, **13**, 667–673. [3.8.2]

Drezner, Z., and Farnum, N. (1993). A generalized binomial distribution, *Communications in Statistics—Theory and Methods*, **22**, 3051–3063. [3.12.6]

Drezner, Z., and Farnum, N. (1994). A correlated Poisson distribution for correlated events, *Communications in Statistics—Theory and Methods*, **23**, 841–857. [3.12.6]

Dubey, S. D. (1966a). Compound Pascal distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **18**, 357–365. [6.2.3, 6.7.3]

Dubey, S. D. (1966b). Graphical tests for discrete distributions, *American Statistician*, **20**, 23–24. [4.7.1]

Dunin-Barkovsky, I. V., and Smirnov, N. V. (1955). *Theory of Probability and Mathematical Statistics in Engineering*, Moscow: Nauka (in Russian). [3.6.1]

Dunkl, C. F. (1981). The absorption distribution and the q-binomial theorem, *Communications in Statistics—Theory and Methods*, **10**, 1915–1920. [10.8.1, 10.8.2]

Ebrahimi, N. (1986). Classes of discrete decreasing and increasing mean residual life distributions, *IEEE Transactions on Reliability*, **35**, 403–405. [11.2.15]

Edwards, A. W. F. (1960). The meaning of binomial distribution, *Nature, London*, **186**, 1074. [3.12.6]

Edwards, A. W. F. (1987). *Pascal's Arithmetical Triangle*, London: Griffin. [3.2]

Efron, B. (1986). Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association*, **81**, 709–721. [9.10, 11.1.1, 11.1.8]

Eggenberger, F., and Pólya, G. (1923). Über die Statistik verketteter Vorgänge, *Zeitschrift für angewandte Mathematik und Mechanik*, **3**, 279–289. [5.1, 5.3, 6.1, 6.2.4]

Eggenberger, F., and Pólya, G. (1928). Calcul des probabilités—sur l'interprétation de certaines courbes de frequence, *Comptes Rendus, Académie des Sciences, Paris*, **187**, 870–872. [5.3, 6.1, 6.2.4]

Eideh, A.-H. A.-H., and Ahmed, M. S. (1989). Some tests for the power series distributions in one parameter using the Kullback-Leibler information measure, *Communications in Statistics—Theory and Methods*, **18**, 3649–3663. [2.2.1]

Elliott, J. M. (1979). *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*, Freshwater Biological Association Scientific Publication 25, Ambleside, England: Freshwater Biological Association. [5.10]

Emerson, J. D., and Hoaglin, D. C. (1983). Stem-and-leaf displays, *Understanding Robust and Exploratory Data Analysis*, D. C. Hoaglin, F. W. Mosteller, and J. W. Tukey (editors), 7–32. New York: Wiley. [1.2.12]

Engen, S. (1974). On species frequency models, *Biometrika*, **61**, 263–270. [5.12.2, 6.11.1, 7.1.7]

Engen, S. (1978). *Stochastic Abundance Models*, London: Chapman & Hall. [5.12.2, 6.11.1]

Enneking, E., and Ahuja, J. C. (1978). On minimum variance unbiased estimation for certain left-truncated discrete distributions, *Sankhyā: The Indian Journal of Statistics, Series B*, **40**, 105–110. [7.1.10]

Enns, E. G. (1966). Derivation of the time dependent probability that a subset of identical units is operational, *Proceedings IEEE (Letters)*, **54**, 986–987. [3.12.2]

Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (1953). *Higher Transcendental Functions*, Vols. 1 and 2, New York: McGraw-Hill. [1.1.6–1.1.8, 11.2.20]

Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (1954). *Tables of Integral Transforms*, Vols. 1 and 2, New York: McGraw-Hill. [1.1.8, 1.1.10]

Ericksen, L. (2002). Sum of powers from uniform distribution moments, *Journal of Statistical Planning and Inference*, **101**, 57–79. [6.10.1]

Erlang, A. K. (1918). Solutions of some problems in the theory of probabilities of significance in automatic telephone exchanges, *Post Office Electrical Engineer's Journal*, **10**, 189–197. [11.2.9]

Estoup, J. B. (1916). *Les Gammes Sténographiques*, Paris: Institut Sténographique. [11.2.20]

Evans, D. A. (1953). Experimental evidence concerning contagious distributions in ecology, *Biometrika*, **40**, 186–211. [5.1, 9.6.5, 9.7]

Everitt, B. S. (2003). *Modern Medical Statistics*, London: Arnold. [11.1.1]

Everitt, B. S., and Hand, D. J. (1981). *Finite Mixture Distributions*, London: Chapman & Hall. [8.2.2, 8.2.5, 8.2.6]

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles, *Theoretical Population Biology*, **3**, 87–112. [7.1.12]

Exton, H. (1978). *Handbook of Hypergeometric Integrals*, Chichester: Ellis Horwood. [1.1.8]

Exton, H. (1983). *q-Hypergeometric Functions and Applications*, Chichester: Ellis Horwood. [1.1.12]

Faddy, M. J. (1994). On variation in Poisson processes, *Mathematical Scientist*, **19**, 47–51. [4.12.5]

Faddy, M. J. (1997). Extended Poisson process modelling and analysis of count data, *Biometrical Journal*, **39**, 431–440. [4.12.5]

Faddy, M. J., and Bosch, R. J. (2001). Likelihood-based modelling and analysis of data underdispersed relative to the Poisson distribution, *Biometrics*, **57**, 620–624. [4.12.5]

Famoye, F. (1993). Restricted generalized Poisson regression model, *Communications in Statistics—Theory and Methods*, **22**, 1335–1354. [7.2.6]

Famoye, F. (1995). Generalized binomial regression model, *Biometrical Journal*, **37**, 581–594. [7.2.7, 11.1.1]

Famoye, F. (1997). Parameter estimation for generalized negative binomial distribution, *Communications in Statistics—Simulation and Computation*, **26**, 269–279. [5.12.3, 7.2.7]

Famoye, F. (2000). Goodness-of-fit tests for generalized logarithmic series distribution, *Computational Statistics and Data Analysis*, **33**, 59–67. [7.2.8]

Famoye, F., and Consul, P. C. (1989). Confidence interval estimation in the class of modified power series distributions, *Statistics*, **20**, 141–148. [2.2.2, 5.12.3, 7.2.7]

Famoye, F., and Consul, P. C. (1993). The truncated generalized negative binomial distribution, *Journal of Applied Statistical Science*, **1**, 141–157. [7.2.7]

Famoye, F., and Singh, K. P. (1995). Regression models for cytogenetic dose-response data, *Journal of Applied Statistical Science*, **2**, 319–333. [7.2.6]

Fang, K.-T. (1985). Occupancy problems, *Encyclopedia of Statistical Sciences*, Vol. 6, S. Kotz, N. L. Johnson, and C. B. Read (editors), 402–406. New York: Wiley. [10.4.1]

Farrington, C. P. (1998). Communicable diseases, *Encyclopedia of Biostatistics*, P. Armitage and T. Colton (editors), Vol. 1, 795–815. Chichester: Wiley. [3.12.5]

Farrington, C. P., and Grant, A. D. (1999). The distribution of time to extinction in subcritical branching processes: Applications to outbreaks in infectious disease, *Journal of Applied Probability*, **36**, 771–779. [10.8.2]

Fazal, S. S. (1976). A test for quasi-binomial distribution, *Biometrische Zeitschrift*, **18**, 619–622. [7.2.5]

Feigin P. D., and Lumelskii, Y. P. (2000). On confidence limits for the difference of two binomial parameters, *Communications in Statistics—Theory and Methods*, **29**, 131–141. [3.8.3]

Feix, M. (1955). Théorie d'enregistrement d'évenements aléatoires, *Journal Physique Radium*, **16**, 719–727. [11.2.9]

Feldman, D., and Fox, M. (1968). Estimation of the parameter n in the binomial distribution, *Journal of the American Statistical Association*, **63**, 150–158. [3.8.2]

Feller, W. (1943). On a general class of "contagious" distributions, *Annals of Mathematical Statistics*, **14**, 389–400. [4.11, 5.9.2, 8.2.5, 8.3.2, 8.3.3, 9.0, 9.3, 9.9]

Feller, W. (1945). On the normal approximation to the binomial distribution, *Annals of Mathematical Statistics*, **16**, 319–329. [3.6.2]

Feller, W. (1948). On probability problems in the theory of counters, *Studies Presented to R. Courant*, 105–115. New York: Interscience. [11.2.9]

Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*, Vol. 1, New York: Wiley. [4.11, 11.2.7]

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications* (second edition), Vol. 1, New York: Wiley. [3.2, 4.2.1, 4.11, 5.3, 5.12.5, 6.4, 6.5, 6.10.2, 7.1.10, 8.3.2, 9.3, 10.1, 10.2, 10.4.1, 10.7.1, 10.7.5]

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (third edition), Vol. 1, New York: Wiley. [3.12.3, 4.2.1, 4.9, 4.11, 5.3, 5.12.2, 6.11.1, 10.7.2, 10.8.2]

Ferguson, T. S. (1964). A characterization of the exponential distribution, *Annals of Mathematical Statistics*, **35**, 1199–1207. [5.9.1]

Ferguson, T. S. (1965). A characterization of the geometric distribution, *American Mathematical Monthly*, **72**, 256–260. [5.9.1]

Ferguson, T. S. (1967). *Mathematical Statistics*, New York: Academic Press. [6.4]

Ferreri, C. (1997). On the ML-estimator of the positive and negative two-parameter binomial distribution, *Statistics and Probability Letters*, **33**, 129–134. [5.8.5]

Fienberg, S. E. (1982). Contingency tables, *Encyclopedia of Statistical Sciences*, Vol. 2, S. Kotz, N. L. Johnson, and C. B. Read (editors), 161–171. New York: Wiley. [4.9]

Finney, D. J. (1949). The truncated binomial distribution, *Annals of Eugenics, London*, **14**, 319–328. [3.11]

Finney, D. J., Latscha, R., Bennett, B. M., and Hsu, P. (1963). *Tables for Testing Significance in a* 2×2 *Contingency Table*, Cambridge: Cambridge University Press. [6.9.1]

Fisher, A. (1936). *The Mathematical Theory of Probabilities and Its Application to Frequency-Curves and Statistical Methods*, Vol. 1, New York: Macmillan. [3.12.3]

Fisher, R. A. (1929). Moments and product moments of sampling distributions, *Proceedings of the London Mathematical Society, Series 2*, **30**, 199–238. [1.2.8]

Fisher, R. A. (1931). Properties of the functions, (Part of introduction to) *British Association Mathematical Tables*, Vol. 1, London: British Association. [9.4]

Fisher, R. A. (1934). *Statistical Methods for Research Workers* (fifth edition), Edinburgh: Oliver and Boyd. [6.11.1]

Fisher, R. A. (1941). The negative binomial distribution, *Annals of Eugenics, London*, **11**, 182–187. [3.8.2, 5.1, 5.8.3]

Fisher, R. A. (1951). Properties of the functions, (Part of introduction to) *British Association Mathematical Tables* (third edition), Vol. 1, London: British Association. [1.1.7, 9.4]

Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population, *Journal of Animal Ecology*, **12**, 42–58. [5.12.2, 7.1.1, 7.1.2, 7.1.6, 7.1.7, 7.1.9, 11.2.16]

Fisher, R. A., and Mather, K. (1936). A linkage test with mice, *Annals of Eugenics, London*, **7**, 265–280. [3.10]

Fishman, G. S. (1973). *Concepts and Methods in Discrete Event Simulation*, New York: Wiley. [6.6]

Fishman, G. S. (1978). *Principles of Discrete Event Simulation*, New York: Wiley. [5.7, 6.6]

Fishman, G. S. (1979). Sampling from the binomial distribution on a computer, *Journal of the American Statistical Association*, **74**, 418–423. [3.7.2]

Fisz, M. (1953). The limiting distribution of the difference of two Poisson random variables, *Zastosowania Matematyki*, **1**, 41–45 (in Polish). [4.12.3]

Fisz, M., and Urbanik, K. (1956). The analytical characterisation of the composed nonhomogeneous Poisson process, *Bulletin de L'Academie Polonaise des Sciences, Classe III*, **3**, 149–150. [4.2.1]

Foster, F. G. (1952). A Markov chain derivation of discrete distributions, *Annals of Mathematical Statistics*, **23**, 624–627. [5.3]

Foster, F. G., and Stuart, A. (1954). Distribution-free tests in time-series based on the breaking of records, *Journal of the Royal Statistical Society, Series B*, **16**, 1–13. [10.5]

Fox, B. L., and Glynn, P. W. (1988). Computing Poisson probabilities, *Communications of the ACM*, **31**, 440–445. [4.6.1]

Fox, W. R., and Lasker, G. W. (1983). The distribution of surname frequencies, *International Statistical Review*, **51**, 81–87. [11.2.20]

Frame, J. S. (1945). Mean deviation of the binomial distribution, *American Mathematical Monthly*, **52**, 377–379. [3.3]

Francis, B. J. (1991). AS R88: A remark on Algorithm AS 121: The trigamma function, *Applied Statistics*, **40**, 514–515. [11.2.2]

Fréchet, M. (1940). *Les Probabilités Associées a un Système d'Événements Compatibles et Dépendants*, Vol. 1: *Événements en Nombre Fini Fixe*, Actualités scientifiques et industrielles No. 859, Paris: Hermann. [1.2.11, 2.4.2, 10.2, 10.3]

Fréchet, M. (1943). *Les Probabilités Associées a un Système d'Événements Compatibles et Dépendants*, Vol. 2: *Cas Particuliers et Applications*, Actualités scientifiques et industrielles No. 942, Paris: Hermann. [1.2.11, 2.4.2, 10.2, 10.3]

Freedman, D. A. (1965). Bernard Friedman's urn, *Annals of Mathematical Statistics*, **36**, 956–970. [11.2.19]

Freeman, G. H. (1980). Fitting two-parameter discrete distributions to many data sets with one common parameter, *Applied Statistics*, **29**, 259–267. [5.8.4]

Freeman, M. F., and Tukey, J. W. (1950). Transformations related to the angular and the square root, *Annals of Mathematical Statistics*, **21**, 607–611. [3.6.3, 4.5]

Freeman, P. R. (1973). Algorithm AS 59: Hypergeometric probabilities, *Applied Statistics*, **22**, 130–133. [6.6]

Freimer, M., Gold, B., and Tritter, A. L. (1959). The Morse distribution, *Transactions of the IRE—Information Theory*, **5**, 25–31. [10.5]

Friedman, B. (1949). A simple urn model, *Communications in Pure and Applied Mathematics*, **2**, 59–70. [11.2.19]

Frisch, R. (1924). Solution d'un problém du calcul des probabilitités, *Skandinavisk Aktuarietidskrift*, **7**, 153–174. [3.3]

Froggatt, P. (1966). One day absence in industry, *Journal of the Statistical and Social Enquiry Society of Ireland*, **21**, 166–178. [9.9]

From, S. G. (1993). Some recurrence relation-based estimators of the parameters of a generalized log-series distribution, *Communications in Statistics—Theory and Methods*, **22**, 1653–1664. [7.1.7]

From, S. G. (1995). A weighted least-squares procedure for estimating the parameters of Altham's multiplicative generalization of the binomial distribution, *Statistics and Probability Letters*, **25**, 193–199. [3.12.6]

Frost, W. H. (1976). Some conceptions of epidemics in general, *American Journal of Epidemiology*, **103**, 141–151. [3.12.5]

Fu, J., and Sproule, R. (1995). A generalization of the binomial distribution, *Communications in Statistics—Theory and Methods*, **24**, 2645–2658. [3.12.6]

Fu, J. C., and Koutras, M. V. (1995). Reliability bounds for coherent structures with independent components, *Statistics and Probability Letters*, **22**, 137–148. [10.7.2]

Fu, Y. X., and Arnold, J. (1992). A table of exact sample size for use with Fisher's exact test for 2×2 tables, *Biometrics*, **48**, 1103–1112. [6.11.1]

Furry, W. H. (1937). On fluctuation phenomena in the passage of high energy electrons through lead, *Physical Review*, **52**, 569–581. [5.2, 5.3, 5.10]

Gabriel, K. R., and Neumann, J. (1962). On a distribution of weather cycles by length, *Quarterly Journal of the Royal Meteorological Society*, **83**, 375–380; **88**, 90–95. [5.2]

Galambos, J. (1973). A general Poisson limit theorem of probability theory, *Duke Mathematical Journal*, **40**, 581–586. [3.6.1]

Galambos, J. (1975). Methods for proving Bonferroni type inequalities, *Journal of the London Mathematical Society*, **9**, 561–564. [10.2]

Galambos, J. (1977). Bonferroni inequalities, *Annals of Probability*, **5**, 577–581. [10.2]

Galambos, J., and Kotz, S. (1978). *Characterizations of Probability Distributions*, Berlin: Springer-Verlag. [4.8]

Galliher, H. P., Morse, P. M., and Simond, M. (1959). Dynamics of two classes of continuous review inventory systems, *Operations Research*, **7**, 362–384. [4.11, 9.3, 9.7]

Galloway, T. (1839). *A Treatise on Probability*, Edinburgh: Adam and Charles Black. [5.3]

Gani, J. (1957). Problems in the probability theory of storage systems, *Journal of the Royal Statistical Society, Series B*, **19**, 182–206. [6.11.1]

Gani, J. (1975). Stochastic models for type counts in a literary text, *Perspectives in Probability and Statistics*, J. Gani (editor), 313–324. London: Academic Press. [3.12.3]

Gart, J. J. (1963). A median test with sequential application, *Biometrika*, **50**, 55–62. [6.9.2]

Gart, J. J. (1968). A simple nearly efficient alternative to the simple sib method in the complete ascertainment case, *Annals of Human Genetics, London*, **31**, 283–291. [3.11]

Gart, J. J. (1970). Some simple graphically oriented statistical methods for discrete data, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 171–191. University Park: Pennsylvania State University Press. [3.8.1, 4.7.1, 7.1.7]

Gart, J. J. (1974). Exact moments of the variance test for left-truncated Poisson distributions, *Sankhyā, Series B*, **36**, 406–416. [4.7.4, 4.12.3]

Gart, J. J., and Pettigrew, H. M. (1970). On the conditional moments of the k-statistics for the Poisson distribution, *Biometrika*, **57**, 661–664. [4.4]

Gasper, G., and Rahman, M. (1990). *Basic Hypergeometric Series*, Cambridge: Cambridge University Press. [1.1.12, 2.9, 10.8, 10.8.1–10.8.3]

Gastwirth, J. L., and Greenhouse, S. W. (1987). Estimating a common relative risk: Application in equal employment, *Journal of the American Statistical Association*, **82**, 38–45. [6.11.1]

Gauss, C. F. (1900). Schönes Theorem der Wahrscheinlichkeitsrechnung, *Werke*, **8**, 88. Koniglichen Gesellenschaft der Wissenschaften zu Göttingen, Leipzig: Teubner. [1.2.10]

Gebhardt, F. (1969). Some numerical comparisons of several approximations to the binomial distribution, *Journal of the American Statistical Association*, **64**, 1638–1648. [3.6.1]

Gebhardt, F. (1971). Incomplete beta-integral $B(x;\, 2/3,\, 2/3)$ and $[p(1-p)]^{-1/6}$ for use with Borges' approximation to the binomial distribution, *Journal of the American Statistical Association*, **66**, 189–191. [3.6.1]

Geiringer, H. (1938). On the probability theory of arbitrarily linked events, *Annals of Mathematical Statistics*, **9**, 260–271. [10.2]

Geisser, S. (1984). On prior distributions for binary trials, *Annals of Statistics*, **38**, 244–247. [3.8.2]

Gelfand, A. E., and Soloman, H. (1975). Analysing the decision making process of the American jury, *Journal of the American Statistical Association*, **70**, 305–310. [8.2.6]

Gentle, J. E. (1998). *Random Number Generation and Monte Carlo Methods*, New York: Springer. [1.2.17]

Gentle, J. E. (2002). *Elements of Computational Statistics*, New York: Springer. [1.2.17]

George, E. O., and Bowman, D. (1995). A full likelihood procedure for analysing exchangeable binary data, *Biometrics*, **51**, 512–523. [3.12.6]

Gerrard, D. J., and Cook, R. D. (1972). Inverse binomial sampling as a basis for estimating negative binomial population densities, *Biometrics*, **28**, 971–980. [5.8.2]

Gerstenkorn, T. (1962). On the generalized Poisson distribution, *Prace Lódzkie Towarzystwa Naukowe, Wydzial*, **3**(85), 7–41. [4.12.5, 11.2.13]

Ghosh, B. K. (1980). The normal approximation to the binomial distribution, *Communications in Statistics—Theory and Methods*, **A9**, 427–438. [3.6.1]

Ghosh, M., and Yang, M.-C. (1988). Simultaneous estimation of Poisson means under entropy loss, *Annals of Statistics*, **16**, 278–291. [4.7.2]

Gibbons, J. D. (1983). Fisher's exact test, *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz, N. L. Johnson, and C. B. Read (editors), 130–133. New York: Wiley. [6.9.1]

Gibbons, J. D. (1986). Randomness, tests of, *Encyclopedia of Statistical Sciences*, Vol. 7, S. Kotz, N. L. Johnson, and C. B. Read (editors), 555–562. New York: Wiley. [10.6.1, 10.6.2]

Gilbert, E. J. (1956). The matching problem, *Psychometrika*, **21**, 253–266. [10.3]

Giltay, J. (1943). A counter arrangement with constant resolving time, *Physica*, **10**, 725–734. [11.2.9]

Giorno, V., Negri, C., and Nobile, A. G. (1985). A solvable model for a finite capacity queuing system, *Journal of Applied Probability*, **22**, 903–911. [3.12.2]

Glasser, G. J. (1962). Minimum variance unbiased estimators for Poisson probabilities, *Technometrics*, **4**, 409–418. [4.7.2]

Gleeson, A. C., and Douglas, J. B. (1975). Quadrat sampling and the estimation of Neyman type A and Thomas distributional parameters, *Australian Journal of Statistics*, **17**, 103–113. [9.10]

Glick, N. (1978). Breaking records and breaking boards, *American Mathematical Monthly*, **85**, 2–26. [10.5]

Gnedenko, B. V. (1961). *Probability Theory* (third edition), Moscow: GITTL. [5.3]

Gnedenko, B. V., and Kovalenko, I. N. (1989). *Introduction to Queueing Theory* (second edition), translated by S. Kotz, Boston: Birkhäuser. [4.9]

Göb, R. (1994) . Bounds for median and 50 percentage point of binomial and negative binomial distribution, *Metrika*, **41**, 43–54. [5.5]

Godambe, A. V. (1977). On representation of Poisson mixtures as Poisson sums and a characterization of the gamma distribution, *Mathematical Proceedings of the Cambridge Philosophical Society*, **82**, 297–300. [8.2.5]

Godambe, A. V., and Patil, G. P. (1975). Some characterisations involving additivity and infinite divisibility and their applications to Poisson mixtures and Poisson sums, *Statistical Distributions in Scientific Work*, Vol. 3: *Characterizations and Applications*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 339–351. Dordrecht: Reidel. [4.11, 8.3.2]

Godbole, A. P. (1990a). Specific formulae for some success run distributions, *Statistics and Probability Letters*, **10**, 119–124. [10.7.4]

Godbole, A. P. (1990b). On hypergeometric and related distributions of order k, *Communications in Statistics—Theory and Methods*, **19**, 1291–1301. [10.7.6]

Godwin, H. J. (1967). Review of T. Gerstenkorn: On the generalized Poisson distribution, *Mathematical Reviews*, **34**(3608). [4.12.5]

Gokhale, D. V. (1980). On power series distributions with mean-variance relationships, *Journal of the Indian Statistical Association*, **18**, 81–84. [2.2.2, 4.8]

Gold, L. (1957). Generalized Poisson distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **9**, 43–47. [4.12.5, 11.2.13]

Goldberg, K., Leighton, F. T., Newman, M., and Zuckerman, S. L. (1976). Tables of binomial coefficients and Stirling numbers, *Journal of Research of the National Bureau of Standards*, **80B**, 99–171. [1.1.3]

Goldie, C. M. (1995). Asymptotics of the Luria-Delbrück distribution, *Journal of Applied Probability*, **32**, 840–841. [11.2.11]

Goldie, C. M., and Rogers, L. C. G. (1984). The k-record processes are i.i.d., *Zeitschrift für Wahrscheinlichskeitstheorie und Verwandte Gebeite*, **67**, 197–211. [10.5]

Goldstein, L. (1990). Poisson approximation and DNA sequence matching, *Communications in Statistics—Theory and Methods*, **19**, 4167–4179. [4.9, 10.1]

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters, *Biometrika*, **40**, 237–264. [4.7.2, 6.11.1, 11.2.20]

Good, I. J. (1957). Distribution of word frequencies, *Nature*, **179**, 595. [6.11.2, 11.2.20]

Goodhardt, G. J., Ehrenberg, A. S. C., and Chatfield, C. (1984). The Dirichlet: A comprehensive model of buying behaviour, *Journal of the Royal Statistical Society, Series A*, **147**, 621–655. [5.10]

Gould, H. W. (1962). Congruences involving sums of binomial coefficients and a formula of Jensen, *Mathematical Notes,* May, 400–402. [2.6]

Govindarajulu, Z. (1962). The reciprocal of the decapitated negative binomial variable, *Journal of the American Statistical Association*, **57**, 906–913 (erratum **58**, 1162). [5.11, 6.11.2]

Gower, J. C. (1961). A note on some asymptotic properties of the logarithmic series distribution, *Biometrika*, **48**, 212–215. [7.1.4, 7.1.5]

Grab, E. L., and Savage, I. R. (1954). Tables of the expected value of $1/x$ for positive Bernoulli and Poisson variables, *Journal of the American Statistical Association*, **49**, 169–177. [3.11, 4.10.1]

Gradsteyn, I. S., and Ryzhik, I. M. (1980). *Tables of Integrals, Series and Products* (enlarged edition), London: Academic Press. [11.2.20]

Grandell, J. (1997). *Mixed Poisson Processes*, New York: Chapman & Hall/CRC. [8.1.2, 8.3.2]

Grassia, A. (1977). On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions, *Australian Journal of Statistics*, **19**, 108–114. [10.4.3]

Greenberg, R. A., and White, C. (1965). The detection of a correlation between the sexes of adjacent sibs in human families, *Journal of the American Statistical Association*, **60**, 1035–1045. [3.12.6]

Greenwood, M. (1931). The statistical measure of infectiveness, *Journal of Hygiene*, **31**, 336–351. [3.12.5]

Greenwood, M., and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *Journal of the Royal Statistical Society, Series A*, **83**, 255–279. [4.12.5, 5.3, 5.10, 8.3.3, 9.4, 11.2.19]

Greenwood, R. E., and Glasgow, M. O. (1950). Distribution of maximum and minimum frequencies in a sample drawn from a multinomial population, *Annals of Mathematical Statistics*, **21**, 416–424. [3.8.2]

Greig-Smith, P. (1964). *Quantitative Plant Ecology* (second edition), London: Butterworth. [4.9]

Greville, T. N. E. (1941). The frequency distribution of a general matching problem, *Annals of Mathematical Statistics*, **12**, 350–354. [10.3]

Griffiths, D. A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease, *Biometrics*, **29**, 637–648. [6.7.2, 6.9.2]

Grimm, H. (1962). Tafeln der negativen Binomialverteilung, *Biometrische Zeitschrift*, **4**, 239–262. [5.7]

Grimm, H. (1964). Tafeln der Neyman-Verteilung Typ A, *Biometrische Zeitschrift*, **6**, 10–23. [9.6.3]

Grimm, H. (1970). Graphical methods for the determination of type and parameters of some discrete distributions, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 193–206. University Park: Pennsylvania State University Press. [3.8.1, 4.7.1, 5.8.1, 9.6.4]

Gross, D., and Harris, C. M. (1985). *Fundamentals of Queuing Theory* (second edition), New York: Wiley. [3.12.2]

Guenther, W. C. (1969). Use of the binomial, hypergeometric and Poisson tables to obtain sampling plans, *Journal of Quality Technology*, **1**, 105–109. [6.9.2]

Guenther, W. C. (1971). The average sample number for truncated double sample attribute plans, *Technometrics*, **13**, 811–816. [6.9.2]

Guenther, W. C. (1972). A simple approximation to the negative binomial (and regular binomial), *Technometrics*, **14**, 385–389. [5.6]

Guenther, W. C. (1975). The inverse hypergeometric—a useful model, *Statistica Neerlandica*, **29**, 129–144. [6.2.2, 6.9.2]

Guenther, W. C. (1977). *Sampling Inspection in Statistical Quality Control*, London: Griffin. [6.7.1]

Guenther, W. C. (1983). Hypergeometric distributions, *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz, N. L. Johnson and C. B. Read (editors), 707–712. New York: Wiley. [6.2.5, 6.6, 6.7.1]

Guess, F. M., and Park, D. H. (1988). Modeling discrete bathtub and upside-down bathtub mean residual-life functions, *IEEE Transactions on Reliability*, **37**, 545–549. [11.2.15]

Guldberg, A. (1931). On discontinuous frequency functions and statistical series, *Skandinavisk Aktuarietidskrift*, **14**, 167–187. [2.3.1, 2.3.3, 2.4.1]

Gumbel, E. J. (1958). *Statistics of Extremes*, New York: Columbia University Press. [6.2.2]

Gumbel, E. J., and von Schelling, H. (1950). The distribution of the number of exceedances, *Annals of Mathematical Statistics*, **21**, 247–262. [6.2.2]

Gupta, D., and Gupta, R. C. (1984). On the distribution of order statistics for a random sample size, *Statistica Neerlandica*, **38**, 13–19. [7.2.6]

Gupta, M. K. (1967). Unbiased estimate for $1/p$, *Annals of the Institute of Statistical Mathematics, Tokyo*, **19**, 413–416. [3.8.2]

Gupta, P. L. (1982). Probability generating functions of a MPSD with applications, *Mathematische Operationforschung und Statistik, Series Statistics*, **13**, 99–103. [2.2.2]

Gupta, P. L., and Gupta, R. C. (1981). Probability of ties and Markov property in discrete order statistics, *Journal of Statistical Planning and Inference*, **5**, 273–279. [5.2]

Gupta, P. L., and Gupta R. C. (1986). A characterization of the Poisson process, *Journal of Applied Probability*, **23**, 233–235. [4.8]

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1995). Inflated modified power series distributions with applications, *Communications in Statistics—Theory and Methods*, **24**, 2355–2374. [8.2.4]

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1996). Analysis of zero-adjusted count data, *Computational Statistics and Data Analysis*, **23**, 207–218 . [8.2.4]

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1997). On the monotone properties of discrete failure rates, *Journal of Statistical Planning and Inference*, **65**, 255–268. [1.2.4, 5.12.2, 7.1.4, 10.8.2, 11.2.15]

Gupta, P. L., and Singh, J. (1981). On the moments and factorial moments of a MPSD, *Statistical Distributions in Scientific Work*, Vol. 4: *Models, Structures, and Characterizations*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 189–195. Dordrecht: Reidel. [2.2.2, 7.2.6]

Gupta, R. C. (1970). *Some Characterizations of Probability Distributions*, PhD Thesis. Detroit, MI: Wayne State University. [5.9.1]

Gupta, R. C. (1974). Modified power series distributions and some of its applications, *Sankhyā, Series B*, **35**, 288–298. [2.2.2, 5.12.3, 7.2.6, 7.2.7]

Gupta, R. C. (1975a). Maximum-likelihood estimation of a modified power series distribution and some of its applications, *Communications in Statistics*, **4**, 689–697. [2.2.2]

Gupta, R. C. (1975b). Some characterizations of discrete distributions by properties of their moment distributions, *Communications in Statistics*, **4**, 761–765. [2.2.2]

Gupta, R. C. (1976). Distribution of the sum of independent generalized logarithmic series variables, *Communications in Statistics—Theory and Methods*, **5**, 45–48. [7.2.8]

Gupta, R. C. (1977a). Minimum variance unbiased estimation in modified power series distribution and some of its applications, *Communications in Statistics—Theory and Methods*, **A6**, 977–991. [2.2.2, 7.2.8]

Gupta, R. C. (1977b). On characterizing distributions by the ratio of variance and mean, *Mathematische Operationsforschung und Statistik*, **8**, 523–527. [4.8]

Gupta, R. C. (1981). On the Rao-Rubin characterization of the Poisson distribution, *Statistical Distributions in Scientific Work*, Vol. 4: *Models, Structures, and Characterizations*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 341–347. Dordrecht: Reidel. [4.8]

Gupta, R. C. (1984). Estimating the probability of winning (losing) in a gambler's ruin problem with applications, *Journal of Statistical Planning and Inference*, **9**, 55–62. [2.2.2, 11.2.10]

Gupta, R. C., and Singh, J. (1982). Estimation of probabilities in the class of modified power series distributions, *Mathematische Operationforschung und Statistik, Series Statistics*, **13**, 71–77. [2.2.2]

Gupta, R. C., and Tripathi, R. C. (1985). Modified power series distributions, *Encyclopedia of Statistical Sciences*, Vol. 5, S. Kotz, N. L. Johnson and C. B. Read (editors), 593–599. New York: Wiley. [2.2.2]

Gupta, R. C., and Tripathi, R. C. (1992). Statistical inference based on the length-biased data for the modified power series distributions, *Communications in Statistics—Theory and Methods*, **21**, 519–537. [2.2.2]

Gupta, R. P., and Jain, G. C. (1974). A generalized Hermite distribution and its properties, *SIAM Journal of Applied Mathematics*, **27**, 359–363. [9.4]

Gupta, S. S. (1960a). Order statistics from the gamma distribution, *Technometrics*, **2**, 243–262. [5.5]

Gupta, S. S. (1960b). *Binomial Order Statistics*, Allentown, PA: Bell Laboratory. [3.5]

Gupta, S. S. (1965). Selection and ranking procedures and order statistics for the binomial distribution, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 219–230. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [3.5]

Gupta, S. S., and Liang T. (2002). Selecting the most reliable Poisson population provided it is better than a control: A nonparametric empirical Bayes approach, *Journal of Statistical Planning and Inference*, **103**, 191–203. [4.7.2]

Gupta, S. S., and Panchapakesan, P. (1974). On moments of order statistics from independent binomial populations, *Annals of the Institute of Statistical Mathematics, Tokyo*, **8**(Suppl.), 95–113. [3.5]

Gurland, J. (1957). Some interrelations among compound and generalized distributions, *Biometrika*, **44**, 265–268. [8.3.2, 9.0, 9.1]

Gurland, J. (1958). A generalized class of contagious distributions, *Biometrics*, **14**, 229–249. [2.4.2, 6.11.3, 8.3.3, 8.4, 9.9, 9.13]

Gurland, J. (1965). A method of estimation for some generalized Poisson distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 141–158. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [5.8.1, 5.8.3, 9.3, 9.6.4]

Gurland, J., Chen, E. E., and Hernandez, F. M. (1983). A new discrete distribution involving Laguerre polynomials, *Communications in Statistics—Theory and Methods*, **A12**, 1987–2004. [5.12.5, 9.12]

Gurland, J., and Tripathi, R. C. (1975). Estimation of parameters on some extensions of the Katz family of discrete distributions involving hypergeometric functions, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 59–82. Dordrecht: Reidel. [2.3.1, 4.12.4, 5.8.3, 6.11.4]

Gurmu, S., and Trivedi, P. K. (1992). Overdispersion tests for truncated Poisson regression models, *Journal of Econometrics*, **54**, 347–370. [5.11]

Gürtler, N., and Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution, *Journal of Statistical Planning and Inference*, **90**, 207–225. [4.7.4]

Gutiérrez-Jáimez, R., and Rodríguez Avi, J. (1997). Family of Pearson discrete distributions generated by the univariate hypergeometric function ${}_3F_2(\alpha_1; \alpha_2; \alpha_3; \gamma_1; \gamma_2; \lambda)$, *Applied Stochastic Models and Data Analysis*, **13**, 115–125. [6.11.2]

Guttman, I. (1958). A note on a series solution of a problem in estimation, *Biometrika*, **45**, 565–567. [5.8.2]

Guzman, J. (1985). Some comments on the Poisson approximation to the binomial distribution, *American Statistician*, **39**, 157–158. [3.6.2]

Haag, J. (1924). Sur un probléme général de probabilités et ses applications, *Proceedings of the International Mathematical Congress, Toronto*, **1**, 659–676. [10.3]

Hadley, G., and Whitin, T. M. (1961). Useful properties of the Poisson distribution, *Operations Research*, **9**, 408–410. [4.4]

Hahn, G. J., and Meeker, W. Q. (1991). *Statistical Intervals*, New York: Wiley. [1.2.16]

Haight, F. A. (1957). Queueing with balking, *Biometrika*, **44**, 360–369. [3.2, 5.3, 11.2.20]

Haight, F. A. (1961a). A distribution analogous to the Borel-Tanner, *Biometrika*, **48**, 167–173. [7.2.1, 11.2.10, 11.2.14]

Haight, F. A. (1961b). Index to the distributions of mathematical statistics, *Journal of Research of the National Bureau of Standards*, **65B**, 23–60. [9.3, 11.2.13]

Haight, F. A. (1965). On the effect of removing persons with N or more accidents from an accident prone population, *Biometrika*, **52**, 298–300. [2.4.2, 8.3.3]

Haight, F. A. (1966). Some statistical problems in connection with word association data, *Journal of Mathematical Psychology*, **3**, 217–233. [6.10.3, 11.2.21]

Haight, F. A. (1967). *Handbook of the Poisson Distribution*, New York: Wiley. [4.2.1, 4.3, 4.4, 4.9, 5.12.8]

Haight, F. A. (1969). Two probability distributions connected with Zipf's rank-size conjecture, *Zastosowania Matematyki*, **10**, 225–228. [11.2.21]

Haight, F. A. (1972). The characterizations of discrete distributions by conditional distributions, *Zastosowania Matematyki*, **13**, 207–213. [3.9, 4.8]

Haight, F. A., and Breuer, M. A. (1960). The Borel-Tanner distribution, *Biometrika* **47**, 145–150. [7.2, 7.2.2, 9.11]

Hald, A. (1952). *Statistical Theory with Engineering Applications*, New York: Wiley. [3.6.1, 5.3]

Hald, A. (1960). The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs, *Technometrics*, **2**, 275–340. [6.8, 6.9.2, 8.3.6]

Hald, A. (1967). The determination of single sampling attribute plans with given producer's and consumer's risk, *Technometrics*, **9**, 401–415. [3.6.1]

Hald, A. (1968). The mixed binomial distribution and the posterior distribution of p for a continuous prior distribution, *Journal of the Royal Statistical Society, Series B*, **30**, 359–367. [8.3.4]

Hald, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*, New York: Wiley. [3.2, 10.2, 10.4.1]

Hald, A., and Kousgaard, E. (1967). A table for solving the binomial equation $B(c, n, p) = P$, *Matematisk-fysiske Skrifter, Det Kongelige Danske Videnskabernes Selskab*, **3**(4), 1–48. [3.8.3, 4.6.1]

Haldane, J. B. S. (1941). The fitting of binomial distributions, *Annals of Eugenics, London*, **11**, 179–181. [3.8.2]

Haldane, J. B. S. (1949). Some statistical problems arising in genetics, *Journal of the Royal Statistical Society, Series B*, **11**, 1–14. [11.2.16]

Hall, P. (1983). On the roles of the Bessel and Poisson distributions in chemical kinetics, *Journal of Applied Probability*, **20**, 585–599. [6.11.2]

Hall, W. J. (1956). Some hypergeometric series distributions occurring in birth-and-death processes at equilibrium, *Annals of Mathematical Statistics*, **27**, 221 (abstract). [4.12.4]

Hamdan, M. A. (1975). Correlation between the numbers of two types of children when the family size distribution is zero-truncated negative binomial, *Biometrics*, **31**, 765–766. [5.11]

Hamedani, G. G., and Walter, G. G. (1990). Empirical Bayes estimation of the binomial parameter n, *Communications in Statistics—Theory and Methods*, **19**, 2065–2084. [3.8.2]

Hamza, K. (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions, *Statistics and Probability Letters*, **23**, 21–25. [3.4]

Handa, B. R., and Mohanty, S. G. (1980). On q-binomial coefficients and some statistical applications, *SIAM Journal of Mathematical Analysis*, **11**, 1027–1035. [2.9]

Hannan, J., and Harkness, W. (1963). Normal approximation to the distribution of two independent binomials, conditional on a fixed sum, *Annals of Mathematical Statistics*, **34**, 1593–1595. [3.4, 6.11.1]

Hansen, B. G., and Willekens, E. (1990). The generalized logarithmic series distribution, *Statistics and Probability Letters*, **9**, 311–316. [7.2.8]

Harkness, W. L. (1965). Properties of the extended hypergeometric distribution, *Annals of Mathematical Statistics*, **36**, 938–945. [6.11.1]

Harkness, W. L. (1970). The classical occupancy problem revisited, *Random Counts in Scientific Work*, Vol. 3: *Random Counts in Physical Science, Geo Science, and Business*, G. P. Patil (editor), 107–126. University Park: Pennsylvania State University Press. [10.4.1, 10.4.3]

Harris, C. M. (1983). On finite mixtures of geometric and negative binomial distributions, *Communications in Statistics—Theory and Methods*, **12**, 987–1007. [8.2.7]

Hart, J. F. (1968). *Computer Approximations*, New York: Wiley. [4.5]

Harter, H. L. (1964). *New Tables of Incomplete Gamma Function Ratio*, United States Air Force Aerospace Research Laboratories, Washington, D.C.: Government Printing Office. [1.1.5]

Harter, H. L. (1988). History and role of order statistics, *Communications in Statistics—Theory and Methods*, **17**, 2091–2107. [1.2.12]

Hartley, H. O., and Fitch, E. R. (1951). A chart for the incomplete beta function and cumulative binomial distribution, *Biometrika*, **38**, 423–426. [3.7.1]

Haseman, J. K., and Kupper, L. L. (1979). Analysis of dichotomous response data from certain toxicological experiments, *Biometrics*, **35**, 281–293. [3.12.6, 6.9.2]

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family, *Journal of the American Statistical Association*, **64**, 1459–1471. [8.2.5, 8.2.6]

Hauck, W. W. (1984). A comparative study of conditional maximum likelihood estimation of a common odds ratio, *Biometrics*, **40**, 1117–1123. [6.11.1]

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents R&D relationship, *Econometrica*, **52**, 909–938. [4.9]

Hawkins, D. M., and Kotz, S. (1976). A clocking property of the exponential distribution, *Australian Journal of Statistics*, **18**, 170–172. [5.9.1]

Heffernan, P. M. (1996). Improved sequential probability ratio tests for negative binomial populations, *Biometrics*, **52**, 152–157. [5.8.5]

Heilbron, D. C. (1981). The analysis of ratios of odds ratios in stratified contingency tables, *Biometrics*, **37**, 55–66. [6.11.1]

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros, *Biometrical Journal*, **36**, 531–547. [8.2.3, 11.1.1]

Hemelrijk, J. (1967). The hypergeometric, the normal and chi-squared, *Statistica Neerlandica*, **21**, 225–229. [6.5]

Herdan, G. (1961). A critical examination of Simon's model of certain distribution functions in linguistics, *Applied Statistics*, **10**, 65–76. [6.10.3]

Heyde, C. C., and Seneta, E. (1972). The simple branching process, a turning point test and a fundamental inequality: A historical note on I. J. Bienaymé, *Biometrika*, **59**, 680–683. [9.1]

Hinde, J. (1982). Compound Poisson regression models, *GLIM 82: Proceedings of the International Conference on Generalized Linear Models,* R. Gilchrist (editor), 109–121. New York: Springer. [11.1.4]

Hinz, P., and Gurland, J. (1967). Simplified techniques for estimating parameters of some generalized Poisson distributions, *Biometrika*, **54**, 555–566. [3.8.1, 5.8.1, 5.8.3, 9.5, 9.6.4, 9.7, 9.8]

Hinz, P., and Gurland, J. (1968). A method of analysing untransformed data from the negative binomial and other contagious distributions, *Biometrika*, **55**, 163–170. [9.6.4]

Hinz, P., and Gurland, J. (1970). A test of fit for the negative binomial and other contagious distributions, *Journal of the American Statistical Association*, **65**, 887–903. [9.6.4]

Hirano, K. (1986). Some properties of the distributions of order k, *Fibonacci Numbers and Their Applications*, A. N. Philippou, G. E. Bergum, and A. F. Horadam (editors), 43–53. Dordrecht: Reidel. [10.7.1, 10.7.4–10.7.6]

Hirano, K., and Aki, S. (1987). Properties of the extended distributions of order k, *Statistics and Probability Letters*, **6**, 67–69. [4.11, 10.7.3, 10.7.6]

Hirano, K., Aki, S., Kashiwagi, N., and Kuboki, H. (1991). On Ling's binomial and negative binomial distributions of order k, *Statistics and Probability Letters*, **11**, 503–509. [10.7.4, 10.7.5]

Hirano, K., Kuboki, H., Aki, S., and Kuribayashi, A. (1984). *Figures of Probability Density Functions in Statistics II—Discrete Univariate Case*, Computer Science Monograph **20**, Tokyo: Institute of Statistical Mathematics. [7.1.12, 10.7.2, 10.7.3]

Hitha, N., and Nair, N. U. (1989). Characterization of some discrete models by properties of residual life function, *Bulletin of the Calcutta Statistical Association*, **38**, 219–223. [11.2.13]

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (editors) (1985). *Exploring Data Tables, Trends, and Shapes*, New York: Wiley. [4.7.1, 5.8.1, 7.1.7]

Hoaglin, D. C., and Tukey, J. W. (1985). Checking the shape of discrete distributions, *Exploring Data Tables, Trends, and Shapes*, D. C. Hoaglin, F. Mosteller, and J. W. Tukey (editors), 345–416. New York: Wiley. [3.8.1]

Hoel, P. G. (1947). Discriminating between binomial distributions, *Annals of Mathematical Statistics*, **18**, 556–564. [3.8.2, 3.8.3]

Hogg, R. V., and Klugman, S. A. (1984). *Loss Distributions*, New York: Wiley. [7.2.8]

Holgate, P. (1964). A modified geometric distribution arising in trapping studies, *Acta Theriologica*, **9**, 353–356. [3.12.2, 8.2.4]

Holgate, P. (1966). Contributions to the mathematics of animal trapping, *Biometrics*, **22**, 925–936. [3.12.2, 8.2.4]

Holgate, P. (1970). The modality of some compound Poisson distributions, *Biometrika*, **57**, 666–667. [8.3.2]

Holgate, P. (1989). Approximate moments of the Adès distribution, *Biometrical Journal*, **31**, 875–883. [11.2.3]

Holla, M. S. (1966). On a Poisson-inverse Gaussian distribution, *Metrika*, **11**, 115–121. [8.3.3, 11.1.5]

Holla, M. S., and Bhattacharya, S. K. (1965). On a discrete compound distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **15**, 377–384. [7.1.11, 8.4]

Holland, P. W. (1989). A note on the covariance of the Mantel-Haenszel log-odds-ratio estimator and the sample marginal rates, *Biometrics*, **45**, 1009–1016. [6.11.1]

Holst, L. (1986). On birthday, collectors', occupancy and other classical urn problems, *International Statistical Review*, **54**, 15–27. [10.4.1]

Hopkins, J. W. (1955). An instance of negative hypergeometric sampling in practice, *Bulletin of the International Statistical Institute*, **34**(4), 298–306. [6.9.2]

Hoppe, F. M., and Seneta, E. (1990). A Bonferroni-type identity and permutation bounds, *International Statistical Review*, **58**, 253–261. [10.2]

Horsnell, G. (1957). Economic acceptance sampling schemes, *Journal of the Royal Statistical Society, Series A*, **120**, 148–191. [8.3.5, 8.3.6]

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions, *Biometrika*, **73**, 387–396 (erratum **75**, 395). [8.3.3, 11.1.2]

Hougaard, P., Lee, M.-L. T., and Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, *Biometrics*, **53**, 1225–1238. [8.3.3, 11.1.2, 11.1.4]

Høyland, A., and Rausand, M. (1994). *System Reliability: Models and Statistical Methods*, New York: Wiley. [11.2.15]

Huang, M.-L., and Fung, K. Y. (1989). Intervened truncated Poisson distribution, *Sankhyā, Series B*, **51**, 302–310. [4.12.3, 4.12.7]

Hubbard, D. J., and Allen, O. B. (1991). Robustness of the SPRT for a negative binomial to misspecification of the dispersion parameter, *Biometrics*, **47**, 419–427. [5.8.5]

Hudson, H. M., and Tsui, K.-W. (1981). Simultaneous Poisson estimators for á priori hypotheses about means, *Journal of the American Statistical Association*, **76**, 182–187. [4.7.2]

Hunter, D. (1976). An upper bound for the probability of a union, *Journal of Applied Probability*, **13**, 597–603. [10.2]

Hunter, J. J. (1983a). Filtering of Markov renewal queues, I: Feedback queues, *Advances in Applied Probability*, **15**, 349–375. [9.12]

Hunter, J. J. (1983b). Filtering of Markov renewal queues, II: Birth-death queues, *Advances in Applied Probability*, **15**, 376–391. [9.12]

Hurwitz, H., and Kac, M. (1944). Statistical analysis of certain types of random functions, *Annals of Mathematical Statistics*, **15**, 173–181. [4.2.1]

Hwang, J. S., and Lin, G. D. (1984). Characterizations of distributions by linear combinations of order statistics, *Bulletin of the Institute of Mathematics, Academia Sinica*, **12**, 179–202. [1.2.12]

Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases, *Annals of Statistics*, **10**, 857–867. [4.7.2]

Ijiri, Y., and Simon, H. A. (1977). *Skew Distributions and the Size of Business Firms*, Amsterdam: North-Holland. [11.2.20]

Irony, T. Z. (1992). Bayesian estimation for discrete distributions, *Journal of Applied Statistics*, **19**, 533–549. [3.8.2, 4.7.2, 5.8.2]

Irwin, J. O. (1937). The frequency distribution of the difference between two independent variates following the same Poisson distribution, *Journal of the Royal Statistical Society, Series A*, **100**, 415–416. [4.12.3]

Irwin, J. O. (1941). Discussion on Chambers and Yule's paper, *Journal of the Royal Statistical Society Supplement*, **7**, 101–107. [5.3]

Irwin, J. O. (1953). On the "transition probabilities" corresponding to any accident distribution, *Journal of the Royal Statistical Society, Series B*, **15**, 87–89. [4.12.5]

Irwin, J. O. (1954). A distribution arising in the study of infectious diseases, *Biometrika*, **41**, 266–268. [6.2.2, 6.9.2, 6.10.1]

Irwin, J. O. (1955). A unified derivation of some well-known frequency distributions of interest in biometry and statistics, *Journal of the Royal Statistical Society, Series A*, **118**, 389–404. [10.2, 10.3]

Irwin, J. O. (1959). On the estimation of the mean of a Poisson distribution from a sample with the zero class missing, *Biometrics*, **15**, 324–326. [4.10.1]

Irwin, J. O. (1963). The place of mathematics in medical and biological statistics, *Journal of the Royal Statistical Society, Series A*, **126**, 1–44. [3.12.2, 6.2.3, 6.10.4, 9.4, 11.2.7]

Irwin, J. O. (1964). The personal factor in accidents—A review article, *Journal of the Royal Statistical Society, Series A*, **127**, 438–451. [9.6.5]

Irwin, J. O. (1967). William Allen Whitworth and a hundred years of probability, *Journal of the Royal Statistical Society, Series A*, **130**, 147–176. [10.2]

Irwin, J. O. (1968). The generalized Waring distribution applied to accident theory, *Journal of the Royal Statistical Society, Series A*, **131**, 205–225. [6.2.3, 6.9.3]

Irwin, J. O. (1975a). The generalized Waring distribution, Part I, *Journal of the Royal Statistical Society, Series A*, **138**, 18–31. [6.2.3]

Irwin, J. O. (1975b). The generalized Waring distribution, Part II, *Journal of the Royal Statistical Society, Series A*, **138**, 204–227. [6.2.3, 6.7.3]

Irwin, J. O. (1975c). The generalized Waring distribution, Part III, *Journal of the Royal Statistical Society, Series A*, **138**, 374–384. [6.2.3]

Ishii, G., and Hayakawa, R. (1960). On the compound binomial distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **12**, 69–80. [6.9.2]

Ising, E. (1925). Beitrag zür Theorie des Ferromagnetismus, *Zeitschrift für Physik*, **31**, 253–258. [10.6.1]

Ivchenko, G. I. (1974). On comparison of binomial and Poisson distributions, *Theory of Probability and Its Applications*, **19**, 584–586. [3.6.1]

Iwase, K. (1986). UMVU estimator of the probability in the geometric distribution with unknown truncation parameter, *Communications in Statistics—Theory and Methods*, **15**, 2449–2453. [5.2]

Iyer, P. V. K. (1949). Calculation of factorial moments of certain probability distributions, *Nature, London*, **164**, 282. [10.2]

Iyer, P. V. K. (1958). A theorem on factorial moments and its applications, *Annals of Mathematical Statistics*, **29**, 254–261. [10.2]

Jackson, F. H. (1921). Summation of q-hypergeometric series, *Messenger of Mathematics*, **50**, 101–112. [1.1.12]

Jackson, R. R. P., and Nichols, D. G. (1956). Some equilibrium results for the queuing process $E_k/M/1$, *Journal of the Royal Statistical Society, Series B*, **18**, 275–279. [3.12.2]

Jagers, P. (1973). How many people pay their tram fares? *Journal of the American Statistical Association*, **68**, 801–804. [5.2]

Jaggia, S., and Thosar, S. (1993). Multiple bids as a consequence of target management resistance: A count data approach, *Review of Quantitative Finance and Accounting*, **3**, 447–457. [11.1.1]

Jain, G. C., and Consul, P. C. (1971). A generalized negative binomial distribution, *SIAM Journal of Applied Mathematics*, **21**, 501–513. [2.6, 5.8.5, 5.12.3, 7.2, 7.2.7]

Jain, G. C., and Gupta, R. P. (1973). A logarithmic series type distribution, *Trabajos de Estadística*, **24**, 99–105. [2.2.2, 7.1.11, 7.2, 7.2.8]

Jain, G. C., and Plunkett, I. G. (1977). A Borel-Hermite distribution and its applications, *Biometrical Journal*, **19**, 347–354. [9.4]

Janardan, K. G. (1973a). *A Characterization of Multivariate Hypergeometric and Inverse Hypergeometric Models*, Mathematical Systems Program Report, Sangamon, IL: Sangamon State University. [6.8]

Janardan, K. G. (1973b). On an alternative expression for the hypergeometric moment generating function, *American Statistician*, **27**, 242. [6.3]

Janardan, K. G. (1975). Markov-Polya urn model with predetermined strategies, *Gujurat Statistical Review*, **2**, 17–32. [2.6]

Janardan, K. G. (1978). On generalized Markov-Polya distribution, *Gujurat Statistical Review*, **5**, 16–32. [6.11.4]

Janardan, K. G. (1984). Moments of certain series distributions and their applications, *SIAM Journal of Applied Mathematics*, **44**, 854–868. [7.2.6]

Janardan, K. G. (1987). Weighted Lagrange distributions and their characterizations, *SIAM Journal on Applied Mathematics*, **47**, 411–415. [7.2.5]

Janardan, K. G. (1988). Relationship between Morisita's model for estimating the environmental density and the generalized Eulerian numbers, *Annals of the Institute of Statistical Mathematics, Tokyo*, **40**, 439–450. [2.6]

Janardan, K. G. (1993). Some properties of the generalized Eulerian distribution, *Journal of Statistical Planning and Inference*, **34**, 159–169. [2.6]

Janardan, K. G. (1997). A wider class of Lagrange distributions of the second kind, *Communications in Statistics—Theory and Methods*, **26**, 2087–2091. [2.5, 7.2, 7.2.9]

Janardan, K. G. (1998). Generalized Polya Eggenberger family of distributions and its relation to Lagrangian Katz family, *Communications in Statistics—Theory and Methods*, **27**, 2423–2443. [7.2.4]

Janardan, K. G. (1999). Estimation of parameters of the GPED, *Communications in Statistics—Theory and Methods*, **28**, 2167–2179. [7.2.4]

Janardan, K. G. (2001). Comment: "On Lagrangian distributions of the second kind," *Communications in Statistics—Theory and Methods*, **30**, 2721–2722. [7.2.9]

Janardan, K. G., Mohan, N. R., Vasudeva, R., and Hebbar, H. V. (1997). A stochastic model with random change point for the study of oviposition tactics of weevils on beans, *Journal of Statistical Planning and Inference*, **61**, 193–201. [4.12.5]

Janardan, K. G., and Patil, G. P. (1970). Location of modes for certain univariate and multivariate distributions, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 57–76. University Park: Pennsylvania State University Press. [2.3.3]

Janardan, K. G., and Patil, G. P. (1972). A unified approach for a class of multivariate hypergeometric models, *Sankhyā, Series A*, **34**, 363–376. [6.3]

Janardan, K. G., and Rao, B. R. (1982). Characterization of generalized Markov-Polya and generalized Polya-Eggenberger distributions, *Communications in Statistics*, **11**, 2113–2124. [4.8]

Janardan, K. G., and Rao, B. R. (1983). Lagrange distributions of the second kind and weighted distributions, *SIAM Journal of Applied Mathematics*, **43**, 302–313. [2.5, 7.2, 7.2.9]

Janardan, K. G., and Schaeffer, D. J. (1977). A generalization of Markov-Polya distribution its extensions and applications, *Biometrical Journal*, **19**, 87–106. [11.2.19]

Janardan, K. G., and Schaeffer, D. J. (1981). Application of discrete distributions for estimating the number of organic compounds in water, *Statistical Distributions in Scientific Work*, Vol. 6: *Applications in Physical, Social, and Life Sciences*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 79–94. Dordrecht: Reidel. [5.10]

Jani, P. N. (1977). Minimum variance unbiased estimation for some left-truncated modified power series distributions, *Sankhyā, Series B*, **39**, 258–278. [2.2.2, 7.2.8]

Jani, P. N. (1978a). New numbers appearing in minimum variance unbiased estimation for decapitated negative binomial and Poisson distributions, *Journal of the Indian Statistical Association*, **16**, 41–48. [2.2.2]

Jani, P. N. (1978b). On modified power series distributions, *Metron*, **36**, 173–186. [2.2.2]

Jani, P. N. (1986). The generalized logarithmic series distribution with zeroes, *Journal of the Indian Society for Agricultural Statistics*, **38**, 345–351. [7.2.8]

Jani, P. N., and Shah, S. M. (1979a). Integral expressions for the tail probabilities of the modified power series distributions, *Metron*, **37**, 75–79. [2.2.2]

Jani, P. N., and Shah, S. M. (1979b). Misclassifications in modified power series distribution in which the value one is sometimes reported as zero, and some of its applications. *Metron*, **37**, 121–136. [2.2.2]

Janko, J. (editor) (1958). *Statistical Tables*, (Republished by Soviet Central Statistical Bureau, 1961): Prague. [4.6.1]

Jánossy, L., Rényi, A., and Aczél, J. (1950). On composed Poisson distributions, I, *Acta Mathematica Academiae Scientiarum Hungaricae*, **1**, 209–224. [9.3]

Jeffreys, H. (1941). Some applications of the method of minimum chi-squared, *Annals of Eugenics, London*, **11**, 108–114. [5.1]

Jensen, A. (1948). An elucidation of Erlang's statistical works through the theory of stochastic processes, *The Life and Works of A. K. Erlang*. Copenhagen. [3.12.2]

Jing, S.-C. (1994). The q-deformed binomial distribution and its asymptotic behaviour, *Journal of Physics A: Mathematics and General*, **27**, 493–499. [10.8.1]

Jing, S.-C., and Fan, H. Y. (1994). q-Deformed state, *Physics Review A*, **49**, 2277–2279. [10.8.1]

John, S. (1970). On analysing mixed samples, *Journal of the American Statistical Association*, **65**, 755–762. [8.2.7]

Johnson, N. L. (1951). Estimators of the probability of the zero class in Poisson and certain related populations, *Annals of Mathematical Statistics*, **22**, 94–101. [4.7.2]

Johnson, N. L. (1957). A note on the mean deviation of the binomial distribution, *Biometrika*, **44**, 532–533. [3.3]

Johnson, N. L. (1959). On an extension of the connexion between Poisson and χ^2 distributions, *Biometrika*, **46**, 352–362. [4.12.3]

Johnson, N. L., and Kotz, S. (1969). *Discrete Distributions* (first edition), Boston: Houghton Mifflin. [10.4.1]

Johnson, N. L., and Kotz, S. (1977). *Urn Models and Their Application*, New York: Wiley. [3.10, 4.2.1, 5.1, 10.4.1, 10.4.2, 11.2.19]

Johnson, N. L., and Kotz, S. (1981). Moments of discrete probability distributions derived using finite difference operators, *American Statistician*, **35**, 268. [1.2.7]

Johnson, N. L., and Kotz, S. (1982). Developments in discrete distributions, 1969–1980, *International Statistical Review*, **50**, 71–101. [9.2]

Johnson, N. L., and Kotz, S. (1985). Some distributions arising as a consequence of errors in inspection, *Naval Research Logistics Quarterly*, **32**, 35–43. [6.11.4]

Johnson, N. L., and Kotz, S. (1989). Characterization based on conditional distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **41**, 13–17. [6.10.4]

Johnson, N. L., and Kotz, S. (1990a). Use of moments in deriving distributions and some characterizations, *Mathematical Scientist*, **15**, 42–52. [1.2.7]

Johnson, N. L., and Kotz, S. (1990b). Randomly weighted averages: Some aspects and extensions, *American Statistician*, **44**, 245–249. [1.2.7]

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, New York: Wiley. [1.2.15, 4.2.1, 11.1.4, 11.2.8]

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2, New York: Wiley. [3.6.1, 3.8.3, 3.11]

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, New York: Wiley. [2.8]

Johnson, N. L., Kotz, S., and Rodriguez, R. N. (1985). Statistical effects of imperfect inspection sampling: I Some basic distributions, *Journal of Quality Technology*, **17**, 1–31. [6.11.4]

Johnson, N. L., Kotz, S., and Rodriguez, R. N. (1986). Statistical effects of imperfect inspection sampling: II. Double sampling and link sampling, *Journal of Quality Technology*, **18**, 198–199. [6.11.4]

Johnson, N. L., Kotz, S., and Sorkin, H. L. (1980). "Faulty inspection" distributions, *Communications in Statistics—Theory and Methods*, **A9**, 917–922. [6.11.4]

Johnson, N. L., Kotz, S., and Wu, X. (1991). *Inspection Errors for Attributes in Quality Control*, London: Chapman & Hall. [6.11.4]

Johnson, N. L., and Young, D. H. (1960). Some applications of two approximations to the multinomial distribution, *Biometrika*, **47**, 463–469. [4.5, 4.7.1]

Jones, C. M., and Zhigljavsky A. A. (2004). Approximating the negative moments of the Poisson distribution, *Statistics and Probability Letters*, **66**, 171–181. [4.3]

Jones, G., and Lai, C. D. (2002). Some parametric chain models, *Mathematical Scientist*, **27**, 117–125. [5.12.9]

Jones, M. C. (1987). Inverse factorial moments, *Statistics and Probability Letters*, **6**, 37–42. [3.3]

Jordan, C. (1927). Sur un cas généralisé de la probabilité des épreuves répétées, *Comptes Rendus, Académie des Sciences, Paris*, **184**, 315–317. [6.2.4]

Jordan, C. (1950). *Calculus of Finite Differences* (second edition), New York: Chelsea. [1.1.3, 9.1]

Jordan, K. (1972). *Chapters on the Classical Calculus of Probability*, Budapest: Akadémiai Kiadó (Hungarian edition, 1956). [10.2]

Jordan, M. C. (1867). De quelques formules de probabilité, *Comptes Rendus, Académie des Sciences, Paris*, **65**, 993–994. [10.2]

Jørgensen, B. (1997). *The Theory of Regression Models*, London: Chapman & Hall/CRC. [11.1.1, 11.1.2, 11.1.9]

Joseph, A. W., and Bizley, M. T. L. (1960). The two-pack matching problem, *Journal of the Royal Statistical Society, Series B*, **22**, 114–130. [10.3]

Joshi, S. W. (1974). Integral expressions for the tail probabilities of the power series distributions, *Sankhyā, Series B*, **36**, 462–465. [2.2.1, 3.4]

Joshi, S. W. (1975). Some recent advances with power series distributions, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 9–17. Dordrecht: Reidel. [2.2.1, 3.4]

Jowett, G. H. (1963). The relationship between the binomial and *F* distributions, *Statistician*, **13**, 55–57. [3.6.1]

Kaas, R., and Buhrman, J. M. (1980). Mean, median and mode in binomial distributions, *Statistica Neerlandica*, **34**, 13–18. [3.4]

Kabaila, P., and Byrne, J. (2001). Exact short Poisson confidence intervals, *Canadian Journal of Statistics*, **29**, 99–106. [4.7.3]

Kabe, D. G. (1969). Some distribution problems of order statistics from discrete populations, *Annals of the Institute of Statistical Mathematics, Tokyo*, **21**, 551–556. [5.2]

Kachitvichyanukul, V., and Schmeiser, B. W. (1985). Computer generation of hypergeometric random variates, *Journal of Statistical Computation and Simulation*, **22**, 127–145. [6.6]

Kachitvichyanukul, V., and Schmeiser, B. W. (1988). Binomial random variate generation, *Communications of the ACM*, **31**, 216–222. [3.7.2]

Kagan, A. (2001). A discrete version of the Stam inequality and a characterization of the Poisson distribution, *Journal of Statistical Planning and Inference*, **92**, 7–12. [4.8]

Kagan, A. M., Linnik, Y. V., and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, New York: Wiley. [3.9, 5.9.1, 5.9.2, 6.8]

Kahn, W. D. (1987). A cautionary note for Bayesian estimation of the binomial parameter *n*, *American Statistician*, **41**, 38–39. [3.8.2]

Kalbfleisch, J. D. (1986). Pseudo-likelihood, *Encyclopedia of Statistical Sciences*, Vol. 7, S. Kotz, N. L. Johnson, and C. B. Read (editors), 324–327. New York: Wiley. [1.2.16]

Kalbfleisch, J. D., and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, New York: Wiley. [1.2.4, 11.2.15]

Kamat, A. R. (1965). Incomplete and absolute moments of some discrete distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 45–64. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [2.3.1, 2.3.3, 5.4, 6.3, 7.1.3]

Kamat, A. R. (1966). A generalization of Johnson's property of the mean deviation for a class of distributions, *Biometrika*, **53**, 285–287. [2.3.3]

Kambo, N. S., and Kotz, S. (1966). On exponential bounds for binomial probabilities, *Annals of the Institute of Statistical Mathematics, Tokyo*, **18**, 277–287. [3.6.2]

Kapadia, C. H., and Thomasson, R. L. (1975). On estimating the parameter of a truncated geometric distribution by the method of moments, *Annals of the Institute of Statistical Mathematics, Tokyo*, **27**, 269–272. [5.2]

Kaplan, N., and Risko, K. (1982). A method for estimating rates of nucleotide substitution using DNA sequence data, *Theoretical Population Biology*, **21**, 318–328. [3.10]

Kaplansky, I., and Riordan, J. (1945). Multiple matching and runs by the symbolic method, *Annals of Mathematical Statistics*, **16**, 272–277. [10.3]

Kapur, J. N. (1978a). On generalized birth and death processes and generalized hypergeometric functions, *Indian Journal of Mathematics*, **20**, 57–69. [2.4.1]

Kapur, J. N. (1978b). Application of generalized hypergeometric functions to generalized birth and death processes, *Indian Journal of Pure and Applied Mathematics*, **9**, 1059–1069. [2.4.1]

Kapur, J. N. (1989). *Maximum Entropy Models in Science and Engineering*, New Delhi: Wiley Eastern. [10.8.3, 11.2.20]

Karlin, S., and McGregor, J. (1965). Ehrenfest urn models, *Journal of Applied Probability*, **2**, 352–376. [3.12.2]

Karlin, S., and Taylor, H. M. (1975). *A First Course in Stochastic Processes* (second edition), New York: Academic Press. [11.2.7]

Karlis, D., and Xekalaki, E. (1998). Minimum Hellinger distance estimation for Poisson mixtures, *Computational Statistics and Data Analysis*, **29**, 81–103. [8.2.5]

Karlis, D., and Xekalaki, E. (1999). On testing for the number of components in finite Poisson mixtures, *Annals of the Institute of Statistical Mathematics, Tokyo*, **51**, 149–161. [8.2.5]

Karlis, D., and Xekalaki, E. (2001). Robust inference for finite Poisson mixtures, *Journal of Statistical Planning and Inference*, **93**, 93–115. [8.2.5]

Kathman, S. J., and Terrell, G. R. (2002). Improving Poisson approximations through tilting and expansion, *Journal of Statistical Planning and Inference*, **104**, 315–327. [4.2.2]

Kathman, S. J., and Terrell, G. R. (2003). Poisson approximation by constrained exponential tilting, *Statistics and Probability Letters*, **61**, 83–89. [4.2.2]

Katsis, A. (2001). Calculating the optimal sample size for binomial populations, *Communications in Statistics—Theory and Methods*, **30**, 665–678. [3.8.4]

Katsis, A., and Toman, B. (1999). Bayesian sample size calculations for binomial experiments, *Journal of Statistical Planning and Inference*, **81**, 349–362. [3.8.4]

Katti, S. K. (1960). Moments of the absolute difference and absolute deviation of discrete distributions, *Annals of Mathematical Statistics*, **31**, 78–85. [2.3.3, 3.3, 4.12.3]

Katti, S. K. (1965). *Some Estimation Procedures For Discrete Distributions*, Statistical Report, Tallahassee, FL: Florida State University. [9.6.4]

Katti, S. K. (1966). Interrelations among generalized distributions and their components, *Biometrics*, **22**, 44–52. [2.4.2, 3.10, 5.3, 6.11.3, 8.3.3, 8.4, 9.1, 9.3, 9.13]

Katti, S. K. (1967). Infinite divisibility of integer valued random variables, *Annals of Mathematical Statistics*, **38**, 1306–1308. [7.1.4, 9.3]

Katti, S. K., and Gurland, J. (1961). The Poisson Pascal distribution, *Biometrics*, **17**, 527–538. [9.8, 11.1.2]

Katti, S. K., and Gurland, J. (1962a). Efficiency of certain methods of estimation for the negative binomial and the Neyman type A distributions, *Biometrika*, **49**, 215–226. [5.8.3, 9.6.4]

Katti, S. K., and Gurland, J. (1962b). Some methods of estimation for the Poisson binomial distribution, *Biometrics*, **18**, 42–51. [9.5]

Katti, S. K., and Rao, A. V. (1970). The log-zero-Poisson distribution, *Biometrics*, **26**, 801–813. [8.2.4, 8.3.3]

Katz, L. (1945). *Characteristics of Frequency Functions Defined by First Order Difference Equations*, PhD Thesis, Ann Arbor, MI: University of Michigan. [2.3.1, 3.4]

Katz, L. (1946). On the class of functions defined by the difference equation $(x+1)f(x+1) = (a+bx)f(x)$, *Annals of Mathematical Statistics*, **17**, 501 (abstract). [2.3.1]

Katz, L. (1948). Frequency functions defined by the Pearson difference equation, *Annals of Mathematical Statistics*, **19**, 120 (abstract). [2.3.1]

Katz, L. (1965). Unified treatment of a broad class of discrete probability distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 175–182. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [2.3.1]

Kay, R. (1985). Partial likelihood, *Encyclopedia of Statistical Sciences*, Vol. 6, S. Kotz, N. L. Johnson, and C. B. Read (editors), 591–593. New York: Wiley. [1.2.16]

Kemp, A. W. (1968a). *Studies in Univariate Discrete Distribution Theory Based on the Generalized Hypergeometric Function and Associated Differential Equations*, PhD Thesis, Belfast: The Queen's University of Belfast. [2.4.1, 2.4.2, 3.4, 3.10, 4.12.4, 5.3, 6.2.5, 8.3.3, 8.4]

Kemp, A. W. (1968b). A wide class of discrete distributions and the associated differential equations, *Sankhyā, Series A*, **30**, 401–410. [2.4.1, 2.4.2, 3.4, 3.10, 4.12.4, 6.2.5, 6.10.2, 11.2.10, 11.2.12]

Kemp, A. W. (1968c). A limited risk cPp, *Skandinavisk Aktuarietidskrift*, **51**, 198–203. [2.4.2, 8.3.3]

Kemp, A. W. (1974). *Towards a Unification of the Theory of Counts*, School of Mathematics, Statistics Reports and Preprints, 12, Bradford, U.K.: University of Bradford. [2.4.2]

Kemp, A. W. (1978a). Cluster size probabilities for generalized Poisson distributions, *Communications in Statistics—Theory and Methods*, **7**, 1433–1438. [7.1.2, 7.1.4, 9.3]

Kemp, A. W. (1978b). On probability generating functions for matching and occupancy distributions, *Zastosowania Matematyki*, **16**, 207–213. [2.4.2, 10.3, 10.4.1]

Kemp, A. W. (1979). Convolutions involving binomial pseudo-variables, *Sankhyā Series A*, **41**, 232–243. [3.12.2, 5.12.5, 7.1.6, 9.12, 11.2.7]

Kemp, A. W. (1981a). Conditionality properties for the bivariate logarithmic distribution with an application to goodness of fit, *Statistical Distributions in Scientific Work*, Vol. 5: *Inferential Problems and Properties*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 57–73. Dordrecht: Reidel. [7.1.2]

Kemp, A. W. (1981b). Efficient generation of logarithmically distributed pseudorandom variables, *Applied Statistics*, **30**, 249–253. [7.1.4, 7.1.6]

Kemp, A. W. (1986). Weighted discrepancies and maximum likelihood estimation for discrete distributions, *Communications in Statistics—Theory and Methods*, **15**, 783–803. [1.2.16, 5.8.3, 7.1.7, 8.2.4]

Kemp, A. W. (1987a). A Poissonian binomial model with constrained parameters, *Naval Research Logistics*, **34**, 853–858. [3.12.3, 10.8.1]

Kemp, A. W. (1987b). Families of discrete distributions satisfying Taylor's power law, *Biometrics*, **43**, 693–699. [11.2.3]

Kemp, A. W. (1988a). Families of distributions for repeated samples of animal counts: Response, *Biometrics*, **44**, 888–890. [11.2.3]

Kemp, A. W. (1988b). Simple algorithms for the Poisson modal cumulative probability, *Communications in Statistics—Simulation and Computation*, **17**, 1495–1508. [4.4, 4.5, 4.6.1, 4.6.2]

Kemp, A. W. (1989). A note on Stirling's expansion for factorial n, *Statistics and Probability Letters*, **7**, 139–143. [4.5, 4.6.1]

Kemp, A. W. (1992a). Heine-Euler extensions of the Poisson distribution, *Communications in Statistics—Theory and Methods*, **21**, 571–588. [4.12.6, 10.8.2]

Kemp, A. W. (1992b). Steady-state Markov chain models for the Heine and Euler distributions, *Journal of Applied Probability*, **29**, 869–876. [4.12.6, 10.8.2]

Kemp, A. W. (1992c). On counts of individuals able to signal the presence of an observer, *Biometrical Journal*, **34**, 595–604. [11.2.7]

Kemp, A. W. (1994). Comments on the Luria-Delbrück distribution, *Journal of Applied Probability*, **31**, 822–828. [11.2.11]

Kemp, A. W. (1995). Splitters, lumpers and species per genus, *Mathematical Scientist*, **20**, 107–118. [7.1.12, 11.2.20]

Kemp, A. W. (1997a). Characterizations of a discrete normal distribution, *Journal of Statistical Planning and Inference*, **63**, 223–229. [10.8.3]

Kemp, A. W. (1997b). On modified q-Bessel functions and some statistical applications, *Advances in Combinatorial Methods and Applications to Probability and Statistics*, N. Balakrishnan (editor), 465–470. Boston: Birkhäuser. [10.8.3]

Kemp, A. W. (1998). Absorption sampling and the absorption distribution, *Journal of Applied Probability*, **35**, 1–6. [10.8.1]

Kemp, A. W. (2001a). The q-beta-geometric distribution as a model for fecundability, *Communications in Statistics—Theory and Methods*, **30**, 2373–2384. [10.8.2]

Kemp, A. W. (2001b). A characterization of a distribution arising from absorption sampling, *Probability and Statistical Models with Applications*, Ch. A. Charalambides, M. V. Koutras, and N. Balakrishnan (editors), 239–246. Boca Raton, FL: Chapman & Hall/CRC. [10.8.1]

Kemp, A. W. (2002a). Certain q-analogues of the binomial distribution, *Sankhyā, Series A*, **109**, 293–305. [10.8.1]

Kemp, A. W. (2002b). Random walks and other aspects of the Bailey-Daum distribution, *Journal of Statistical Planning and Inference*, **101**, 163–168. [10.8.2]

Kemp, A. W. (2002c). Existence conditions and properties for the generalized Euler family of distributions, *Journal of Statistical Planning and Inference*, **101**, 169–178. [10.8.2]

Kemp, A. W. (2003). Characterizations involving $U|(U + V = m)$ for certain discrete distributions, *Journal of Statistical Planning and Inference*, **109**, 31–41. [10.8.1]

Kemp, A. W. (2004). Classes of discrete lifetime distributions, *Communications in Statistics—Theory and Methods*, **33**, 3069–3093. [1.2.4]

Kemp, A. W. (2005). Steady-state Markov chain models for certain q-confluent hypergeometric distributions, *Journal of Statistical Planning and Inference* (in press). [10.8.2]

Kemp, A. W., and Kemp, C. D. (1966). An alternative derivation of the Hermite distribution, *Biometrika*, **53**, 627–628. [9.4, 11.1.4]

Kemp, A. W., and Kemp, C. D. (1969a). Branching and clustering models associated with the "lost-games" distribution, *Journal of Applied Probability*, **6**, 700–703. [7.2.1, 11.2.10]

Kemp, A. W., and Kemp, C. D. (1969b). Some distributions arising from an inventory decision problem, *Bulletin of the International Statistical Institute*, **43**(2), 336–338. [2.4.2, 6.2.5, 11.2.13]

Kemp, A. W., and Kemp, C. D. (1971). On mixing processes and the lost-games distribution, *Zastosowania Matematyki*, **12**, 167–173. [11.2.10]

Kemp, A. W., and Kemp, C. D. (1974). A family of distributions defined via their factorial moments, *Communications in Statistics*, **3**, 1187–1196. [2.4.1, 2.4.2, 3.4]

Kemp, A. W., and Kemp, C. D. (1975). Models for Gaussian hypergeometric distributions, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 31–40. Dordrecht: Reidel. [6.2.5]

Kemp, A. W., and Kemp, C. D. (1986). Distributional properties of a model for the spread of drug abuse, *Communications in Statistics—Theory and Methods*, **15**, 3287–3298. [5.12.5]

Kemp, A. W., and Kemp, C. D. (1987). A rapid and efficient estimation procedure for the negative binomial distribution, *Biometrical Journal*, **29**, 865–873. [5.8.3]

Kemp, A. W., and Kemp, C. D. (1989). Even-point estimation, *Encyclopedia of Statistical Sciences Supplement*, S. Kotz, N. L. Johnson, and C. B. Read (editors). New York: Wiley. [9.4]

Kemp, A. W., and Kemp, C. D. (1990). A composition-search algorithm for low-parameter Poisson generation, *Journal of Statistical Computation and Simulation*, **35**, 239–244. [4.4, 4.6.2]

Kemp, A. W., and Kemp, C. D. (1991). Weldon's dice data revisited, *American Statistician*, **45**, 216–222. [3.12.3, 10.8.1]

Kemp, A. W., and Kemp, C. D. (1992). A group-dynamic model and the lost-games distribution, *Communications in Statistics—Theory and Methods*, **21**, 791–798. [11.2.10]

Kemp, A. W., and Kemp, C. D (2004). Factorial moment characterizations for certain binomial-type distributions, *Communications in Statistics—Theory and Methods*, **33**, 3059–3068. [10.4.3]

Kemp, A. W., and Newton, J. (1990). Certain state-dependent processes for dichotomised parasite populations, *Journal of Applied Probability*, **27**, 251–258. [2.4.1, 10.8.1]

Kemp, C. D. (1967a). "Stuttering-Poisson" distributions, *Journal of the Statistical and Social Enquiry Society of Ireland*, **21**(5), 151–157. [4.11, 5.1, 9.3, 9.7]

Kemp, C. D. (1967b). On a contagious distribution suggested for accident data, *Biometrics*, **23**, 241–255. [9.6.5, 9.9]

Kemp, C. D. (1970). "Accident proneness" and discrete distribution theory, *Random Counts in Scientific Work*, Vol. 2: *Random Counts in Biomedical and Social Sciences*, G. P. Patil (editor), 41–65. University Park: Pennsylvania State University Press. [5.10]

Kemp, C. D. (1986). A modal method for generating binomial variables, *Communications in Statistics—Theory and Methods*, **15**, 805–813. [3.6.1, 3.7.1, 3.7.2]

Kemp, C. D. (1997). A q-logarithmic distribution, *Advances in Combinatorial Methods and Applications to Probability and Statistics*, N. Balakrishnan (editor), 465–470. Boston: Birkhäuser. [7.1.12, 10.8.2]

Kemp, C. D. (2002). q-Analogues of the hyper-Poisson distribution, *Journal of Statistical Planning and Inference*, **101**, 179–183. [10.8.2]

Kemp, C. D. (2004). A q-generalization of the logarithmic distribution (submitted). [10.8.4]

Kemp, C. D., and Kemp, A. W. (1956a). Generalized hypergeometric distributions, *Journal of the Royal Statistical Society, Series B*, **18**, 202–211. [2.4.1, 6.2.2, 6.2.3, 6.2.5, 6.4, 8.4]

Kemp, C. D., and Kemp, A. W. (1956b). The analysis of point quadrat data, *Australian Journal of Botany*, **4**, 167–174. [6.7.2, 6.9.2]

Kemp, C. D., and Kemp, A. W. (1965). Some properties of the "Hermite" distribution, *Biometrika*, **52**, 381–394. [9.3, 9.4]

Kemp, C. D., and Kemp, A. W. (1967). A special case of Fisher's "modified Poisson series," *Sankhyā, Series A*, **29**, 103–104. [9.4]

Kemp, C. D., and Kemp, A. W. (1968). On a distribution associated with certain stochastic processes, *Journal of the Royal Statistical Society, Series B*, **30**, 401–410. [11.2.10]

Kemp, C. D., and Kemp, A. W. (1987). Rapid generation of frequency tables, *Applied Statistics*, **36**, 277–282. [1.2.17]

Kemp, C. D., and Kemp, A. W. (1988). Rapid estimation for discrete distributions, *Statistician*, **37**, 243–255. [4.10.1, 7.1.7, 8.2.4]

Kemp, C. D., and Kemp, A. W. (1991). Poisson random variate generation, *Applied Statistics*, **40**, 143–158. [4.6.1, 4.6.2]

Kemp, C. D., and Kemp, A. W. (1997). Aspects of the Skellam-Haldane distribution, *Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz*, N. L. Johnson and N. Balakrishnan (editors), 325–338. New York: Wiley. [11.2.16]

Kemp, C. D., Kemp, A. W., and Loukas, S. (1979). *Sampling from Discrete Distributions*, Statistics Reports and Preprints, 42, Bradford, U.K.: University of Bradford. [7.1.6]

Kempton, R. A. (1975). A generalized form of Fisher's logarithmic series, *Biometrika*, **62**, 29–38. [7.1.11]

Kendall, D. G. (1948). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, *Biometrika*, **35**, 6–15. [5.3, 7.1.2, 7.1.10, 9.2, 11.2.16]

Kendall, D. G. (1949). Stochastic processes and population growth, *Journal of the Royal Statistical Society, Series B*, **11**, 230–282. [5.3, 5.10]

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, Vol. 1, London: Griffin. [3.3, 4.3, 11.2.8]

Kendall, M. G. (1960). The bibliography of operational research, *Operational Research Quarterly*, **11**, 31–36. [6.10.3]

Kendall, M. G. (1961). Natural law in the social sciences, *Journal of the Royal Statistical Society, Series A*, **124**, 1–18. [6.10.3, 6.10.4, 11.2.20]

Kendall, M. G. (1968). Studies in the history of probability and statistics XVIII, Thomas Young on coincidences, *Biometrika*, **55**, 249–250. [10.3]

Kendall, M. G., and Stuart, A. (1961). *The Advanced Theory of Statistics*, Vol. 2, London: Griffin. [3.8.3]

Kendall, M. G., and Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, London: Griffin. [10.8.4]

Khalil, Z., Dimitrov, B., and Dion, J. P. (1991). A characterization of the geometric distribution related to random sums, *Stochastic Models*, **7**, 321–326. [5.9.1]

Khamis, S. H. (1960). Incomplete gamma function expansions of statistical distribution functions, *Bulletin of the International Statistical Institute*, **37**, 385–396. [1.1.5]

Khamis, S. H., and Rudert, W. (1965). *Tables of the Incomplete Gamma Function Ratio: Chi-Square Integral, Poisson Distribution*, Darmstadt: von Liebig. [1.1.5, 4.6.1]

Khan, A. H., and Ali, M. M. (1987). Characterizations of probability distributions through higher order gap, *Communications in Statistics—Theory and Methods*, **16**, 1281–1287. [1.2.12]

Kharshikar, A. V. (1970). On the expected value of $S^2/\bar{x}$, *Biometrics*, **26**, 343–346. [4.8]

Khatri, C. G. (1959). On certain properties of power-series distributions, *Biometrika*, **46**, 486–490. [2.2.1, 7.1.3]

Khatri, C. G. (1961). On the distributions obtained by varying the number of trials in a binomial distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **13**, 47–51. [7.1.10, 8.2.4]

Khatri, C. G. (1962). Distributions of order statistics for the discrete case, *Annals of the Institute of Statistical Mathematics, Tokyo*, **14**, 167–171. [3.5, 9.13]

Khatri, C. G. (1983). Some remarks on the moments of discrete distributions using difference operators, *American Statistician*, **37**, 96–97. [1.2.7]

Khatri, C. G., and Patel, I. R. (1961). Three classes of univariate discrete distributions, *Biometrics*, **17**, 567–575. [3.10, 8.2.4, 9.3, 9.13]

King, G. (1902). *Institute of Actuaries' Text Book of the Principles of Interest, Life Annuities, and Assurances, and Their practical Application*, Vol 2: *Life Contingencies* (second edition), London: Leyton. [10.2]

Kirkman, T. P. (1857). On the k partitions of the r-gon and r-ace, *Philosophical Transactions of the Royal Society of London*, **147**, 217–272. [6.10.5]

Kitagawa, T. (1952). *Tables of Poisson Distribution*, Tokyo: Baifukan. [4.6.1]

Klar, B. (1999). Model selection for a family of discrete failure time distributions, *Australian and New Zealand Journal of Statistics*, **41**, 337–352. [11.2.15]

Kleinrock, L. (1975). *Queueing Systems*, Vol. 1, New York: Wiley. [11.2.14]

Kleinrock, L. (1976). *Queueing Systems*, Vol. 2, New York: Wiley. [11.2.14]

Klotz, J. (1970). The geometric density with unknown location parameter, *Annals of Mathematical Statistics*, **41**, 1078–1082. [5.2]

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models: From Data to Decisions*, New York: Wiley. [2.3.2, 5.8.5, 5.12.2]

Knessel, Ch. (1991). On the transient behavior of the repairman problem, *Advances in Applied Probability*, **23**, 327–354. [6.11.2]

Knopp, K. (1951). *Theory and Application of Infinite Series*, London: Blackie & Son. [11.2.20]

Koch, G. G., Atkinson, S. S., and Stokes, M. E. (1986). Poisson regression, *Encyclopedia of Statistical Sciences*, Vol. 7, S. Kotz, N. L. Johnson, and C. B. Read (editors), 32–41. New York: Wiley. [4.9]

Kocherlakota, K., and Kocherlakota, S. (1993). A statistical problem arising in ascertainment, *Journal of Applied Statistics*, **20**, 363–374. [3.12.4]

Kocherlakota, S., and Kocherlakota, K. (1986). Goodness of fit tests for discrete distributions, *Communications in Statistics—Theory and Methods*, **15**, 815–829. [1.2.16]

Kocherlakota, S., and Kocherlakota, K. (1990). Tests of hypothesis for the weighted binomial distribution, *Biometrics*, **46**, 645–656. [3.12.4]

Kokonendji, C. C. (1999). Anscombe's problem for generalized negative-binomial distribution, *Canadian Journal of Statistics*, **27**, 199–205 (in French). [5.8.5]

Kokonendji, C. C., and Khoudar, M. (2004). On strict arcsine distribution, *Communications in Statistics—Theory and Methods*, **33**, 993–1006. [5.12.9]

Kolchin, V. F., Sevast'yanov, B. A., and Chistyakov, V. P. (1978). *Random Allocations*, New York: Wiley. [10.4.1]

Korwar, R. M. (1975). On characterizing some discrete distributions by linear regression, *Communications in Statistics*, **4**, 1133–1147. [1.2.9, 4.8]

Kosambi, D. D. (1949). Characteristic properties of series distributions, *Proceedings of the National Institute for Science, India*, **15**, 109–113. [2.2.1, 2.2.2, 3.4, 4.8]

Kotz, S. (1974). Characterizations of statistical distributions: A supplement to recent surveys, *International Statistical Review*, **42**, 39–65. [4.8]

Kotz, S., and Johnson, N. L. (1990). Regression relations in random sum theory, *Proceedings of the 1990 International Statistical Symposium, Taipei.* [1.2.9]

Kotz, S., and Johnson, N. L. (1991). A note on renewal (partial sums) distributions for discrete variables, *Statistics and Probability Letters*, **12**, 229–231. [11.2.13]

Kourouklis, S. (1986). Characterizations of some discrete distributions based on a variant of the Rao-Rubin condition. *Communications in Statistics—Theory and Methods*, **15**, 839–851. [4.8]

Koutras, M. V. (1997a). Waiting time distributions associated with runs of fixed length in two-state Markov chains, *Annals of the Institute of Statistical Mathematics, Tokyo*, **49**, 123–139. [10.7.2, 10.7.3]

Koutras, M. V. (1997b). Waiting times and number of appearances of events in a sequence of discrete random variables, *Advances in Combinatorial Methods and Applications to Probability and Statistics*, N. Balakrishnan (editor), 465–470. Boston: Birkhäuser. [10.7.5]

Krafft, O. (1969). A note on exponential bounds for binomial probabilities, *Annals of the Institute of Statistical Mathematics, Tokyo*, **21**, 219–220. [3.6.2]

Kreweras, G. (1979). Some finite distributions tending towards Poisson distributions, *Bulletin of the International Statistical Institute*, **49**, 296–299. [4.2.2]

Krishnaji, N. (1974). Characterization of some discrete distributions based on a damage model, *Sankhyā, Series A*, **36**, 204–213. [4.8]

Krishnamoorthy, K., and Thomson J. (2003). A more powerful test for comparing two Poisson means, *Journal of Statistical Planning and Inference*, **119**, 23–35. [4.7.2]

Kronmal, R. A., and Peterson, A. V. (1979). On the alias method for generating random variables from a discrete distribution, *American Statistician*, **33**, 214–218. [1.2.17, 6.6]

Kulasekera, K. B., and Tonkyn, D. W. (1992). A new distribution, with applications to survival, dispersal and dispersion, *Communications in Statistics—Simulation and Computation*, **21**, 499–518. [7.1.12, 11.2.20]

Kulkarni, P. M., and Shah, A. K. (1995). Testing the equality of several binomial proportions to a prespecified standard, *Statistics and Probability Letters*, **25**, 213–219. [3.8.4]

Kumar, A., and Consul, P. C. (1979). Negative moments of a modified power series distribution and bias of the maximum likelihood estimator, *Communications in Statistics—Theory and Methods*, **A8**, 151–166. [2.2.2, 5.12.3, 7.2.7]

Kumar, A., and Consul, P. C. (1980). Minimum variance unbiased estimation for modified power series distribution, *Communications in Statistics—Theory and Methods*, **A9**, 1261–1275. [2.2.2, 5.12.3, 7.2.6, 7.2.7]

Kumar, C. S. (2002). Extended generalized hypergeometric probability distributions, *Statistics and Probability Letters*, **59**, 1–7. [6.11.4]

Kupper, L. L., and Haseman, J. K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments, *Biometrics*, **34**, 69–76. [3.12.6]

Kyriakoussis, A., Li, G., and Papadopoulos, A. (1998). On characterization and goodness-of-fit test of some discrete distribution families, *Journal of Statistical Planning and Inference*, **74**, 215–228. [4.7.4]

Kyriakoussis, A., and Papadopoulos, A. S. (1993). On the Bayes estimators of the probability of “success” and reliability function of the zero-truncated binomial and negative binomial distributions, *Sankhyā: Indian Journal of Statistics, Series B*, **55**, 171–185. [3.8.2]

Kyriakoussis, A., and Papageorgiou, H. (1991a). Characterization of hypergeometric type distributions by regression, *Statistics*, **22**, 467–477. [6.8]

Kyriakoussis, A., and Papageorgiou, H. (1991b). Characterizations of logarithmic series distributions, *Statistica Neerlandica*, **45**, 1–8. [7.1.8]

Laha, R. G. (1982). Characteristic functions, *Encyclopedia of Statistical Sciences*, Vol. 1, S. Kotz, N. L. Johnson, and C. B. Read (editors), 415–422. New York: Wiley. [1.2.10]

Lai, C. D., and Wang, D. Q. (1995). A finite range discrete life distribution, *International Journal of Reliability, Quality and Safety Engineering*, Vol. 2, 147–160. [11.2.15]

Lai, C. D., Xie, M., and Murthy, D. N. P. (2001). Bathtub-shaped failure rate life distributions, *Handbook of Statistics*, Vol. 20, N. Balakrishnan and C. R. Rao (editors), 69–104. Amsterdam: Elsevier. [11.2.15]

Lam, H. S., and Lampard, D. G. (1981). Modelling of drug receptor interaction with birth and death processes, *Journal of Mathematical Biology*, **12**, 153–172. [3.12.2]

Laplace, P. S. (1820). *Théorie Analytique des Probabilités* (third edition), Paris: Courvier. [3.6.1]

Larson, H. R. (1966). A nomograph of the cumulative binomial distribution, *Industrial Quality Control*, **23**, 270–278. [3.7.1]

Laubscher, N. F. (1961). On stabilizing the binomial and negative binomial variances, *Journal of the American Statistical Association*, **56**, 143–150. [5.6]

Laurent, A. G. (1965). Probability distributions, factorial moments, empty cell test, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 437–442. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [1.2.11]

Lawler, G. F. (1995). *Introduction to Stochastic Processes*, New York: Chapman & Hall. [11.2.13]

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, New York: Wiley. [1.2.4, 11.2.15]

Lea, D. E., and Coulson, C. A. (1949). The distribution of the numbers of mutants in bacterial populations, *Journal of Genetics*, **49**, 264–285. [11.2.11]

L’Ecuyer, P., (1990). Random numbers for simulation, *Communications of the ACM*, **33**, 85–97. [1.2.17]

Lee, J. J., Serachitopol, D. M., and Brown, B. W. (1997). Likelihood-weighted confidence intervals for the difference of two binomial proportions, *Biometrical Journal*, **39**, 387–407. [3.8.3]

Lee, L. F. (1986). Specification test for Poisson regression models, *International Economic Review*, **27**, 689–706. [4.9]

Lee, M.-L. T., and Whitmore, G. A. (1993). Stochastic processes directed by randomized time, *Journal of Applied Probability*, **30**, 302–314. [5.12.2]

Lee, P. A. (1967). The Ehrenfest urn model and a machine maintenance problem, *Proceedings IEEE*, **55**, 455–456. [3.12.2]

Leemis, J. F. (1995). *Reliability: Probabilistic Models and Statistical Methods*, Englewood Cliffs, NJ: Prentice-Hall. [1.2.4, 11.2.15]

Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (second edition), New York: Wiley. [1.2.16, 2.2.1]

León, C. A., and Perron, F. (2003). Extremal properties of sums of Bernoulli random variables, *Statistics and Probability Letters*, **62**, 345–354. [3.12.3]

Lepage, Y. (1978). Negative factorial moments of positive random variables, *Industrial Mathematics*, **28**, 95–100. [3.3]

Lessing, R. (1973). An alternative expression for the hypergeometric moment generating function, *American Statistician*, **27**, 115. [6.3]

Letac, G. (1991). Contre exemple au théorème de P C Consul sur la factorisation des distributions de Poisson généralisées, *Canadian Journal of Statistics*, **19**, 229–231. [7.2.6]

Levene, H. (1949). On a matching problem arising in genetics, *Annals of Mathematical Statistics*, **20**, 91–94. [10.3]

Levin, B., and Reeds, J. (1977). Compound multinomial likelihood functions are unimodal: Proof of a conjecture of I. J. Good, *Annals of Statistics*, **5**, 79–87. [5.8.5]

Lévy, P. (1937a). Sur les exponentielles de polynomes et sur l'arithmétique des produits de lois de Poisson, *Annales Scientifique de l'Ecole Normale Supérieure, Series III*, **54**, 231–292. [9.3]

Lévy, P. (1937b). *Théorie de l'Addition des Variables Aléatoires*, Paris: Gauthier-Villars. [4.2.1]

Lévy, P. (1954). *Théorie de l'Addition des Variables Aléatoires* (second edition), Paris: Gauthier-Villars. [3.6.2]

Lewis, P. A., and Orav, E. J. (1989). *Simulation Methodology for Statisticians, Operations Analysts and Engineers*, Vol. 1, Pacific Grove, CA: Wadsworth & Brooks/Cole. [1.2.17]

Lexis, W. (1877). *Zur Theorie der Massenerscheiungen in der Menschlichen Gesellschaft*, Freiburg: Wagner. [3.12.3]

Liang, T.-C. (1999). Monotone empirical Bayes tests for a discrete normal distribution, *Statistics and Probability Letters*, **44**, 241–249. [10.8.3]

Lidstone, G. J. (1942). Notes on the Poisson frequency distribution, *Journal of the Institute of Actuaries*, **71**, 284–291. [4.4]

Lieberman, G. J., and Owen, D. B. (1961). *Tables of the Hypergeometric Probability Distribution*, Stanford: Stanford University Press. [6.4, 6.5, 6.6]

Liebermeister, C. (1877). Ueber Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik, *Sammlung Klinischer Vorträge* (Innere Medicin No. 31–64), **110**, 935–962. [6.11.1]

Lin, G. D. (1987). Characterizations of distributions via relationships between two moments of order statistics, *Journal of Statistical Planning and Inference*, **19**, 73–80. [1.2.12]

Lin, G. D., and Hu, C.-Y. (2001). The Riemann zeta distribution, *Bernoulli*, **7**, 817–828. [11.2.20]

Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B*, **20**, 102–107. [8.3.3]

Lindley, D. V. (1990). The present position in Bayesian statistics, *Statistical Science*, **5**, 44–65. [1.2.2]

Lindsay, B. (1995). *Mixture Models: Theory, Geometry and Applications*, Alexandria, VA: Institute of Mathematical Statistics and the American Statistical Association. [8.1.2]

Lindsey, J. K. (1995). *Modelling Frequency and Count Data*, Oxford: Oxford University Press. [6.9.2, 11.1.8]

Lindsey, J. K. (1996). *Parametric Statistical Inference*, Oxford: Oxford University Press. [11.1.1]

Ling, K. D. (1988). On binomial distributions of order k, *Statistics and Probability Letters*, **6**, 371–376. [10.7.3, 10.7.5]

Ling, K. D. (1989). A new class of negative binomial distributions of order k, *Statistics and Probability Letters*, **7**, 247–250. [10.7.3]

Ling, K. D. (1990). On geometric distributions of order $(k_1, \ldots, k_m)$, *Statistics and Probability Letters*, **9**, 163–171. [2.8, 10.7.6]

Ling, R. F., and Pratt, J. W. (1984). The accuracy of Peizer approximations to the hypergeometric distribution with comparisons to some other approximations, *Journal of the American Statistical Association*, **79**, 49–60. [6.5]

Lingappaiah, G. S. (1987). Some variants of the binomial distribution, *Bulletin of the Malaysian Mathematical Society*, **10**, 82–94. [5.12.7, 7.2.5]

Lingappaiah, G. S. (1992). Distribution of the sum of independent right truncated negative binomial variates, *Revista de Matematica e Estatadistica*, **10**, 155–163. [5.11]

Lisman, J. H. C., and van Zuylen, M. C. A. (1972). Note on the generation of most probable frequency distributions, *Statistica Neerlandica*, **26**, 19–23. [10.8.3, 11.2.20]

Littlewood, J. E. (1969). On the probability in the tail of a binomial distribution, *Advances in Applied Probability*, **1**, 43–72. [3.6.1]

Lloyd, E. L. (1980). *Handbook of Applicable Mathematics*, Vol. 2: *Probability*, Chichester: Wiley. [4.11]

Lloyd, E. L. (editor) (1984). *Handbook of Applicable Mathematics*, Vol. 6A: *Statistics*, Chichester: Wiley. [3.8.4]

Lo, H.-P., and Wani, J. K. (1983). Maximum likelihood estimation of the parameters of the invariant abundance distributions, *Biometrics*, **39**, 977–986. [7.1.1, 7.1.7]

Loève, M. (1963). *Probability theory* (third edition), New York: van Nostrand. [10.2, 10.4.2]

Logunov, P. L. (1991). Estimates for the convergence rate to the Poisson distribution for random sums of independent indicators, *Theory of Probability and Its Application*, **35**, 587–590. [4.2.2]

Longini, I. M. (1998). Chain binomial model, *Encyclopedia of Biostatistics*, Vol. 1, 593–597. Chichester: Wiley. [3.12.5]

López-Blázquez, F., and Salamanca Miño, B. (2000). Binomial approximation to hypergeometric probabilities, *Journal of Statistical Planning and Inference*, **87**, 21–29. [6.5]

López-Blázquez, F., and Salamanca-Miño, F. (2001). Exact and approximated relations between negative hypergeometric and negative binomial probabilities, *Communications in Statistics—Theory and Methods*, **30**, 957–967. [5.6]

Lotka, A. J. (1926). The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, **16**, 317–323. [11.2.20]

Lotka, A. J. (1939). *Théorie Analytique des Associations Biologiques, Vol. 2: Analyse Démographique avec Applications Particulière à l'Espèce Humaine*, *Actualitiés Scientifiques et Industrielles*, Paris: Hermann. [3.12.2]

Louis, T. A. (1981). Confidence intervals for a binomial parameter after observing no successes, *American Statistician*, **35**, 154. [3.8.3]

Louv, W. C., and Littell, R. C. (1986). Combining one-sided binomial tests, *Journal of the American Statistical Association*, **81**, 550–554. [3.8.4]

Lovison, G. (1998). An alternative representation of Altham's multiplicative-binomial distribution, *Statistics and Probability Letters*, **36**, 415–420. [3.12.6]

Lu, W.-S. (2000). Improved confidence intervals for a binomial parameter using the Bayesian method, *Communications in Statistics—Theory and Methods*, **29**, 2835–2847. [3.8.3]

Lüders, R. (1934). Die Statistik der seltenen Ereignisse, *Biometrika*, **26**, 108–128. [5.3, 5.12.5, 7.1.2, 9.3]

Lukacs, E. (1956). Characterization of populations by properties of suitable statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, 195–214. Berkeley: University of California Press. [4.8]

Lukacs, E. (1965). Characterization problems for discrete distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 65–74. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [3.9, 4.8, 5.9.1]

Lukacs, E. (1970). *Characteristic Functions* (second edition), London: Griffin. [1.2.10, 9.3]

Lukacs, E. (1983). *Developments in Characteristic Function Theory*, London: Griffin. [1.2.10]

Luke, Y. L. (1975). *Mathematical Functions and Their Approximations*, New York: Academic Press. [1.1.6]

Lund, R. E. (1980). Algorithm AS 152: Cumulative hypergeometric probabilities, *Applied Statistics*, **29**, 221–223. [6.6]

Lundberg, O. (1940). *On Random Processes and Their Application to Sickness and Accident Statistics*, Uppsala: Almqvist and Wicksells (reprinted 1964). [5.3, 8.3.3]

Luria, S. E., and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance, *Genetics*, **28**, 491–511. [11.2.11]

Lutsenko, M. M., and Maloshevskii, S. G. (2003). Minimax confidence intervals for the binomial parameter, *Journal of Statistical Planning and Inference*, **113**, 67–77. [3.8.3]

Lwin, T. (1981). A modified power series distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **33**, 361–374. [7.1.10]

Ma, W. T., Sandri, G. vH., and Sarkar, S. (1991). Novel representation of power series that arise in statistical mechanics and population genetics, *Physics Letters*, **A155**, 103–106. [11.2.11]

Ma, W. T., Sandri, G. vH., and Sarkar, S. (1992). Analysis of the Luria–Delbrück distribution using discrete convolution powers, *Journal of Applied Probability*, **29**, 255–267. [11.2.11]

Ma, Y.-L. (1982). A simple binomial approximation for hypergeometric distribution, *Acta Mathematicae Applicatae Sinica*, **5**, 418–425 (in Chinese). [6.5]

Maceda, E. C. (1948). On the compound and generalized Poisson distributions, *Annals of Mathematical Statistics*, **19**, 414–416. [8.3.2]

Macfarlane, A. J. (1989). On q-analogues of the quantum harmonic oscillator and the quantum group $SU(2)_q$, *Journal of Physics A: Mathematics and General*, **22**, 4581–4588. [10.8.2]

Madsen, R. W. (1993). Generalized binomial distributions, *Communications in Statistics—Theory and Methods*, **22**, 3065–3086. [3.12.6]

Mainland, D. (1948). Statistical methods in medical research, *Canadian Journal of Research*, **26**(section E), 1–166. [3.8.3]

Makuch, R. W., Stephens, M. A., and Escobar, M. (1989). Generalized binomial models to examine the historical control assumption in active control equivalence studies, *Statistician*, **38**, 61–70. [3.12.6]

Mandelbrot, B. (1959). A note on a class of skew distribution functions: Analysis and critique of a paper by H. A. Simon, *Information and Control*, **2**, 90–99. [6.10.3, 11.2.20]

Mann, N. M., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*, New York: Wiley. [5.2]

Mantel, N. (1951). Evaluation of a class of diagnostic tests, *Biometrics*, **3**, 240–246. [3.11]

Mantel, N. (1962). (Appendix to Haenzel, W., Loveland, D. B., and Sorken, M. B.) Lung cancer mortality as related to residence and smoking histories, I: White males, *Journal of the National Cancer Institution*, **28**, 947–997. [4.7.3]

Mantel, N., and Pasternack, B. S. (1968). A class of occupancy problems, *American Statistician*, **22**, 23–24. [10.4.1]

Margolin, B. H., Resnick, M. A., Rimpo, J. H., Archer, P., Galloway, S. M., Bloom, A. D., and Zeiger, E. (1986). Statistical analyses for *in vitro* cytogenetics assays using Chinese hamster ovary cells, *Environmental Mutagenesis*, **8**, 183–204. [4.9]

Margolin, B. H., and Winokur, H. S. (1967). Exact moments of the order statistics of the geometric distribution and their relation to inverse sampling and reliability of redundant systems, *Journal of the American Statistical Association*, **62**, 915–925. [5.2]

Maritz, J. S. (1952). Note on a certain family of discrete distributions, *Biometrika*, **39**, 196–198. [5.3, 9.3]

Maritz, J. S. (1969). Empirical Bayes estimation for the Poisson distribution, *Biometrika*, **56**, 349–359. [4.7.2, 4.8]

Maritz, J. S., and Lwin, T. (1989). *Empirical Bayes Methods* (second edition), London: Chapman & Hall. [1.2.16, 4.7.2]

Marlow, W. H. (1965). Factorial distributions, *Annals of Mathematical Statistics*, **36**, 1066–1068. [2.7, 6.10.4]

Martin, D. C., and Katti, S. K. (1962). Approximations to the Neyman type A distribution for practical problems, *Biometrics*, **18**, 354–364. [9.6.3]

Martin, D. C., and Katti, S. K. (1965). Fitting of some contagious distributions to some available data by the maximum likelihood method, *Biometrics*, **21**, 34–48 (erratum **21**, 514). [5.10, 8.2.4, 9.5, 9.6.5]

Maruyama, G. (1955). On the Poisson distribution derived from independent random walks, *Ochanomizu University Natural Science Reports*, **6**, 1–6. [4.2.1]

Mathai, A. M., and Saxena, R. K. (1973). *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Berlin: Springer-Verlag. [1.1.8]

Mathai, A. M., and Saxena, R. K. (1978). *The H Function with Applications in Statistics and Other Disciplines*, New Delhi: Wiley Eastern. [1.1.8]

Matsunawa, T. (1986). Poisson distribution, *Encyclopedia of Statistical Sciences*, Vol. 7, S. Kotz, N. L. Johnson, and C. B. Read (editors), 20–25. New York: Wiley. [4.5]

Matuszewski, T. I. (1962). Some properties of Pascal distribution for finite population, *Journal of the American Statistical Association*, **57**, 172–174 (erratum **57**, 919.) [6.3]

Maynard, J. M., and Chow, B. (1972). An approximate Pitman-type "close" estimator for the negative binomial parameter, *P, Technometrics*, **14**, 77–88. [5.8.2]

McCullagh, P. (1991). Quasi-likelihood and estimating functions, *Statistical Theory and Modelling: In Honour of Sir David Cox, FRS*, D. V. Hinkley, N. Reid, and E. J. Snell (editors), 265–286. London: Chapman & Hall. [1.2.16]

McCullagh, P., and Nelder, J. A. (1983). *Generalized Linear Models*, London: Chapman & Hall. [11.1.1]

McGill, W. (1967). Neural counting mechanisms and energy detection in audition, *Journal of Mathematical Psychology*, **4**, 351–376. [5.12.5]

McGrath, E. J., and Irving, D. C. (1973). *Techniques for Efficient Monte Carlo Simulation 2: Random Number Generators for Selected Probability Distributions*, Technical Report SAI-72-590-LJ, La Jolla, CA: Science Applications. [6.6]

McGuire, J. U., Brindley, T. A., and Bancroft, T. A. (1957). The distribution of European corn borer *Pyrausta Nubilalis* (Hbn.) in field corn, *Biometrics*, **13**, 65–78 [errata and extensions (1958) **14**, 432–434]. [9.4, 9.5]

McKay, I. (1996). A note on density estimation for Poisson mixtures, *Statistics and Probability Letters*, **27**, 255–258. [8.3.2]

McKendrick, A. G. (1914). Studies on the theory of continuous probabilities, with special reference to its bearing on natural phenomena of a progressive nature, *Proceedings of the London Mathematical Society*, **13**(2), 401–416. [3.2, 4.2.1, 5.3]

McKendrick, A. G. (1926). Applications of mathematics to medical problems, *Proceedings of the Edinburgh Mathematical Society*, **44**, 98–130. [3.12.2, 4.10.1, 4.12.5, 9.4, 11.2.7, 11.2.10]

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions, *Advances in Applied Probability*, **18**, 679–705. [5.10]

McKenzie, E. (1991). Linear characterizations of the Poisson distribution, *Statistics and Probability Letters*, **11**, 459–461. [4.8]

McLachlan, G. J., and Krishnan T. (1997). *The EM Algorithm*, New York: Wiley. [5.8.5]

McLachlan, G. J., and Peel, D. (2000). *Finite Mixture Models*, New York: Wiley. [6.9.2, 8.1.2, 8.2.2, 11.1.1]

McMahon, P. A. (1894). A certain class of generating functions in the theory of numbers, *Philosophical Transactions of the Royal Society of London, Series A*, **185**, 111–160. [10.3]

McMahon, P. A. (1898). A new method in combinatory analysis with applications to Latin Squares and associated questions, *Transactions of the Cambridge Philosophical Society*, **16**, 262–290. [10.3]

McMahon, P. A. (1902). The problem of derangement in the theory of permutations, *Transactions of the Cambridge Philosophical Society*, **21**, 467–481. [10.3]

McMahon, P. A. (1915). *Combinatory Analysis*, Vol. 1, London: Cambridge University Press (reprinted by Chelsea, New York, 1960). [10.2, 10.3]

McMahon, P. A. (1916). *Combinatory Analysis*, Vol. 2, London: Cambridge University Press (reprinted by Chelsea, New York, 1960). [10.2]

McNally, R. J. Q. (1990). Maximum likelihood estimation of the parameters of the prior distributions of three variables that strongly influence reproductive performance in cows, *Biometrics*, **46**, 501–514. [6.10.4]

Medgyessy, P. (1977). *Decomposition of Superpositions of Density Functions and Discrete Distributions*, Budapest: Akadémia Kiadó; Bristol: Adam Hilger. [8.2.5, 8.2.7]

Medhi, J. (1975). On the convolutions of left-truncated generalized negative binomial and Poisson variables, *Sankhyā, Series B*, **37**, 293–299. [7.2.6]

Medhi, J., and Borah, M. (1984). On generalized Gegenbauer polynomials and associated probabilities, *Sankhyā, Series B*, **46**, 157–165. [7.1.11, 11.2.7]

Meelis, E. (1974). Testing for homogeneity of k independent negative binomial distributed random variables, *Journal of the American Statistical Association*, **69**, 181–186. [5.8.4]

Meilijson, I., Newborn, M. R., Tenenbein, A., and Yechieli, U. (1982). Number of matches and matched people in the birthday problem, *Communications in Statistics—Simulation and Computation*, **11**, 361–370. [10.1, 10.4.1]

Mendenhall, W., and Lehman, E. H. (1960). An approximation to the negative moments of the positive binomial useful in life testing, *Technometrics*, **2**, 233–239. [3.11]

Meredith, W. (1971). Poisson distributions of error in mental test theory, *British Journal of Mathematical and Statistical Psychology*, **24**, 49–82. [5.9.2]

Merlini, D., Sprugnoli, R., and Verri, M. C. (2002). Some statistics on Dyck paths, *Journal of Statistical Planning and Inference*, **101**, 211–227. [6.10.1]

Mertz, D. B., and Davies, R. B. (1968). Cannibalism of the pupal stage by adult flour beetles: An experiment and a stochastic model, *Biometrics*, **24**, 247–275. [10.4.1]

Metzler, D., Grossmann, S., and Wakolbinger, A. (2002). A Poisson model for gapped local alignments, *Statistics and Probability Letters*, **60**, 91–100. [4.2.1]

Meyer, A. (1879). *Vorlesungen über Wahrscheinlichkeitsrechnung* (translated by E. Czuber), Leipzig: Teubner. [5.3]

Mi, J. (1993). Discrete bathtub failure rate and upside-down bathtub mean residual life, *Naval Research Logistics*, **40**, 361–371. [11.2.15]

Mikulski, P. W., and Smith, P. J. (1976). A variance bound for unbiased estimation in inverse sampling, *Biometrika*, **63**, 216–217. [5.8.2]

Miller, A. J. (1961). A queueing model for road traffic flow, *Journal of the Royal Statistical Society, Series B*, **23**, 64–90. [6.10.3, 6.10.4]

Milne-Thompson, L. M. (1933). *The Calculus of Finite Differences*, London: Macmillan. [1.1.3, 1.1.9]

Minkova, L. D. (2002). A generalization of the classical discrete distributions, *Communications in Statistics—Theory and Methods*, **31**, 871–888. [11.2.7]

Mishra, A., and Sinha, J. K. (1981). A generalization of binomial distribution, *Journal of the Indian Statistical Association*, **19**, 93–98. [7.2.5]

Mishra, A., and Tiwary, D. (1985). On generalized logarithmic series distribution, *Journal of the Indian Society of Agricultural Statistics*, **37**, 219–225. [7.2.8]

Misra, N., Singh, H., and Harner, E. J. (2003). Stochastic comparisons of Poisson and binomial random variables with their mixtures, *Statistics and Probability Letters*, **65**, 279–290. [8.3.3]

Mohanty, S. G. (1966). On a generalized two-coin tossing problem, *Biometrische Zeitschrift*, **8**, 266–272. [7.2]

Mohanty, S. G. (1994). Success runs of length-*k* in Markov dependent trials, *Annals of the Institute of Statistical Mathematics, Tokyo*, **46**, 777–796. [10.7.2]

Mohanty, S. G., and Panny, W. (1990). A discrete-time analogue of the M/M/1 queue and the transient solution: A geometric approach, *Sankhyā, Series A (Indian Journal of Statistics)*, **52**, 364–370. [11.2.14]

Molenaar, W. (1970a). *Approximations to the Poisson, Binomial and Hypergeometric Functions*, Tract 31. Amsterdam: Mathematical Centre. [3.6.1, 4.5, 6.4, 6.5]

Molenaar, W. (1970b). Normal approximations to the Poisson distribution, *Random Counts in Scientific Work*, Vol. 2: *Random Counts in Biomedical and Social Sciences*, G. P. Patil (editor), 237–254. University Park: Pennsylvania State University Press. [4.5, 4.7.3]

Molenaar, W. (1973). Simple approximations to the Poisson, binomial, and hypergeometric distributions, *Biometrics*, **29**, 403–408. [4.5, 4.7.3, 6.5]

Molina, E. C. (1942). *Poisson's Exponential Binomial Limit*, New York: Van Nostrand. [4.6.1]

Mood, A. M. (1940). The distribution theory of runs, *Annals of Mathematical Statistics*, **11**, 367–392. [10.6.1]

Mood, A. M. (1943). On the dependence of sampling inspection plans upon population distributions, *Annals of Mathematical Statistics*, **14**, 415–425. [6.8]

Mood, A. M. (1950). *Introduction to the Theory of Statistics*, New York: McGraw-Hill. [6.9.2]

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics* (third edition), New York: McGraw-Hill. [1.1.1]

Moore, P. G. (1954). A note on truncated Poisson distributions, *Biometrics*, **10**, 402–406. [4.10.1, 4.10.2]

Moore, P. G. (1956). The geometric, logarithmic and discrete Pareto forms of series, *Journal of the Institute of Actuaries*, **82**, 130–136. [11.2.20]

Moran, P. A. P. (1952). A characteristic property of the Poisson distribution, *Proceedings of the Cambridge Philosophical Society*, **48**, 206–207. [4.8]

Moran, P. A. P. (1955). A probability theory of dams and storage systems: Modifications of the release rule, *Australian Journal of Applied Science*, **6**, 117–130. [6.11.1]

Moran, P. A. P. (1968). *An Introduction to Probability Theory*, Oxford: Oxford University Press (reprinted, with corrections, 1984). [10.2, 10.4.1]

Moran, P. A. P. (1973). Asymptotic properties of homogeneity tests, *Biometrika*, **60**, 79–85. [4.7.4]

Morgan, B. J. T. (1984). *Elements of Simulation*, London: Chapman & Hall. [1.2.17]

Morgan, B. J. T. (2000). *Applied Stochastic Modelling*, London: Arnold. [11.1.1]

Morice, E., and Thionet, P. (1969). Loi binomiale et loi de Poisson, *Revue de Statistique Appliquée*, **17**, 75–89. [3.6.1]

Morlat, G. (1952). Sur une généralisation de la loi de Poisson, *Comptes Rendus, Académie des Sciences, Paris, Series A*, **235**, 933–935. [4.12.5]

Morris, C. N. (1982). Natural exponential families with quadratic variance functions, *Annals of Statistics*, **10**, 65–80. [3.4]

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: Statistical theory, *Annals of Statistics*, **11**, 515–529. [3.4]

Morris, K. W. (1963). A note on direct and inverse sampling, *Biometrika*, **50**, 544–545. [5.6]

Morse, P. M. (1958). *Queues, Inventories and Maintenance*, New York: Wiley. [10.8.2]

Morton, R. (1991). Analysis of extra-multinomial data derived from extra-Poisson variables conditional on their total, *Biometrika*, **78**, 1–6. [6.11.1]

Mosteller, F., and Youtz, C. (1961). Tables of the Freeman-Tukey transformations for the binomial and Poisson distributions, *Biometrika*, **48**, 433–440. [3.6.3, 4.5]

Mosteller, F. W., and Parunak, A. (1985). Identifying extreme cells in a sizable contingency table: Probabilistic and exploratory approaches, *Exploring Data Tables, Trends and Shapes*, D. C. Hoaglin, F. W. Mosteller, and J. W. Tukey (editors), 189–224. New York: Wiley. [11.1.8]

Muench, H. (1936). The probability distribution of protection test results, *Journal of the American Statistical Association*, **31**, 677–689. [6.9.2]

Muench, H. (1938). Discrete frequency distributions arising from mixtures of several single probability values, *Journal of the American Statistical Association*, **33**, 390–398. [6.9.2]

Mulekar, M. S. (1995). Selecting best negative binomial population using two distance measures, *Statistics and Decisions*, **13**, 53–70. [5.8.5]

Mulekar, M. S., and Young, L. J. (1991). Approximations for a fixed sample size selection procedure for negative binomial populations, *Communications in Statistics—Theory and Methods*, **20**, 1767–1776. [5.8.5]

Mulekar, M. S., and Young, L. J. (1993). A fixed sample size selection procedure for negative binomial populations, *Metrika*, **40**, 25–35. [5.8.5]

Mullahy, J. (1986). Specification and testing in some modified count data models, *Journal of Econometrics*, **33**, 341–365. [3.12.2, 8.2.3]

Murakami, M. (1961). Censored sample from truncated Poisson distribution, *Journal of the College of Arts and Sciences, Chiba University*, **3**, 263–268. [4.10.1]

Murat, M., and Szynal, D. (2003). Moments of certain deformed probability distributions, *Communications in Statistics—Theory and Methods*, **32**, 291–313. [2.2.2, 2.7]

Murthi, B. P. S., Srinivasan, K., and Tadikamalla, P. R. (1993). Robustness of the negative binomial distribution to model purchase incidence: A simulation study, *Communications in Statistics—Simulation and Computation*, **22**, 399–420. [5.12.8]

Muselli M. (2000). New improved bounds for reliability of consecutive-k-out-of-n:F systems, *Journal of Applied Probability*, **37**, 1164–1170. [10.7.2]

Muthu, S. K. (1982). *Problems and Errors for the Physical Sciences*, New Delhi: Orient Longman; London: Sangam Books. [10.4.2]

Nagaraja, H. N. (1982). On the non-Markovian structure of discrete order statistics, *Journal of Statistical Planning and Inference*, **7**, 29–33. [5.2]

Nagaraja, H. N. (1988a). Record values and related statistics—a review, *Communications in Statistics—Theory and Methods*, **17**, 2223–2238. [5.9.1, 10.5]

Nagaraja, H. N. (1988b). Some characterizations of discrete distributions based on linear regressions of adjacent order statistics, *Journal of Statistical Planning and Inference*, **20**, 65–75. [6.10.1]

Nagaraja, H. N. (1990). *Order Statistics from Discrete Distributions*, Technical Report 45, Columbus, OH: Ohio State University Department of Statistics. [1.2.12, 5.2, 5.9.1]

Nagaraja, H. N. (1992). Order statistics from discrete distributions (with discussions and rejoinder), *Statistics* (formerly a series of *Mathematische Operationsforschung und Statistik*), **23**, 189–216. [1.2.12]

Nagaraja, H. N., and Srivastava, R. C. (1987). Some characterizations of geometric type distributions based on order statistics, *Journal of Statistical Planning and Inference*, **17**, 181–191. [5.9.1]

Nair, N. U., and Hitha, N. (1989). Characterization of discrete models by distribution based on their partial sums, *Statistics and Probability Letters*, **8**, 335–337. [6.10.1, 11.2.13]

Nakagawa, T., and Osaki, S. (1975). The discrete Weibull distribution, *IEEE Transactions on Reliability*, **R-24**, 300–301. [11.2.13, 11.2.15]

Nakashima, E. (1997). Some methods for estimation in a negative-binomial model, *Annals of the Institute of Statistical Mathematics, Tokyo*, **49**, 101–115. [5.8.5]

Nam, J.-M. (1995). Confidence limits for the ratio of two binomial proportions based on likelihood scores: Non-iterative method, *Biometrical Journal*, **37**, 375–379. [3.8.3]

Nanapoulos, P. (1977). Zeta laws and arithmetical functions, *Comptes Rendus, Académie des Sciences, Paris, Series A*, **285**, 875–878. [11.2.20]

Nandi, S. B., and Dutta, D. K. (1988). Some developments in the generalized Bell distribution, *Sankhyā, Series B*, **50**, 362–375. [2.6]

Naor, P. (1956). On machine interference, *Journal of the Royal Statistical Society, Series B*, **18**, 280–287. [11.2.12]

Naor, P. (1957). Normal approximation to machine interference with many repairmen, *Journal of the Royal Statistical Society, Series B*, **1**, 334–341. [11.2.12, 11.2.19]

Narayana, T. V. (1959). A partial order and its application to probability theory, *Sankhyā*, **21**, 91–98. [6.10.5]

Narayana, T. V. (1979). *Lattice Path Combinatorics with Statistical Applications*, Toronto: University of Toronto Press. [6.10.5]

National Bureau of Standards (1950). *Tables of the Binomial Probability Distribution*, Washington, DC: U.S. Government Printing Office. [3.7.1]

Nayatani, Y., and Kurahara, B. (1964). A condition for using the approximation by the normal and the Poisson distribution to compute the confidence intervals for the binomial parameter, *Reports of Statistical Application Research, JUSE*, **11**, 99–105. [3.8.3]

Nedelman, J., and Wallenius, T. (1986). Bernoulli trials, Poisson trials, surprising variances, and Jensen's inequality, *American Statistician*, **40**, 286–289. [3.12.3]

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society, Series A*, **135**, 370–384. [11.1.2]

Nelson, D. L. (1975). Some remarks on generalizations of the negative binomial and Poisson distributions, *Technometrics*, **17**, 135–136. [7.2.6]

Nelson, W. C., and David, H. A. (1967). The logarithmic distribution: A review, *Virginia Journal of Science*, **18**, 95–102. [7.1.2]

Neuman, P. (1966). Uber den Median der Binomial und Poissonverteilung, *Wissenschaftliche Zeitschrift der Technischen Universität Dresden*, **15**, 223–226. [3.6.2]

Neville, A. M., and Kemp, C. D. (1975). On characterizing the hypergeometric and multivariate hypergeometric distributions, *Statistical Distributions in Scientific Work*, Vol. 3: *Characterizations and Applications*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 353–357. Dordrecht: Reidel. [6.8]

Nevzorov, V. B. (1987). "Records," *Theory of Probability and Its Applications*, **32**, 201–228. [10.5]

Newby, M. (1999). Moments and generating functions for the absorption distribution and its negative binomial analogue, *Communications in Statistics—Theory and Methods*, **28**, 2935–2945. [10.8.1, 10.8.2]

Newell, D. J. (1965). Unusual frequency distributions, *Biometrics*, **21**, 159–168. [3.11, 4.10.2]

Neyman, J. (1939). On a new class of "contagious" distributions applicable in entomology and bacteriology, *Annals of Mathematical Statistics*, **10**, 35–57. [9.1, 9.9]

Neyman, J. (1965). Certain chance mechanisms involving discrete distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 4–14. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [7.2.1]

Neyman, J., and Scott, E. L. (1964). A stochastic model of epidemics, *Stochastic Models in Medicine and Biology*, J. Gurland (editor), 45–83. Madison, WI: University of Wisconsin Press. [7.2.1, 11.2.10]

Neyman, J., and Scott, E. L. (1966). On the use of $C(\alpha)$ optimal tests of composite hypotheses, *Bulletin of the International Statistical Institute*, **41**(1), 477–497. [4.7.4]

Ng, T.-H. (1989). A new class of modified binomial distributions with applications to certain toxicological experiments, *Communications in Statistics—Theory and Methods*, **18**, 3477–3492. [3.12.6]

Nicholson, W. L. (1956). On the normal approximation to the hypergeometric distribution, *Annals of Mathematical Statistics*, **27**, 471–483. [6.4, 6.5]

Nicholson, W. L. (1961). Occupancy probability distribution critical points, *Biometrika*, **48**, 175–180. [10.4.1]

Nishimura, K., and Sibuya, M. (1997). Extended Stirling family of discrete probability distributions, *Communications in Statistics—Theory and Methods*, **26**, 1727–1744. [7.1.12]

Nisida, T. (1962). On the multiple exponential channel queueing system with hyper-Poisson arrivals, *Journal of the Operations Research Society, Japan*, **5**, 57–66. [4.12.4]

Noack, A. (1950). A class of random variables with discrete distributions, *Annals of Mathematical Statistics*, **21**, 127–132. [2.2.1, 3.4, 6.2.5]

Nörlund, N. E. (1923). *Vorlesungen über Differenzenrechnung*, New York: Chelsea (1954 reprint). [1.1.9, 11.2.2]

O'Carroll, F. M. (1962). Fitting a negative binomial distribution to coarsely grouped data by maximum likelihood, *Applied Statistics*, **11**, 196–201. [5.8.3]

O'Hagan, A. (1994). *Bayesian Inference, Kendall's Advanced Theory of Statistics*, London: Arnold. [1.2.2, 1.2.16]

Okamoto, M. (1958). Some inequalities relating to the partial sum of binomial probabilities, *Annals of the Institute of Statistical Mathematics, Tokyo*, **10**, 29–35. [3.6.2]

Olds, E. G. (1938). A moment generating function which is useful in solving certain matching problems, *Bulletin of the American Mathematical Society*, **44**, 407–413. [10.3]

Oliver, R. M. (1961). A traffic counting distribution, *Operations Research*, **10**, 105–114. [11.2.9]

Olkin, I., Petkau, A. J., and Zidek, J. V. (1981). A comparison of *n* estimators for the binomial distribution, *Journal of the American Statistical Association*, **76**, 637–642. [3.8.2]

Olofsson, P. (1999). A Poisson approximation with applications to the number of maxima in a discrete sample, *Statistics and Probability Letters*, **44**, 23–27. [4.2.2]

Olshen, R. A., and Savage, L. J. (1970). A generalized unimodality, *Journal of Applied Probability*, **7**, 21–34. [1.2.7]

Oluyede, B. O. (1994). Tests for equality of several binomial populations against an order restricted alternative and model selection for one-dimensional multinomials, *Biometrical Journal*, **36**, 17–32. [3.8.4]

Ong, S. H. (1987). Soms notes on the non-central negative binomial distribution, *Metrika*, **34**, 225–236. [7.1.11]

Ong, S. H. (1988). A discrete Charlier distribution, *Biometrical Journal*, **8**, 1003–1009. [5.12.5, 11.1.6]

Ong, S. H. (1995a). Some stochastic models leading to the convolution of two binomial variables, *Statistics and Probability Letters*, **22**, 161–166. [3.12.2, 11.2.7]

Ong, S. H. (1995b). Computation of probabilities of a generalized log-series and related distributions, *Communications in Statistics—Theory and Methods*, **24**, 253–271. [7.1.11]

Ong, S. H. (1995c). Characterization of the negative binomial and gamma distributions, *Biometrical Journal*, **37**, 251–254. [5.9.2]

Ong, S. H., and Lee, P. A. (1979). The non-central negative binomial distribution, *Biometrical Journal*, **21**, 611–628. [5.12.5]

Ong, S. H., and Lee, P. A. (1986). On a generalized non-central negative binomial distribution, *Communications in Statistics—Theory and Methods*, **15**, 1065–1079. [5.12.5]

Ong, S. H., and Lee P. A. (1992). Orthogonal polynomials of a class of discrete distributions with application, *Sankhyā, Series B*, **54**, 170–183. [9.12]

Ong, S. H., and Muthaloo, S. (1995). A class of discrete distributions suited to fitting very long-tailed data, *Communications in Statistics—Simulation and Computation*, **24**, 929–945. [7.1.11]

Ong, S. H., and Toh, K. K. (2001). Cluster size distribution of the non-central negative binomial distribution, *Communications in Statistics—Theory and Methods*, **30**, 2415–2426. [5.12.5]

Ord, J. K. (1967a). Graphical methods for a class of discrete distributions, *Journal of the Royal Statistical Society, Series A*, **130**, 232–238. [2.3.3, 3.8.1, 4.7.1, 5.8.1, 6.1, 6.2.5, 7.1.7]

Ord, J. K. (1967b). On a system of discrete distributions, *Biometrika*, **54**, 649–656. [2.3.3, 3.4]

Ord, J. K. (1967c). *On Families of Discrete Distributions*, PhD Thesis, London: University of London. [2.3.3, 2.4.1, 11.2.6]

Ord, J. K. (1968a). Approximations to distribution functions which are hypergeometric series, *Biometrika*, **55**, 243–248. [6.5]

Ord, J. K. (1968b). The discrete Student's t distribution, *Annals of Mathematical Statistics*, **39**, 1513–1516. [11.2.6]

Ord, J. K. (1970). The negative binomial model and quadrat sampling, *Random Counts in Scientific Work*, Vol. 2: *Random Counts in Biomedical and Social Sciences*, G. P. Patil (editor), 151–163. University Park: Pennsylvania State University Press. [3.8.1]

Ord, J. K. (1972). *Families of Frequency Distributions*, London: Griffin. [2.3.3, 4.7.1, 5.8.1, 7.1.7, 9.10]

Ord, J. K. (1985). Pearson systems of distributions, *Encyclopedia of Statistical Sciences*, Vol. 6, S. Kotz, N. L. Johnson, and C. B. Read (editors), 655–659. New York: Wiley. [2.3.3]

Ord, J. K., and Whitmore, G. (1986). The Poisson-inverse Gaussian distribution as a model for species abundance, *Communications in Statistics—Theory and Methods*, **15**, 853–871. [11.1.5]

Otter, R. (1949). The multiplicative process, *Annals of Mathematical Statistics*, **20**, 206–224. [7.2.1, 11.2.10]

Ottestad, P. (1939). On the use of the factorial moments in the study of discontinuous frequency distributions, *Skandinavisk Aktuarietidskrift*, **22**, 22–31. [2.3.1, 5.8.1]

Ottestad, P. (1943). On Bernoullian, Lexis, Poisson and Poisson-Lexis series, *Skandinavisk Aktuarietidskrift*, **26**, 15–67. [3.12.3]

Ottestad, P. (1944). On certain compound frequency distributions, *Skandinavisk Aktuarietidskrift*, **27**, 32–42. [5.9.2, 8.3.2]

Owen, A. R. G. (1965). The summation of class frequencies (Appendix to Bliss, C. I.: An analysis of some insect trap records), *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 385–397. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [7.1.5]

Owen, D. B. (1962). *Handbook of Statistical Tables*, Reading, MA: Addison-Wesley. [6.7.1, 6.9.1, 7.2.2, 10.3, 10.4.1]

Pachares, J. (1960). Tables of confidence limits for the binomial distribution, *Journal of the American Statistical Association*, **55**, 521–533. [3.8.3]

Padgett, W. J., and Spurrier, J. D. (1985). On discrete failure models, *IEEE Transactions on Reliability*, **34**, 253–256. [11.2.15]

Pakes, A. G. (1993). Remarks on the Luria–Delbrück distribution, *Journal of Applied Probability*, **30**, 991–994. [11.2.11]

Pakes, A. G. (1995). Characterization of discrete laws via mixed sums and Markov branching processes, *Stochastic Processes and Their Applications*, **55**, 285–300. [11.2.5]

Paloheimo, J. E. (1963). On statistics of search, *Bulletin of the International Statistical Institute*, **40**(2), 1060–1061. [7.1.10]

Panaretos, J. (1982). An extension of the damage model, *Metrika*, **29**, 189–194. [9.2]

Panaretos, J. (1983a). A generating model involving Pascal and logarithmic series distributions, *Communications in Statistics—Theory and Methods*, **12**, 841–848. [7.1.11]

Panaretos, J. (1983b). On Moran's property of the Poisson distribution, *Biometrical Journal*, **25**, 69–76. [4.8]

Panaretos, J. (1987a). Convolution and mixing properties of the damage model, *Statistica*, **47**, 1–8. [9.2]

Panaretos, J. (1987b). On a functional equation for the generating function of the logarithmic series distribution, *Revue Roumaine de Mathématiques Pures et Appliquées*, **32**, 365–367. [7.1.2]

Panaretos, J. (1989a). A probability model involving the use of the zero-truncated Yule distribution for analysing surname data, *IMA Journal of Mathematics Applied in Medicine and Biology*, **6**, 133–136. [6.10.3]

Panaretos, J. (1989b). On the evolution of surnames, *International Statistical Review*, **57**, 161–167. [6.10.3]

Panaretos, J., and Xekalaki, E. (1986a). On some distributions arising from certain generalized sampling schemes, *Communications in Statistics—Theory and Methods*, **15**, 873–891. [6.9.3, 7.1.12, 9.4, 10.7.4, 10.7.6]

Panaretos, J., and Xekalaki, E. (1986b). On generalized binomial and multinomial distributions and their relation to generalized Poisson distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **38**, 223–231. [10.7.6]

Panaretos, J., and Xekalaki, E. (1986c). The stuttering generalized Waring distribution, *Statistics and Probability Letters*, **4**, 313–318. [10.7.6]

Panaretos, J., and Xekalaki, E. (1989). A probability distribution associated with events with multiple occurrences, *Statistics and Probability Letters*, **8**, 389–396. [10.7.6]

Papageorgiou, H. (1985). On characterizing some discrete distributions by a conditional distribution and a regression function, *Biometrical Journal*, **27**, 473–479. [1.2.9, 6.8]

Parthasarathy, P. R. (1987). A transient solution to an M/M/1 queue: A simple solution, *Advances in Applied Probability*, **19**, 997–998. [11.2.14]

Parzen, E. (1960). *Modern Probability Theory and Its Applications*, New York: Wiley. [10.2, 10.4.1]

Parzen, E. (1962). *Stochastic Processes with Applications to Science and Engineering*, San Francisco: Holden-Day. [4.2.1, 4.9]

Pascal, B. (1679). *Varia opera Mathematica D. Petri de Fermat*, Tolossae. [5.1, 5.3]

Patel, J. (1973). A catalogue of failure distributions, *Communications in Statistics*, **1**, 281–284. [6.10.1, 7.1.4, 11.2.15]

Patel, J. K., Kapardia, C. H., and Owen, D. B. (1976). *Handbook of Statistical Distributions*, New York: Dekker. [1.2.6, 1.2.7, 3.8.2, 3.8.3]

Patel, S. R., and Jani, P. N. (1977). On minimum variance unbiased estimation of generalized Poisson distributions and decapitated generalized Poisson distributions, *Journal of the Indian Statistical Association*, **15**, 157–159. [2.2.2]

Patel, Y. C. (1971). *Some Problems in Estimation for the Parameters of the Hermite Distribution*, PhD Thesis, Athens, GA: University of Georgia. [9.4]

Patel, Y. C. (1976a). Estimation of the parameters of the triple and quadruple stuttering-Poisson distributions, *Technometrics*, **18**, 67–73. [9.3]

Patel, Y. C. (1976b). Even point estimation and moment estimation in Hermite distribution, *Biometrics*, **32**, 865–873. [9.4]

Patel, Y. C. (1977). Higher moments of moment estimators and even point estimators for the parameters of the Hermite distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **29A**, 119–130. [9.4]

Patel, Y. C. (1985). An asymptotic expression for cumulative sum of probabilities of the Hermite distribution, *Communications in Statistics—Theory and Methods*, **14**, 2233–2241. [9.4]

Patel, Y. C., Shenton, L. R., and Bowman, K. O. (1974). Maximum likelihood estimation for the parameters of the Hermite distribution, *Sankhyā, Series B*, **36**, 154–162. [9.4]

Patil, G. P. (1961). *Contributions to Estimation in a Class of Discrete Distributions*, PhD Thesis, Ann Arbor, MI: University of Michigan. [2.2.1]

Patil, G. P. (1962a). Certain properties of the generalized power series distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **14**, 179–182. [2.2.1, 2.2.2, 4.8]

Patil, G. P. (1962b). On homogeneity and combined estimation for the generalized power series distribution and certain applications, *Biometrics*, **18**, 365–374. [2.2.1]

Patil, G. P. (1962c). Maximum-likelihood estimation for generalized power series distributions and its application to a truncated binomial distribution, *Biometrika*, **49**, 227–237. [2.2.1]

Patil, G. P. (1962d). Some methods of estimation for the logarithmic series distribution, *Biometrics*, **18**, 68–75. [7.1.6, 7.1.7]

Patil, G. P. (1963a). On the equivalence of the binomial and inverse binomial acceptance sampling plans and an acknowledgement, *Technometrics*, **5**, 119–121. [5.6]

Patil, G. P. (1963b). Minimum variance unbiased estimation and certain problems of additive number theory, *Annals of Mathematical Statistics*, **34**, 1050–1056. [2.2.1]

Patil, G. P. (1964a). Estimation for the generalized power series distribution with two parameters and its application to binomial distribution, *Contributions to Statistics*, C. R. Rao (editor), 335–344. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [2.2.1]

Patil, G. P. (1964b). On certain compound Poisson and compound binomial distributions, *Sankhyā, Series A*, **26**, 293–294. [7.1.7, 7.1.10, 8.2.4, 9.4]

Patil, G. P. (1985). Logarithmic series distribution, *Encyclopedia of Statistical Sciences*, Vol. 5, S. Kotz, N. L. Johnson, and C. B. Read (editors), 111–114. New York: Wiley. [7.1.6]

Patil, G. P. (1986). Power series distributions, *Encyclopedia of Statistical Sciences*, Vol. 7, S. Kotz, N. L. Johnson, and C. B. Read (editors), 130–134. New York: Wiley. [3.4]

Patil, G. P., and Bildikar, S. (1966). Identifiability of countable mixtures of discrete probability distributions using methods of infinite matrices, *Proceedings of the Cambridge Philosophical Society*, **62**, 485–494. [8.3.1]

Patil, G. P., Boswell, M. T., Joshi, S. W., and Ratnaparkhi, M. V. (1984). *Dictionary and Bibliography of Statistical Distributions in Scientific Work*, Vol. 1: *Discrete Models*, Fairland, MD: International Co-operative Publishing House. [1.1.3, 2.4.1, 2.4.2, 5.1, 6.9.1, 6.9.2, 7.1.7, 10.7.1, 11.2.1, 11.2.18, 11.2.21]

Patil, G. P., and Joshi, S. W. (1968). *A Dictionary and Bibliography of Discrete Distributions*, Edinburgh: Oliver and Boyd. [2.4.1, 2.4.2, 5.3, 11.2.13]

Patil, G. P., and Joshi, S. W. (1970). Further results on minimum variance unbiased estimation and additive number theory, *Annals of Mathematical Statistics*, **41**, 567–575. [2.2.1]

Patil, G. P., Kamat, A. R., and Wani, J. K. (1964). *Certain Studies on the Structure and Statistics of the Logarithmic Series Distribution and Related Tables*, Report No. ARL 64–197, Wright-Patterson Air Force Base: OH: Aerospace Research Laboratories. [7.1.6, 7.1.7]

Patil, G. P., and Rao, C. R. (1978). Weighted distributions and size biased sampling with application to wildlife populations and human families, *Biometrics*, **34**, 179–189. [3.12.4]

Patil, G. P., Rao, C. R., and Ratnaparkhi, M. V. (1986). On discrete weighted distributions and their use in model choice for observed data, *Communications in Statistics—Theory and Methods*, **15**, 907–918. [2.4.1, 3.12.4]

Patil, G. P., Rao, C. R., and Zelen, M. (1986). *A Computerized Bibliography of Weighted Distributions and Related Weighted Methods for Statistical Analysis and Interpretations of Encountered Data, Observational Studies, Representativeness Issues, and Resulting Inferences*, University Park, PA: Centre for Statistical Ecology and Environmental Statistics, Pennsylvania State University. [3.12.4]

Patil, G. P., Rao, C. R., and Zelen, M. (1988). Weighted distributions, *Encyclopedia of Statistical Sciences*, Vol. 9, S. Kotz, N. L. Johnson, and C. B. Read (editors), 565–571. New York: Wiley. [3.12.4]

Patil, G. P., and Ratnaparkhi, M. V. (1977). Characterizations of certain statistical distributions based on additive damage models involving Rao-Rubin condition and some of its variants, *Sankhyā, Series B*, **39**, 65–75. [4.8, 6.8]

Patil, G. P., and Seshadri, V. (1964). Characterization theorems for some univariate probability distributions, *Journal of the Royal Statistical Society, Series B*, **26**, 286–292. [3.9, 4.8, 5.9.1, 5.9.2, 6.8, 6.10.1]

Patil, G. P., and Seshadri, V. (1975). A characteristic property of certain generalized power series distributions, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 19–30. Dordrecht: Reidel. [7.1.4]

Patil, G. P., and Wani, J. K. (1965a). Maximum likelihood estimation for the complete and truncated logarithmic series distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 398–409. Calcutta: Statistical Publishing Society; Oxford: Pergamon. (Also *Sankhyā, Series A*, **27**, 281–292.) [7.1.6, 7.1.7, 7.1.10]

Patil, G. P., and Wani, J. K. (1965b). On certain structural properties of the logarithmic series distribution and the first type Stirling distribution, *Sankhyā, Series A*, **27**, 271–280. [7.1.7, 7.1.8, 7.1.12]

Patil, S. A., and Raghunandanan, K. (1990). Compound Hermite and stuttering Poisson distributions, *Bulletin of the Calcutta Statistical Association*, **39**, 97–103. [9.4, 11.2.7]

Paul, S. R. (1985). A three-parameter generalization of the binomial distribution, *Communications in Statistics—Theory and Methods*, **14**, 1497–1506. [3.12.6, 11.1.1]

Paul, S. R. (1987). On the beta-correlated binomial (BCB) distribution: A three-parameter generalization of the binomial distribution, *Communications in Statistics—Theory and Methods*, **16**, 1473–1478. [3.12.6, 11.1.1]

Paulson, A. S., and Uppuluri, V. R. R. (1972). A characterization of the geometric distribution and a bivariate geometric distribution, *Sankhyā, Series A*, **34**, 297–301. [5.9.1]

Pearson, E. S. (1925). Bayes' theorem, examined in the light of experimental sampling, *Biometrika*, **17**, 388–442. [6.9.2]

Pearson, E. S., and Hartley, H. O. (1976). *Biometrika Tables for Statisticians* (third edition), Vol. 1, London: Biometrika Trust. [1.1.5, 3.7.1, 3.8.3, 4.6.1, 4.7.3, 6.9.1]

Pearson, K. (1895). Contributions to the mathematical theory of evolution I. Skew distribution in homogeneous material, *Philosophical Transactions of the Royal Society of London, Series A*, **186**, 343–414. [2.3.1, 6.2.1]

Pearson, K. (1899). On certain properties of the hypergeometrical series, and on the fitting of such series to observation polygons in the theory of chance, *Philosophical Magazine, 5th series*, **47**, 236–246. [6.2.1, 6.3]

Pearson, K. (1906). On the curves which are most suitable for describing the frequency of random samples of a population, *Biometrika*, **5**, 172–175. [6.5]

Pearson, K. (1907). On the influence of past experience on future expectation, *Philosophical Magazine, Series B*, **13**, 365–378. [6.2.2]

Pearson, K. (1915). On certain types of compound frequency distributions in which the components can be individually described by binomial series, *Biometrika*, **11**, 139–144. [8.1.2]

Pearson, K. (editor) (1922). *Tables of the Incomplete* Γ*-Function*, London: Her Majesty's Stationery Office. [1.1.5]

Pearson, K. (1924). On the moments of the hypergeometrical series, *Biometrika*, **16**, 157–162. [6.2.1, 6.3]

Pearson, K. (editor) (1934). *Tables of the Incomplete Beta-Function*, London: Cambridge University Press. [1.1.5, 3.7.1, 3.8.3]

Peizer, D. B., and Pratt, J. W. (1968). A normal approximation for binomial, F, beta and other common related tail probabilities, I, *Journal of the American Statistical Association*, **63**, 1416–1456. [3.6.1, 4.5, 5.6]

Peng, J. C. M. (1975). *Simultaneous Estimation of the Parameters of Independent Poisson Distributions*, Technical Report No. 78, Stanford, CA: Stanford University Department of Statistics. [4.7.2]

Pérez-Abreu, V. (1991). Poisson approximation to power series distributions, *American Statistician*, **45**, 42–45. [4.2.2]

Perline, R. (1998). Mixed Poisson distributions tail equivalent to their mixing distributions, *Statistics and Probability Letters*, **38**, 229–233. [8.3.2]

Perry, J. N. (1984). Negative binomial model for mosquitoes, *Biometrics*, **40**, 863. [5.10]

Perry, J. N., and Taylor, L. R. (1985). Adès: New ecological families of species-specific frequency distributions that describe repeated spatial samples with an intrinsic power-law variance-mean property, *Journal of Animal Ecology*, **54**, 931–953. [11.2.3]

Perry, J. N., and Taylor, L. R. (1988). Families of distributions for repeated samples of animal counts, *Biometrics*, **44**, 881–888. [11.2.3]

Pessin, V. (1961). Some asymptotic properties of the negative binomial distribution, *Annals of Mathematical Statistics*, **32**, 922–923 (abstract). [5.5]

Pessin, V. (1965). Some discrete distribution limit theorems using a new derivative, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 109–122. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [5.5]

Peterson, A. V., and Kronmal, R. A. (1980). A representation for discrete distributions by equiprobable mixtures, *Journal of Applied Probability*, **17**, 102–111. [8.1.2]

Peterson, A. V., and Kronmal, R. A. (1982). On mixture methods for the computer generation of random variables, *American Statistician*, **36**, 184–191. [8.1.2]

Peterson, C. G. J. (1896). The yearly immigration of young plaice into the Linfjord from the German Sea, *Danish Biological Station Report*, **6**, 5–48. [6.9.1]

Peto, S. (1953). A dose-response equation for the invasion of microorganisms, *Biometrics*, **9**, 320–335. [10.4.1]

Pettigrew, H. M., and Mohler, W. C. (1967). A rapid test for the Poisson distribution using the range, *Biometrics*, **23**, 685–692. [4.5, 4.7.1]

Phatarfod, R. M., and Mardia, K. V. (1973). Some results for dams with Markovian inputs, *Journal of Applied Probability*, **10**, 166–180. [3.12.2]

Philippou, A. N. (1983). The Poisson and compound Poisson distributions of order k and some of their properties, *Zapisky Nauchnyka Seminarov Leningradskogo Otdelinya Matematscheskogo Instituta im V. A. Steklova AN SSSR*, **130**, 175–180 (in Russian). [4.11, 10.7.1, 10.7.6]

Philippou, A. N. (1984). The negative binomial distribution of order k and some of its properties, *Biometrical Journal*, **26**, 789–794. [10.7.1, 10.7.3]

Philippou, A. N. (1986). Distributions and Fibonacci polynomials of order k, longest runs, and reliability of consecutive-k-out-of-n:F systems, *Fibonacci Numbers and Their Applications*, A. N. Philippou, G. E. Bergum, and A. F. Horadam (editors), 203–227. Dordrecht: Reidel. [10.7.1, 10.7.5]

Philippou, A. N. (1988). On multiparameter distributions of order k, *Annals of the Institute of Statistical Mathematics, Tokyo*, **40**, 467–475. [10.7.6]

Philippou, A. N. (1989). Mixtures of distributions by the Poisson distribution of order k, *Biometrical Journal*, **31**, 67–74. [2.8, 10.7.6]

Philippou, A. N., and Antzoulakos, D. L. (1990). Multivariate distributions of order k on a generalized sequence, *Statistics and Probability Letters*, **9**, 453–463. [10.7.6]

Philippou, A. N., Georghiou, C., and Philippou, G. N. (1983). A generalized geometric distribution and some of its probabilities, *Statistics and Probability Letters*, **1**, 171–175. [2.8, 10.7.1]

Philippou, A. N., and Makri, F. S. (1986). Success runs and longest runs, *Statistics and Probability Letters*, **4**, 101–105 (corrected version 211–215). [10.7.5]

Philippou, A. N., and Muwafi, A. A. (1982). Waiting for the k-th consecutive success and the Fibonacci sequence of order k, *Fibonacci Quarterly*, **20**, 28–32. [10.7.1, 10.7.2]

Philippou, A. N., Tripsiannis, G. A., and Antzoulakos, D. L. (1989). New Polya and inverse Polya distributions of order k, *Communications in Statistics-Theory and Methods*, **18**, 2125–2137. [10.7.6]

Philipson, C. (1960a). The theory of confluent hypergeometric functions and its application to Poisson processes, *Skandinavisk Aktuarietidskrift*, **43**, 136–162. [9.7]

Philipson, C. (1960b). Note on the application of compound Poisson processes to sickness and accident statistics, *ASTIN Bulletin*, **1**, 224–237. [8.3.3]

Phillips, M. J. (1978). Sums of random variables having the modified geometric distribution with application to two-person games, *Advances in Applied Probability*, **10**, 647–665. [11.2.7]

Philpot, J. W. (1964). *Orthogonal Parameters for Two Parameter Distributions*, MS Thesis, Blacksburg, VA: Virginia Polytechnic Institute. [9.6.3]

Pichon, G., Merlin, M., Fagneaux, G., Riviere, F., and Laigret, J. (1976). *Etude de la Distribution des Densités Microfilariennes dans des Foyers de Filariose Lymphatique*, "Louis Malarde" Technical Report, Papeete, Tahiti: Institut de Recherches Medicale. [5.11]

Piegorsch, W. W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter, *Biometrics*, **46**, 863–867. [5.8.3, 5.8.5]

Pielou, E. C. (1957). The effect of quadrat size on the estimation of the parameters of Neyman's and Thomas's distributions, *Journal of Ecology*, **45**, 31–47. [9.6.4, 9.6.5]

Pielou, E. C. (1962). Runs of one species with respect to another in transects through plant populations, *Biometrics*, **18**, 579–593. [5.2, 6.10.3, 6.10.4]

Pielou, E. C. (1963). Runs of healthy and diseased trees in transects through an infected forest, *Biometrics*, **19**, 603–614. [5.2]

Pieters, E. P., Gates, C. E., Matis, J. H., and Sterling, W. L. (1977). Small sample comparisons of different estimators of negative binomial parameters, *Biometrics*, **33**, 718–723. [5.8.3]

Pillai, R. N. (1990). On Mittag-Leffler functions and related distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **42**, 157–161. [5.12.2, 11.2.5]

Pillai, R. N., and Jayakumar, K. (1995). Discrete Mittag-Leffler distributions, *Statistics and Probability Letters*, **42**, 571–161. [5.12.2, 11.2.5]

Plackett, R. L. (1953). The truncated Poisson distribution, *Biometrics*, **9**, 485–488. [4.10.1]

Plunkett, I. G., and Jain, G. C. (1975). Three generalized negative binomial distributions, *Biometrische Zeitschrift*, **17**, 276–302. [9.4, 11.2.7]

Pohlmeier, W., and Ulrich, V. (1995). An econometric model of the two-part decision making process in the demand for health care, *Journal of Human Resources*, **30**, 339–361. [5.12.9]

Poisson, S. D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Regles Générales du Calcul des Probabilitiés*, Paris: Bachelier, Imprimeur-Libraire pour les Mathematiques, la Physique, etc. [3.12.3, 4.2.1]

Pólya, G. (1930). Sur quelques points de la théorie des probabilités, *Annales de l'Institut H. Poincaré*, **1**, 117–161. [6.2.4, 9.7]

Pólya, G. (1970). Gaussian binomial coefficients and the enumeration of inversions, *Proceedings of the Second Chapel Hill Conference on Combinatorial Mathematics and Its Applications, August 1970*, Chapel Hill: University of North Carolina. [2.9, 10.8.4]

Potthof, R. F., and Whittinghill, M. (1966). Testing for homogeneity II: The Poisson distribution, *Biometrika*, **53**, 183–190. [4.7.4]

Prasad, A. (1957). A new discrete distribution, *Sankhyā*, **17**, 353–354. [6.10.3]

Prasaka Rao, B. L. S. (1992). *Identifiability in Stochastic Models: Characterization of Probability Distributions*, San Diego: Academic Press. [1.2.9]

Pratt, J. W. (1968). A normal approximation for binomial, F, beta and other common related tail probabilities, II, *Journal of the American Statistical Association*, **63**, 1457–1483. [3.6.1, 5.6]

Prekopa, A. (1952). On composed Poisson distributions, IV, *Acta Mathematica Academiae Scientiarum Hungaricae*, **3**, 317–325. [4.12.3]

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by measurement errors, *Journal of the American Statistical Association*, **81**, 321–327. [3.12.6, 6.9.2]

Preston, F. W. (1948). The commonness, and rarity, of species, *Ecology*, **29**, 254–283. [7.1.12]

Prodinger, H. (1996). Asymptotics of the Luria-Delbrück distribution via singularity analysis, *Journal of Applied Probability*, **33**, 282–283. [11.2.11]

Prodinger, H. (2004). On the moments of a distribution defined by the Gaussian polynomials, *Journal of Statistical Planning and Inference*, **119**, 237–239. [10.8.4]

Prohorov, Y. V. (1953). The asymptotic behaviour of the binomial distribution, *Uspekhi Matematicheskii Nauk* (*New Series*), **8**, 135–142. [3.6.1, 3.6.2]

Pulskamp, R. (1990). A note on the estimation of binomial probabilities, *American Statistician*, **44**, 293–295. [3.8.2]

Puri, P. S. (1966). Probability generating function of absolute difference of two random variables, *Proceedings of the National Academy of Science*, **56**, 1059–1061. [5.9.1]

Puri, P. S. (1973). On a property of exponential and geometric distributions and its relevance to multivariate failure rate, *Sankhyā, Series A*, **35**, 61–78. [5.9.1]

Puri, P. S., and Rubin, H. (1970). A characterization based on the absolute difference of two i.i.d. random variables, *Annals of Mathematical Statistics*, **41**, 2113–2122. [5.9.1]

Puri, P. S., and Rubin, H. (1972). On a characterization of the family of distributions with constant multivariate failure rates, *Annals of Probability*, **2**, 738–740. [5.9.1]

Qu, Y., Beck, G. J., and Williams, G. W. (1990). Polya-Eggenberger distribution: Parameter estimation and hypothesis tests, *Biometrical Journal*, **32**, 229–242. [6.7.2, 6.8]

Qu, Y., Greene T., and Piedmonte, M. R. (1993). Symmetric Bernoulli distributions and generalized binomial distributions, *Biometrical Journal*, **35**, 523–533. [3.12.6]

Quenouille, M. H. (1949). A relation between the logarithmic, Poisson, and negative binomial series, *Biometrics*, **5**, 162–164. [5.3, 7.1.2]

Quine, M. P., and Seneta, E. (1987). Bortkiewicz's data and the law of small numbers, *International Statistical Review*, **55**, 173–181. [4.2.1]

Qvale, P. (1932). Remarks on semi-invariants and incomplete moments, *Skandinavisk Aktuarietidskrift*, **15**, 196–210. [2.4.1]

Rabinovitch, N. L. (1973). Studies in the history of probability and statistics, XXII: Probability in the Talmud, *Biometrika*, **56**, 437–441. [11.2.19]

Raff, M. S. (1956). On approximating the point binomial, *Journal of the American Statistical Association*, **51**, 293–303. [3.6.1]

Rahme, E., and Joseph, L. (1998). Exact sample size determination for binomial experiments, *Journal of Statistical Planning and Inference*, **66**, 83–93. [3.8.4]

Raiffa, H., and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, Cambridge, MA: MIT Press. [3.4, 6.2.3, 6.4]

Raikov, D. (1938). On the decomposition of Gauss' and Poisson's laws, *Izvestia Akademie Nauk SSSR, Series A*, **2**, 91–124. [4.8]

Rainville, E. R. (1960). *Special Functions*, New York: Macmillan. [1.1.8, 11.2.7]

Ramasubban, T. A. (1958). The generalized mean differences of the binomial and Poisson distributions, *Biometrika*, **45**, 549–556. [4.3]

Rao, B. L. S. P., and Sreehari, M. (1987). On a characterization of Poisson distribution through inequalities of Chernoff-type, *Australian Journal of Statistics*, **29**, 38–41. [4.8]

Rao, B. R. (1981). Correlation between the numbers of two types of children in a family with the MPSD for the family size, *Communications in Statistics—Theory and Methods*, **10**, 249–254. [7.2.8]

Rao, B. R., and Janardan, K. G. (1984). Use of the generalized Markov-Polya distribution as a random damage model and its identifiability, *Sankhyā, Series A*, **46**, 458–462. [6.11.4]

Rao, B. R., and Janardan, K. G. (1985). An analog of the Rao-Rubin condition for distributions other than the Poisson, *Pakistan Journal of Statistics*, **1**, 1–15. [9.2]

Rao, B. R., Mazumdar, S., Waller, J. H., and Li, C. C. (1973). Correlation between the numbers of two types of children in a family, *Biometrics*, **29**, 271–279. [5.10]

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment, *Classical and Contagious Discrete Distributions*, G. P Patil (editor), 320–332. Calcutta: Statistical Publishing Society; Oxford: Pergamon. (Republished *Sankhyā*, **A27**, 1965, 311–324.) [2.4.1, 3.12.4, 4.8, 9.2]

Rao, C. R. (1971). Some comments on the logarithmic series distribution in the analysis of insect trap data, *Statistical Ecology*, Vol. 1: *Spatial Patterns and Statistical Distributions*, G. P. Patil, E. C. Pielou, and W. E. Waters (editors), 131–142. University Park: Pennsylvania State University Press. [7.1.1, 7.1.7]

Rao, C. R. (1985). Weighted distributions arising out of methods of ascertainment: What populations does a sample represent?, *A Celebration of Statistics: ISI Centenary Volume*, A. C. Atkinson and S. E. Fienberg (editors), 543–569. New York: Springer-Verlag. [3.12.4]

Rao, C. R., and Rubin, H. (1964). On a characterization of the Poisson distribution, *Sankhyā, Series A*, **26**, 295–298. [3.9, 4.8, 4.10.1, 9.2]

Rao, C. R., and Shanbhag, D. N. (1994). *Choquet-Deny Type Functional Equations with Applications to Stochastic Models*, Chichester: Wiley. [10.5, 10.8.1]

Rao, C. R., and Shanbhag, D. N. (1998). Recent approaches to characterizations based on order statistics and record values, *Handbook of Statistics*, Vol. 16, N. Balakrishnan and C. R. Rao (editors), 231–256. Amsterdam: North-Holland. [10.5]

Rao, C. R., and Srivastava, R. C. (1979). Some characterizations based on a multivariate splitting model, *Sankhyā, Series A*, **41**, 124–128. [4.8]

Rao, C. R., Srivastava, R. C., Talwalker, S., and Edgar, G. A. (1980). Characterization of probability distributions based on a generalized Rao-Rubin condition, *Sankhyā, Series A*, **42**, 161–169. [4.8, 9.2]

Rao, M. B., and Shanbhag, D. N. (1982). Damage models, *Encyclopedia of Statistical Sciences*, Vol. 2, S. Kotz, N. L. Johnson, and C. B. Read (editors), 262–265. New York: Wiley. [9.2]

Ratcliffe, J. F. (1964). The significance of the difference between two Poisson variables, *Applied Statistics*, **13**, 84–86. [4.12.3]

Ratnaparkhi, M. V. (1981). On splitting model and related characterizations of some statistical distributions, *Statistical Distributions in Scientific Work*, Vol. 4: *Models, Structures, and Characterizations*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 349–355. Dordrecht: Reidel. [4.8]

Rawlings, D. (1997). Absorption processes: Models for q-identities, *Advances in Applied Mathematics*, **18**, 133–148. [10.8.1, 10.8.2]

Rawlings, D. (2002). A joint absorption-Mahonian process, *Journal of Statistical Planning and Inference*, **101**, 281–289. [10.8.1]

Ray, S. K., and Sahai A. (1978). On variance of the MVU inverse binomial estimator, *Bulletin of the Calcutta Statistical Association*, **27**, 105–108. [5.8.2]

Rayner, J. C. W., and Best, D. J. (1989). *Smooth Tests of Goodness of Fit*, New York: Oxford University Press. [1.2.16, 3.8.4, 4.7.4, 5.8.5]

Reid, D. D. (1981). The Poisson lognormal distribution and its use as a model of plankton aggregation, *Statistical Distributions in Scientific Work*, Vol. 6: *Applications in Physical, Social, and Life Sciences*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 303–316. Dordrecht: Reidel. [11.1.4]

Relles, D. (1972). Simple algorithm for generating binomial random variables when N is large, *Journal of the American Statistical Association*, **67**, 612–613. [3.7.2]

Rényi, A. (1964). On an extremal property of the Poisson process, *Annals of the Institute of Statistical Mathematics, Tokyo*, **16**, 129–133. [4.2.1]

Resnick, S. I. (1973). Extremal processes and record value times, *Journal of Applied Probability*, **10**, 864–868. [10.5]

Richards, P. I. (1968). A generating function (Problem 67–18), *SIAM Review*, **10**, 455–456. [10.4.1]

Rider, P. R. (1953). Truncated Poisson distributions, *Journal of the American Statistical Association*, **48**, 826–830. [4.10.1]

Rider, P. R. (1962a). Estimating the parameters of mixed Poisson, binomial and Weibull distributions, *Bulletin of the International Statistical Institute*, **39**(2), 225–232. [8.2.5, 8.2.6, 8.2.7]

Rider, P. R. (1962b). The negative binomial distribution and the incomplete beta function, *American Mathematical Monthly*, **69**, 302–304. [5.11]

Ridout, M., Hinde, J., and Demétrio, C. G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives, *Biometrics*, **57**, 219–223. [5.8.5]

Riordan, J. (1937). Moment recurrence relations for binomial, Poisson, and hypergeometric frequency distributions, *Annals of Mathematical Statistics*, **8**, 103–111. [4.3]

Riordan, J. (1958). *An Introduction to Combinatorial Analysis*, New York: Wiley. [1.1.3, 9.1, 10.2, 10.4.1]

Ripley, B. D. (1981). *Spatial Statistics*, New York: Wiley. [4.9]

Ripley, B. D. (1987). *Stochastic Simulation*, New York: Wiley. [1.2.17]

Robbins, H. (1956). An empirical Bayes approach to statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 157–163. Berkeley: University of California Press. [4.7.2]

Robinson, G. K. (1982). Confidence intervals and regions, *Encyclopedia of Statistical Sciences*, Vol. 2, S. Kotz, N. L. Johnson, and C. B. Read (editors), 120–127. New York: Wiley. [1.2.16]

Robson, D. S., and Regier, H. A. (1964). Estimation of population number and mortality rates, *Methods for Assessment of Fish Production in Fresh Waters*, IBP Handbook No. 3, W. E. Ricker (editor), 124–158. Oxford: Blackwell Scientific. [6.7.1]

Rodrigues, J. (2003). Bayesian analysis of zero-inflated distributions, *Communications in Statistics—Theory and Methods*, **32**, 281–289. [8.2.4]

Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A. J., and Olmo Jiménez, M. J. (2003). Estimation of parameters in Gaussian hypergeometric distributions, *Communications in Statistics—Theory and Methods*, **32**, 1101–1118. [6.7]

Rogers, A. (1965). A stochastic analysis of the spatial clustering of retail establishments, *Journal of the American Statistical Association*, **60**, 1094–1103. [9.6.5]

Rogers, A. (1969). Quadrat analysis of urban dispersion: 2 Case studies of urban retail systems, *Environment and Planning*, **1**, 155–171. [9.6.5]

Roman, S. M., and Rota, G. C. (1978). The umbral calculus, *Advances in Mathematics*, **27**, 95–188. [2.6]

Romani, J. (1956). Distribución de la suma algebraíca de variables de Poisson, *Trabajos de Estadística*, **7**, 175–181. [4.12.3]

Romanovsky, V. (1923). Note on the moments of the binomial $(p+q)^N$ about its mean, *Biometrika*, **15**, 410–412. (See also *Les Principes de la Statistique Mathematique*, 1933, 39–40 and 320–321.). [3.3]

Romanovsky, V. (1925). On the moments of the hypergeometrical series, *Biometrika*, **17**, 57–60. [6.2.1]

Romanowska, M. (1978). Poisson approximation of some probability distributions, *Bulletin de l'Academie Polonaise des Sciences, Series des Sciences Mathematiques, Astronomiques et Physiques*, **26**, 1023–1026. [3.6.1]

Romig, H. G. (1953). *Binomial Tables*, New York: Wiley. [3.7.1]

Roos, B. (2001). Sharp constants in the Poisson approximation, *Statistics and Probability Letters*, **52**, 155–168. [4.2.2]

Roos, B. (2003). Improvements in the Poisson approximation of mixed Poisson distributions, *Journal of Statistical Planning and Inference*, **113**, 467–483. [4.2.2]

Ross, G. J. S. (1980). *MLP: Maximum Likelihood Program* [a manual], Harpenden: Rothamsted Experimental Station Statistics Department. [5.8.3, 11.2.3]

Ross, G. J. S. (1990). *Nonlinear Estimation*, New York: Springer-Verlag. [1.2.16]

Ross, G. J. S., and Preece, D. A. (1985). The negative binomial distribution, *Statistician*, **34**, 323–335. [5.8.3]

Rowe, J. A. (1942). Mosquito light trap catches from ten American cities, 1940, *Iowa State College Journal of Science*, **16**, 487–518. [7.1.9]

Roy, D. (2003). The discrete normal distribution, *Communications in Statistics—Theory and Methods*, **32**, 1871–1883. [10.8.3, 11.2.6]

Roy, D., and Gupta, R. P. (1999). Characterizations and model selections through reliability measures in the discrete case, *Statistics and Probability Letters*, **43**, 197–206. [5.12.2, 11.2.15]

Roy, J., and Mitra, S. K. (1957). Unbiased minimum variance estimation in a class of discrete distributions, *Sankhyā*, **18**, 371–378. [2.2.1, 5.8.2]

Rubinovitch, M. (1985). The slow server problem: A queue with stalling, *Journal of Applied Probablility*, **22**, 879–892. [3.12.2, 9.12]

Rudolfer, S. M. (1990). A Markov chain model of extrabinomial variation, *Biometrika*, **77**, 255–264. [3.12.6]

Rueda, R., and O'Reilly, F. (1999). Tests of fit for discrete distributions based on the probability generating function, *Communications in Statistics—Simulation and Computation*, **28**, 259–274. [4.7.4]

Rushton, S., and Lang, E. D. (1954). Tables of the confluent hypergeometric function, *Sankhyā*, **13**, 369–411. [1.1.7]

Russell, K. G. (1978). Estimation of the parameters of the Thomas distribution, *Biometrics*, **34**, 95–99. [9.10]

Rutemiller, H. C. (1967). Estimation of the probability of zero failures in m binomial trials, *Journal of the American Statistical Association*, **62**, 272–277. [3.8.2]

Rutherford, E., Chadwick, J., and Ellis, C. D. (1930). *Radiation from Radioactive Substances*, London: Cambridge University Press. [4.2.1]

Rutherford, E., and Geiger, H. (1910). The probability variations in the distribution of α particles, *Philosophical Magazine, 6th Series*, **20**, 698–704. [4.2.1]

Rutherford, R. S. C. (1954). On a contagious distribution, *Annals of Mathematical Statistics*, **25**, 703–713. [11.2.19]

Ryan, F. J. (1952). Contribution to discussion on "The statistical theory of bacterial populations subject to mutation," *Journal of the Royal Statistical Society, Series B*, **14**, 39–40. [11.2.11]

Sadooghi-Alvandi, S. M. (1990). Estimation of the parameter of a Poisson distribution using a LINEX loss function, *Australian Journal of Statistics*, **32**, 393–398. [4.7.2]

Sahai, A., and Buhrman, J. M. (1979). Bounds for the variance of an inverse binomial estimator, *Statistica Neerlandica*, **33**, 213–216. [5.8.2]

Sahai, H., and Khurshid, A. (1993a). Confidence intervals for the ratio of two Poisson means, *Mathematical Scientist*, **18**, 43–50. [4.7.3]

Sahai, H., and Khurshid, A. (1993b). Confidence intervals for the mean of a Poisson distribution: A review, *Biometrical Journal*, **7**, 857–876. [4.7.3]

Said, A. S. (1958). Some properties of the Poisson distribution, *Journal of the American Institute of Chemical Engineering*, **4**, 290–292. [4.4]

Saleh, A. K. M., and Rahim, M. A. (1972). Distribution of the sum of variates from truncated discrete populations, *Canadian Mathematical Bulletin*, **15**, 395–398. [5.11]

Saleh, A. K. M. E. (1981). Decomposition of finite mixture of distributions by minimum chi-square method, *Aligarh Journal of Statistics*, **1**, 86–97. [8.2.5]

Salvia, A. A., and Bollinger, R. C. (1982). On discrete hazard functions, *IEEE Transactions on Reliability*, **31**, 458–459. [5.12.2, 6.10.4, 6.11.1, 11.2.13, 11.2.15]

Samaniego, F. J. (1976). A characterization of convoluted Poisson distributions with applications to estimation, *Journal of the American Statistical Association*, **71**, 475–479. [4.8, 4.12.3, 9.12]

Sambursky, S. (1956). On the possible and probable in ancient Greece, *Osiris*, **12**, 35–48. [11.2.19]

Samiuddin, M., and Mallick, S. A. (1970). On the logit transformation of binomial variate, *Journal of Natural Science and Mathematics*, **10**, 59–63. [3.6.1]

Sampford, M. R. (1955). The truncated negative binomial distribution, *Biometrika*, **42**, 58–69. [5.11]

Samuels, S. M. (1965). On the number of successes in independent trials, *Annals of Mathematical Statistics*, **36**, 1272–1278. [4.5]

Sandelius, M. (1952). A confidence interval for the smallest proportion of a binomial population, *Journal of the Royal Statistical Society, Series B*, **14**, 115–117. [3.8.2]

Sandiford, P. J. (1960). A new binomial approximation for use in sampling from finite populations, *Journal of the American Statistical Association*, **55**, 718–722. [6.5]

Sandland, R. (1974). A note on some applications of the truncated geometric distribution, *Australian Journal of Statistics*, **16**, 57–58. [5.2]

Sandland, R. L., and Cormack, R. M. (1984). Statistical inference for Poisson and multinomial models for capture-recapture experiments, *Biometrika*, **71**, 27–33. [4.9]

Sankaran M. (1968). Mixtures by the inverse Gaussian distribution, *Sankhyā, Series B*, **30**, 455–458. [11.1.5]

Sankaran, M. (1970). The discrete Poisson-Lindley distribution, *Biometrics*, **26**, 145–149. [8.3.3]

Santos Silva, J. M. C. (1997). Generalized Poisson regression for positive count data, *Communications in Statistics—Simulation and Computation*, **26**, 1089–1102. [11.1.1]

Sapatinas, T. (1995). Identifiability of mixtures of power series distributions and related characterizations, *Annals of the Institute of Statistical Mathematics, Tokyo*, **47**, 447–459. [8.2.2, 8.3.2]

Sapatinas, T. (1996). On the generalized Rao-Rubin condition and some variants, *Australian Journal of Statistics*, **38**, 299–306. [4.8]

Sapatinas, T. (1999). A characterization of the negative binomial distribution via α-monotonicity, *Statistics and Probability Letters*, **45**, 49–53. [5.9.2]

Sarkadi, K. (1957a). Generalized hypergeometric distributions, *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közlenényei*, **2**, 59–68. [6.2.2, 6.2.5]

Sarkadi, K. (1957b). On the distribution of the number of exceedances, *Annals of Mathematical Statistics*, **28**, 1021–1022. [6.2.2]

Sasieni, P. (1998). Semiparametric regression, *Encyclopedia of Biostatistics*, Vol. 5, P. Armitage and T. Colton (editors), 4051–4053. Chichester: Wiley. [11.1.1]

Sathe, Y. S., and Dixit, U. J. (2001). Estimation of $P[X \leq Y]$ in the negative binomial distribution, *Journal of Statistical Planning and Inference*, **93**, 83–92. [5.8.5]

Sathe, Y. S., and Ravi, S. (1997). A characterization of negative binomial distribution, *Journal of the Indian Statistical Association*, **35**, 31–32. [5.9.2]

Sato, T. (1990). Confidence limits for the common odds ratio based on the asymptotic distribution of the Mantel-Haenszel estimator, *Biometrics*, **46**, 71–80. [6.11.1]

Satten, G. A., and Kupper, L. L. (1990). Continued fraction representation for expected cell counts of a 2×2 table: A rapid and exact method for conditional maximum likelihood estimation, *Biometrics*, **46**, 217–223. [6.11.1]

Satterthwaite, F. E. (1942). Generalized Poisson distributions, *Annals of Mathematical Statistics*, **13**, 410–417. [9.3]

Satterthwaite, F. E. (1957). Binomial and Poisson confidence limits, *Industrial Quality Control*, **13**, 56–59. [3.8.3]

Schader, M., and Schmid, F. (1985). Computation of M.L. estimates for the parameters of a negative binomial distribution from grouped data: A comparison of the scoring, Newton-Raphson and E-M algorithms, *Applied Stochastic Models and Data Analysis*, **1**, 11–23. [5.8.5]

Schader, M., and Schmid, F. (1989). Two rules of thumb for the approximation of the binomial distribution by the normal distribution, *American Statistician*, **43**, 23–24. [3.6.1]

Scheaffer, R. L. (1976). A note on approximate confidence limits for the negative binomial model, *Communications in Statistics—Theory and Methods*, **A5**, 149–158. [5.8.2]

Scheaffer, R. L., and Leavenworth, R. S. (1976). The negative binomial model for counts in units of varying size, *Journal of Quality Technology*, **8**, 158–163. [5.12.4]

Schenzle, D. (1979). Fitting the truncated negative binomial distribution without the second sample moment, *Biometrics*, **35**, 637–639. [5.11]

Schmeiser, B. W., and Kachitvichyanukul, V. (1981). *Poisson Random Variate Generation*, Research Memorandum 81–4, West Lafayette, IN: Purdue University School of Industrial Engineering. [4.6.2]

Schneider, B. E. (1978). Algorithm AS121: Trigamma function, *Applied Statistics*, **27**, 97–99. [11.2.2]

Schreider, Y. A. (1967). On the possibility of a theoretical model of statistical laws for texts (towards a basis for Zipf's law), *Problemy Peredachi Informatsii*, **3**, 57–63 (in Russian). [11.2.20]

Scollnik, D. P. M. (1995). Bayesian analysis of an intervened Poisson distribution, *Communications in Statistics—Theory and Methods*, **24**, 735–754. [4.12.7]

Scollnik, D. P. M. (1997). Inference concerning the size of the zero class from an incomplete Poisson sample, *Communications in Statistics—Theory and Methods*, **26**, 221–236. [4.7.2]

Seal, H. L. (1947). A probability distribution of deaths at age x when policies are counted instead of lives, *Skandinavisk Aktuarietidskrift*, **30**, 18–43. [11.2.20]

Seal, H. L. (1949a). Mortality data and the binomial probability law, *Skandinavisk Aktuarietidskrift*, **32**, 188–216. [3.10]

Seal, H. L. (1949b). The historical development of the use of generating functions in probability theory, *Bulletin de l'Association des Actuaires Suisses*, **49**, 209–228. [1.2.11]

Seal, H. L. (1952). The maximum likelihood fitting of the discrete Pareto law, *Journal of the Institute of Actuaries*, **78**, 115–121. [11.2.20]

Seber, G. A. F. (1982a). Capture-recapture methods, *Encyclopedia of Statistical Sciences*, Vol. 1, S. Kotz, N. L. Johnson, and C. B. Read (editors), 367–374. New York: Wiley. [6.9.1]

Seber, G. A. F. (1982b). *The Estimation of Animal Abundance* (second edition), London: Griffin. [3.10, 4.9, 5.2, 6.7.1]

Seeber, G. U. H. (1998). Poisson regression, *Encyclopedia of Biostatistics*, Vol. 4, P. Armitage and T. Colton (editors), 3404–3412. Chichester: Wiley. [11.1.1]

Seneta, E. (1988). Degree, iteration and permutation in improving Bonferroni-type bounds, *Australian Journal of Statistics*, **30A**, 27–38. [10.2]

Seneta, E. (1994). Carl Liebermeister's hypergeometric tails, *Historia Mathematica*, **21**, 453–462. [6.11.1]

Serbinowska, M. (1996). Consistency of an estimator of the number of changes in binomial observations, *Statistics and Probability Letters*, **29**, 337–344. [3.8.2]

Serfling, R. J. (1977). The role of the Poisson distribution in approximating system reliability of k-out-of-n structures, *Theory and Applications of Reliability: With Emphasis on Bayesian and Nonparametric Methods*, Vol. 1, C. P. Tsokas and I. N. Shimi (editors), 243–258. New York: Academic Press. [4.2.2]

Shaban, S. A. (1981). Computation of the Poisson–inverse Gaussian distribution, *Communications in Statistics—Theory and Methods*, **A10**, 1389–1399. [11.1.5]

Shaban, S. A. (1988). Poisson-lognormal distributions, *Lognormal Distributions: Theory and Applications*, E. L. Crow and K. Shimizu (editors), 195–210. New York: Dekker. [7.1.12, 11.1.4]

Shah, B. V., and Venkataraman, V. K. (1962). A note on modified Poisson distribution, *Metron*, **22**(3/4), 27–35. [4.10.3]

Shah, S. M. (1966). On estimating the parameter of a doubly truncated binomial distribution, *Journal of the American Statistical Association*, **61**, 259–263. [3.11]

Shah, S. M. (1971). The displaced negative binomial distribution, *Bulletin of the Calcutta Statistical Association*, **20**, 143–152. [5.11]

Shah, S. M., and Kabe, D. G. (1981). Characterizations of exponential, Pareto, power function, Burr and logistic distributions by order statistics, *Biometrical Journal*, **23**, 109–112. [1.2.12]

Shaked, M. (1974). On the distribution of the minimum and of the maximum of a random number of i.i.d. random variables, *Statistical Distributions in Scientific Work*, Vol. 1: *Models and Structures*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 363–380. Dordrecht: Reidel. [5.9.1]

Shaked, M. (1980). On mixtures from exponential families, *Journal of the Royal Statistical Society, Series B*, **42**, 192–198. [8.2.2, 8.3.2]

Shaked, M., and Shantikumar, J. (1994). *Stochastic Orders and Their Applications*, Boston: Academic Press. [3.12.3]

Shanbhag, D. N. (1970a). Another characteristic property of the Poisson distribution, *Proceedings of the Cambridge Philosophical Society*, **68**, 167–169. [4.8]

Shanbhag, D. N. (1970b). Characterizations for exponential and geometric distributions, *Journal of the American Statistical Association*, **65**, 1256–1259. [5.9.1]

Shanbhag, D. N. (1972). Some characterizations based on the Bhattacharya matrix, *Journal of Applied Probability*, **9**, 580–587. [4.8]

Shanbhag, D. N. (1973). Comments on Wang's paper, *Proceedings of the Cambridge Philosophical Society*, **73**, 473–475. [4.8]

Shanbhag, D. N. (1974). An elementary proof for the Rao-Rubin characterization of the Poisson distribution, *Journal of Applied Probability*, **11**, 211–215. [4.8, 5.9.1]

Shanbhag, D. N., and Clark, R. M. (1972). Some characterizations for the Poisson distribution starting with a power series distribution, *Proceedings of the Cambridge Philosophical Society*, **71**, 517–522. [4.8]

Shanbhag, D. N., and Kapoor, S. (1993). Some questions in characterization theory, *Mathematical Scientist*, **18**, 127–133. [10.8.1]

Shanbhag, D. N., and Panaretos, J. (1979). Some results related to the Rao-Rubin characterization of the Poisson distribution, *Australian Journal of Statistics*, **21**, 78–83. [4.8]

Shanbhag, D. N., and Rajamannar, G. (1974). Some characterizations of the bivariate distribution of independent Poisson variables, *Australian Journal of Statistics*, **16**, 119–125. [4.8]

Shane, H. D. (1973). A Fibonacci probability function, *Fibonacci Quarterly*, **11**, 517–522. [10.7.1]

Shanmugam, R. (1985). An intervened Poisson distribution and its medical application, *Biometrics*, **41**, 1025–1029. [4.12.7]

Shanmugam, R. (1986). A characterization of positive binomial distribution and its goodness-of-fit application, *Statistics*, **17**, 33–39. [6.11.1]

Shanmugam, R. (1992). An inferential procedure for the Poisson intervention parameter, *Biometrics*, **48**, 559–565. [4.12.7]

Shanmugam, R., and Singh, J. (1981). On the Stirling distribution of the first kind, *Statistical Distributions in Scientific Work*, Vol. 4: *Models, Structures, and Characterizations*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 181–187. Dordrecht: Reidel. [7.1.12]

Shanthikumar, J. G. (1982). Recursive algorithm to evaluate the reliability of a consecutive-k-out-of-n:F system, *IEEE Transactions on Reliability*, **R-31**, 442–443. [10.7.5]

Shanthikumar, J. G. (1985). Discrete random variate generation using uniformization, *European Journal of Operational Research*, **21**, 387–398. [7.1.6]

Shea, B. L. (1989). Remark AS R77—A remark on algorithm AS 152: Cumulative hypergeometric probabilities, *Applied Statistics*, **38**, 199–204. [6.6]

Shenton, L. R. (1949). On the efficiency of the method of moments and Neyman's type A distribution, *Biometrika*, **36**, 450–454. [9.6.2, 9.6.4]

Shenton, L. R., and Bowman, K. O. (1967). Remarks on large sample estimators for some discrete distributions, *Technometrics*, **9**, 587–598. [9.6.1, 9.6.4]

Shenton, L. R., and Bowman, K. O. (1977). *Maximum Likelihood Estimation in Small Samples*, London: Griffin. [9.6.4, 9.7]

Shenton, L. R., and Skees, P. (1970). Some statistical aspects of amounts and duration of rainfall, *Random Counts in Scientific Work*, Vol. 3: *Random Counts in Physical Science, Geo Science, and Business*, G. P. Patil (editor), 73–94. University Park: Pennsylvania State University Press. [7.1.12]

Sherbrooke, C. C. (1968). Discrete compound Poisson processes and tables of the geometric Poisson distribution, *Naval Research Logistics Quarterly*, **15**, 189–203. [9.7]

Sheu, S. S. (1984). The Poisson approximation to the binomial distribution, *American Statistician*, **38**, 206–207. [3.6.2]

Shimizu, R. (1968). Generalized hypergeometric distributions, *Proceedings of the Institute of Statistical Mathematics, Tokyo*, **16**, 147–165 (in Japanese). [6.2.5]

Shorrock, R. W. (1972a). A limit theorem for inter-record times, *Journal of Applied Probability*, **9**, 219–223. [10.5]

Shorrock, R. W. (1972b). On record values and record times, *Journal of Applied Probability*, **9**, 316–326. [10.5]

Shoukri, M. M., and Consul, P. C. (1987). Some chance mechanisms generating the generalized Poisson probability models, *Advances in the Statistical Sciences, Vol. 5: Biostatistics*, I. B. MacNeill and G. J. Umphrey (editors), 259–268. Dordrecht: Reidel. [7.2.6]

Shumway, R., and Gurland, J. (1960a). A fitting procedure for some generalized Poisson distributions, *Skandinavisk Aktuarietidskrift*, **43**, 87–108. [9.5, 9.8]

Shumway, R., and Gurland, J. (1960b). Fitting the Poisson binomial distribution, *Biometrics*, **16**, 522–533. [9.5, 9.8]

Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **31**, 373–390. [11.2.2]

Sibuya, M. (1983). Generalized hypergeometric distributions, *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz, N. L. Johnson, and C. B. Read (editors), 330–334. New York: Wiley. [6.2.5]

Sibuya, M. (1988). Log-concavity of Stirling numbers and unimodality of Stirling distributions, *Annals of the Institute of Statistical Mathematics, Tokyo*, **40**, 693–714. [7.1.12]

Sibuya, M., and Shimizu, R. (1981). Classification of the generalized hypergeometric family of distributions. *Keio Science and Technology Reports*, **34**, 1–38. [6.2.5]

Sichel, H. S. (1951). The estimation of the parameters of a negative binomial distribution with special reference to psychological data, *Psychometrika*, **16**, 107–127. [5.10]

Sichel, H. S. (1971). On a family of discrete distributions particularly suited to represent long-tailed frequency data, *Proceedings of the Third Symposium on Mathematical Statistics*, N. F. Laubscher (editor), 51–97. Pretoria: CSIR. [11.1.5]

Sichel, H. S. (1973a). Statistical evaluation of diamondiferous deposits, *Journal of the South African Institute of Mining and Metallurgy*, **73**, 235–243. [11.1.5]

Sichel, H. S. (1973b). The density and size distribution of diamonds, *Bulletin of the International Statistical Institute*, **45**(2), 420–427. [11.1.5]

Sichel, H. S. (1974). On a distribution representing sentence-length in written prose, *Journal of the Royal Statistical Society, Series A*, **137**, 25–34. [11.1.5]

Sichel, H. S. (1975). On a distribution law for word frequencies, *Journal of the American Statistical Association*, **70**, 542–547. [11.1.5]

Sichel, H. S. (1982a). Repeat-buying and the generalized inverse-Gaussian distribution, *Applied Statistics*, **31**, 193–204. [11.1.5]

Sichel, H. S. (1982b). Asymptotic efficiencies of three methods of estimation for the inverse Gaussian–Poisson distribution, *Biometrika*, **69**, 467–472. [11.1.5]

Silva, G. (1941). Una generalizazione del problema delle concordanze, *Instituto Veneto de Scienze, Lettere ed Arti, Venezia Classe de Scienze Matematiche e Naturali*, **100**, 689–709. [10.3]

Silverman, B. W. (1985). Penalized maximum likelihood estimation, *Encyclopedia of Statistical Sciences*, Vol. 6, S. Kotz, N. L. Johnson, and C. B. Read (editors), 664–667. New York: Wiley. [1.2.16]

Sim, C. H., and Lee, P. A. (1989). Simulation of negative binomial processes, *Journal of Statistical Computation and Simulation*, **34**, 29–42. [5.7]

Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process, *Annals of Statistics*, **4**, 1200–1209. [8.2.5]

Simon, H. A. (1955). On a class of skew distribution functions, *Biometrika*, **42**, 425–440. [6.10.3, 11.2.21]

Simon, H. A. (1960). Some further notes on a class of skew distribution functions, *Information and Control*, **3**, 80–88. [6.10.3]

Simons, G., and Johnson, N. L. (1971). On the convergence of binomial to Poisson distributions, *Annals of Mathematical Statistics*, **42**, 1735–1736. [3.6.1, 4.2.2]

Singh, H., and Vasudeva, H. L. (1984). A characterization of exponential distribution by Yule distribution, *Journal of the Indian Statistical Association*, **22**, 93–96. [6.10.3, 6.10.4]

Singh, J. (1978). A characterization of positive Poisson distribution and its statistical application, *SIAM Journal of Applied Mathematics*, **34**, 545–548. [4.8]

Singh, S. N. (1963). A note on inflated Poisson distribution, *Journal of the Indian Statistical Association*, **1**, 140–144. [8.2.4]

Singh, S. N. (1964). On the time of first birth, *Sankhyā, Series B*, **26**, 95–102. [11.2.9]

Singh, S. N. (1968). A chance mechanism of the variation in the number of births per couple, *Journal of the American Statistical Association*, **63**, 209–213. [11.2.9]

Singh, S. N., Bhattacharya, B. N., and Yadava, R. C. (1974). A parity dependent model for number of births and its applications, *Sankhyā, Series B*, **36**, 93–102. [11.2.9]

Sinha, B. K., and Bose, A. (1985). Unbiased sequential estimation of $1/p$: Settlement of a conjecture, *Annals of the Institute of Statistical Mathematics, Tokyo*, **37**, 455–460. [3.8.2]

Sinha, B. K., and Sinha, B. K. (1975). Some problems of unbiased sequential binomial estimation, *Annals of the Institute of Statistical Mathematics, Tokyo*, **27**, 245–258. [3.8.2]

Siotani, M. (1956). Order statistics for discrete case with a numerical application to the binomial distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **8**, 95–104. [3.5]

Siotani, M., and Ozawa, M. (1948). Tables for testing the homogeneity of k independent binomial experiments on a certain event based on the range, *Annals of the Institute of Statistical Mathematics, Tokyo*, **10**, 47–63. [3.5]

Siromoney, G. (1962). Entropy of logarithmic series distributions, *Sankhyā, Series A*, **24**, 419–420. [7.1.4]

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations, *Journal of the Royal Statistical Society, Series A*, **109**, 296. [4.12.3]

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials, *Journal of the Royal Statistical Society, Series B*, **10**, 257–261. [5.3, 6.7.2, 6.9.2]

Skellam, J. G. (1949). The probability distribution of gene-frequencies in relation to selection, mutation, and random extinction, *Proceedings of the Cambridge Philosophical Society*, **45**, 364–367. [11.2.16]

Skellam, J. G. (1952). Studies in statistical ecology I: Spatial pattern, *Biometrika*, **39**, 346–362. [9.4, 9.5, 9.8]

Skellam, J. G. (1958). On the derivation and applicability of Neyman's type A distribution, *Biometrika*, **45**, 32–36. [9.6.5]

Skibinsky, M. (1970). A characterization of hypergeometric distributions, *Journal of the American Statistical Association*, **65**, 926–929. [6.8]

Slater, L. J. (1960). *Confluent Hypergeometric Functions*, London: Cambridge University Press (erratum *Mathematics of Computation*, 1963, **17**, 486–487). [1.1.7, 1.1.8]

Slater, L. J. (1966). *Generalized Hypergeometric Functions*, London: Cambridge University Press. [1.1.6, 1.1.12, 2.9]

Slud, E. V. (1977). Distribution inequalities for the binomial law, *Annals of Probability*, **5**, 404–412. [3.6.2]

Smith, A. F. M. (1984). Present position and potential developments: Some personal views: Bayesian statistics, *Journal of the Royal Statistical Society, Series A*, **147**, 245–259. [1.2.2]

Smith, A. F. M. (1985). An overview of some problems relating to finite mixtures, *Rassegna di Metodi Statistici ed Applicazione 5 Cagliari*, 138–149. [8.2.1, 8.2.2]

Smith, L., and Diaconis, P. (1988). Honest Bernoulli excursions, *Journal of Applied Probability*, **25**, 464–477. [6.10.4]

Smith, P. J. (1995). A recursive formulation of the old problem of obtaining moments from cumulants and vice versa, *American Statistician*, **49**, 217–218. [1.2.8]

Sobel, M., and Huyett, M. J. (1957). Selecting the best one of several binomial populations, *Bell System Technical Journal*, **36**, 537–576. [3.5]

Society of Actuaries (1954). *Impairment Study, 1951*, London: Society of Actuaries. [4.7.3]

Solow, R. M. (1960). On a family of lag distributions, *Econometrica*, **28**, 392–406. [5.10]

Somerville, P. N. (1957). Optimum sampling in binomial populations, *Journal of the American Statistical Association*, **52**, 494–502. [3.5]

Song, R., Schlecht, P. C., and Groff, J. H. (2000). A composite binomial model derived from correlated random variables, *Communications in Statistics—Theory and Methods*, **29**, 1179–1195. [3.12.6]

Spinelli, J. J. (2001). Testing for fit for the grouped exponential distribution, *Canadian Journal of Statistics*, **29**, 451–458. [5.2]

Spinelli, J. J., and Stephens, M. A. (1997). Cramer–von Mises tests of fit for the Poisson distribution, *Canadian Journal of Statistics*, **25**, 257–268. [4.7.4]

Springer, M. D. (1979). *The Algebra of Random Variables*, New York: Wiley. [1.1.10, 3.4]

Sprott, D. A. (1957). Probability distributions associated with distinct hits on targets, *Bulletin of Mathematical Biophysics*, **19**, 163–170. [10.4.1]

Sprott, D. A. (1958). The method of maximum likelihood applied to the Poisson binomial distribution, *Biometrics*, **14**, 97–106. [9.4, 9.5]

Sprott, D. A. (1965). A class of contagious distributions and maximum likelihood estimation, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 337–350. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [9.2]

Sreehari, M., and Kalamkar, V. A. (1997). Modelling some stationary Markov processes and related characterizations, *Journal of Statistical Planning and Inference*, **63**, 363–375. [4.8]

Srivastava, H. M., and Kashyap, B. R. K. (1982). *Special Functions in Queueing Theory and Related Stochastic Processes*, New York: Academic Press. [2.4.1, 6.10.1]

Srivastava, M. S. (1965). Characterization theorems for some distributions, *Annals of Mathematical Statistics*, **36**, 361 (abstract). [5.9.1]

Srivastava, R. C. (1971). On a characterization of the Poisson process, *Journal of Applied Probability*, **8**, 615–616. [4.8]

Srivastava, R. C. (1974). Two characterizations of the geometric distribution, *Journal of the American Statistical Association*, **69**, 267–269. [5.9.1]

Srivastava, R. C. (1979). Two characterizations of the geometric distribution by record values, *Sankhyā, Series B*, **40**, 276–279. [5.9.1, 10.5]

Srivastava, R. C. (1981). On some characterizations of the geometric distribution, *Statistical Distributions in Scientific Work*, Vol. 4, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 411–416. Dordrecht: Reidel. [10.5]

Srivastava, R. C., and Singh, J. (1975). On some characterizations of the binomial and Poisson distributions based on a damage model, *Statistical Distributions in Scientific Work*, Vol. 3: *Characterizations and Applications*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 271–277. Dordrecht: Reidel. [4.8]

Srivastava, R. C., and Srivastava, A. B. L. (1970). On a characterization of the Poisson distribution, *Journal of Applied Probability*, **7**, 497–501. [4.8]

Sródka, T. (1963). On approximation of hypergeometric distribution, *Zeszyty Naukowe Politechniki Lódzkiej*, **10**(53), 5–17. [6.5]

Stadlober, E. (1991). Binomial random variate generation: A method based on ratio of uniforms, *The Frontiers of Statistical Computation, Simulation, and Modeling*, P. R. Nelson, E. J. Dudewicz, A. Öztürk, and E. C. van der Meulen (editors), 93–112. Columbus, OH: American Sciences Press. [3.7.2]

Staff, P. J. (1964). The displaced Poisson distribution, *Australian Journal of Statistics*, **6**, 12–20. [4.12.4]

Staff, P. J. (1967). The displaced Poisson distribution—region 3, *Journal of the American Statistical Association*, **62**, 643–654. [4.12.4]

Stancu, D. D. (1968). On the moments of negative order of the positive Bernoulli and Poisson variables, *Studia Universitatis Babes-Bolyai, Series Math-Phys*, **13**, 27–31. [3.3]

Stanley, R. P. (1999). *Enumerative Combinatorics*, Vol. 2, Cambridge: Cambridge University Press. [6.10.5]

Steck, G. P. (1973). Bounding Poisson approximations to the binomial, *Communications in Statistics*, **2**, 189–203. [3.6.1]

Steck, G. P., and Zimmer, W. J. (1968). The relationship between Neyman and Bayes confidence intervals for the hypergeometric parameter, *Technometrics*, **10**, 199–203. [6.7.1, 6.9.2]

Stein, C. (1986). *Approximate Computation of Expectations*, Hayward, CA: Institute of Mathematical Statistics. [4.8]

Stein, C. M. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *Proceedings of the Sixth Berkeley Symposium*, **2**, 583–602. [4.2.2]

Stein, G., Zucchini, W., and Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension, *Journal of the American Statistical Association*, **82**, 938–944. [11.1.5]

Stein, W. E., and Dattero, R. (1984). A new discrete Weibull distribution, *IEEE Transactions on Reliability*, **33**, 196–197. [11.2.15]

Stepanov, A. V. (1992). Limit theorems for weak records, *Theory of Probability and Its Applications*, **37**, 570–573. [10.5]

Stephan, F. F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate, *Annals of Mathematical Statistics*, **16**, 50–61. [3.11]

Stephens, M. A. (1986). Tests based on EDF statistics, *Goodness-of-Fit Techniques*, R. B. D'Agostino and M. A. Stephens (editors), 97–193. New York: Dekker. [4.7.4]

Steutel, F. W. (1970). *Preservation of Infinite Divisibility under Mixing*, Tract 33. Amsterdam: Mathematical Centre. [7.1.4, 9.3]

Steutel, F. W. (1983). Infinite divisibility, *Encyclopedia of Statistical Sciences*, Vol. 4, S. Kotz, N. L. Johnson, and C. B. Read (editors), 114–116. New York: Wiley. [9.3]

Steutel, F. W. (1988). Note on discrete α-unimodality, *Statistica Neerlandica*, **42**, 137–140. [1.2.7]

Steutel, F. W. (1990). *The Set of Geometrically Infinitely Divisible Distributions*, Memorandum COSOR 90–42, Eindhoven: Technische Universiteit Eindhoven. [9.3]

Steutel, F. W., and Thiemann, J. G. F. (1989a). On the independence of integer and fractional parts, *Statistica Neerlandica*, **43**, 53–59. [5.2]

Steutel, F. W., and Thiemann, J. G. F. (1989b). The gamma process and the Poisson distribution, *Colloquia Mathematica Societatis János Bolyai* 57, *Limit Theorems in Probability and Statistics*, Pecs, Hungary, 477–490. [4.4]

Steutel, F. W., and van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *Annals of Probability*, **7**, 893–899. [11.2.5]

Stevens, W. L. (1937). Significance of grouping, *Annals of Eugenics, London*, **8**, 57–60. [10.1, 10.4.1]

Stevens, W. L. (1939). Distribution of groups in a sequence of alternatives, *Annals of Eugenics, London*, **9**, 10–17. [10.1, 10.3, 10.6.1]

Stevens, W. L. (1950). Fiducial limits for the parameter of a discontinuous distribution, *Biometrika*, **37**, 117–129. [3.8.3]

Stevens, W. L. (1951). Mean and variance of an entry in a contingency table, *Biometrika*, **38**, 468–470. [3.4]

Stevens, W. L. (1957). Shorter intervals for the parameter of the binomial and Poisson distributions, *Biometrika*, **44**, 436–440. [4.7.3]

Steyn, H. S. (1956). On the univariable series $F(t) = F(a; b_1, b_2, \ldots, b_k; t, t^2, \ldots, t^k)$ and its applications in probability theory, *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen, Series A*, **59**, 190–197. [11.2.18]

Steyn, H. S., Jr. (1980). A class of multiparameter power series distributions, *South African Statistical Journal*, **14**, 1–15. [2.2.2, 11.2.17]

Steyn, H. S., Jr. (1984). Two-parameter power series distributions with applications, *South African Statistical Journal*, **18**, 29–44. [2.2.2, 3.12.2, 11.2.17]

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA: Harvard University Press. [3.2]

Stoyan, D., Kendall, W. S., and Mecke J. (1987). *Stochastic Geometry and Its Applications*, Chichester: Wiley. [4.9]

Strackee, J., and van der Gon, J. J. D. (1962). The frequency distribution of the difference between two Poisson variates, *Statistica Neerlandica*, **16**, 17–23. [4.12.3]

Streit, F. (1987). Testing for effective intervention, *Biometrics*, **43**, 999–1000. [4.12.7]

Stuart, A. (1963). Standard errors for percentages, *Applied Statistics*, **12**, 87–101. [3.7.1]

Stuart, A., and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics* (fifth edition), Vol. 1, London: Griffin. [1.1.1, 1.1.7, 1.1.11, 1.2.7, 1.2.8, 1.2.15, 1.2.16, 3.12.3, 4.4, 7.1.2, 9.1]

Student (1907). On the error of counting with a haemocytometer, *Biometrika*, **5**, 351–360. [4.2.1, 5.3, 5.12.5]

Student (1919). An example of deviations from Poisson's law in practice, *Biometrika*, **12**, 211–213. [3.8.2]

Subrahmaniam, Kocherlakota (1966). On a general class of contagious distributions: The Pascal-Poisson distribution, *Trabajos de Estadística*, **17**, 109–127. [5.12.6, 9.9, 9.13]

Subrahmaniam, Kathleen (1978). The Pascal-Poisson distribution revisited: Estimation and efficiency, *Communications in Statistics—Theory and Methods*, **A7**, 673–683. [5.12.6, 9.9, 9.13]

Sulanke, R. A. (1998). Catalan path statistics having the Narayana distribution, *Discrete Mathematics*, **189**, 369–389. [6.10.5]

Sulanke, R. A. (1999). Constraint sensitive Catalan path statistics having the Narayana distribution, *Discrete Mathematics*, **204**, 397–414. [6.10.5]

Sulanke, R. A. (2002). The Narayana distribution, *Journal of Statistical Planning and Inference*, **101**, 311–326. [6.10.5]

Sundt, B., and Jewell, W. S. (1981). Further results on recursive evaluation of compound distributions, *ASTIN Bulletin*, **18**, 27–39. [2.3.2, 5.12.2]

Svensson, A. (1969). *Some Remarks on the Moran-Chatterji Theorem*, Stockholm: University of Stockholm Department of Mathematical Statistics. [3.9, 4.8]

Swift, R. J. (2000). A simple immigration–catastrophe process, *Mathematical Scientist*, **25**, 32–36. [4.12.5]

Syski, R. (1988). Further comments on the solution of the M/M/1 queue, *Advances in Applied Probability*, **20**, 693. [11.2.14]

Szabłowski, P. J. (2001). Discrete normal distribution and its relationship with Jacobi Theta functions, *Statistics and Probability Letters*, **44**, 241–249. [10.8.3]

Szegö, G. (1939, 1959, 1967). *Orthogonal Polynomials (three editions)*, Providence, RI: American Mathematical Society. [1.1.11]

Taguti, G. (1952). Tables of 5% and 1% points for the Polya-Eggenberger distribution function, *Reports of Statistical Application Research, JUSE*, **2**. [5.7]

Taillie, C., and Patil, G. P. (1986). The Fibonacci distribution revisited, *Communications in Statistics—Theory and Methods*, **15**, 951–959. [10.7.1]

Takács, L. (1955). Investigation of waiting time problems by reduction to Markov processes, *Acta Mathematica Academiae Scientiarum Hungaricae*, **6**, 101–128. [11.2.10]

Takács, L. (1962). A generalization of the ballot problem and its applications in the theory of queues, *Journal of the American Statistical Association*, **57**, 327–337. [7.2]

Takács, L. (1967). On the method of inclusion and exclusion, *Journal of the American Statistical Association*, **62**, 102–113. [10.2]

Takács, L. (1980). The problem of coincidences, *Archive for the History of Exact Sciences*, **21**, 229–244. [10.3]

Tallis, G. M. (1969). The identifiability of mixtures of distributions, *Journal of Applied Probability*, **6**, 389–398. [8.3.1]

Tallis, G. M. (1983). Goodness of fit, *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz, N. L. Johnson, and C. B. Read (editors), 451–460. New York: Wiley. [4.7.4]

Tallis, G. M., and Chesson, P. (1982). Identifiability of mixtures, *Journal of the Australian Mathematical Society*, **A32**, 339–348. [8.3.1]

Talwalker, S. (1970). A characterization of the double Poisson distribution, *Sankhyā, Series A*, **32**, 265–270. [4.8]

Talwalker, S. (1975). Models in medicine and toxicology, *Statistical Distributions in Scientific Work*, Vol. 2: *Model Building and Model Selection*, G. P. Patil, S. Kotz, and J. K. Ord (editors), 263–274. Dordrecht: Reidel. [4.8]

Talwalker, S. (1980). A note on the generalized Rao-Rubin condition and characterization of certain discrete distributions, *Journal of Applied Probability*, **17**, 563–569. [4.8]

Talwalker, S. (1986). Functional equations in characterizations of discrete distributions by Rao-Rubin condition and its variants, *Communications in Statistics—Theory and Methods*, **15**, 961–979. [9.2]

Tanner, J. C. (1953). A problem of interference between two queues, *Biometrika*, **40**, 58–69. [7.2.2, 9.11, 11.2.14]

Tate, R. F., and Goen, R. L. (1958). Minimum variance unbiased estimation for the truncated Poisson distribution, *Annals of Mathematical Statistics*, **29**, 755–765. [2.2.1, 4.10.1, 4.12.3]

Taylor, H. M., and Karlin, S. (1998). *An Introduction to Stochastic Modeling* (third edition), San Diego: Academic Press. [4.9, 5.2, 5.10]

Taylor, L. R. (1961). Aggregation, variance and the mean, *Nature*, **189**, 732–735. [5.8.4]

Teich, M. C., and McGill, W. J. (1976). Neural counting and photon counting in the presence of dead time, *Physical Review Letters*, **36**, 754–758. [5.12.5]

Teicher, H. (1955). An inequality on Poisson probabilities, *Annals of Mathematical Statistics*, **26**, 147–149. [4.4, 4.5]

Teicher, H. (1960). On the mixture of distributions, *Annals of Mathematical Statistics*, **31**, 55–73. [8.2.2, 8.2.5, 8.3.2]

Teicher, H. (1961). Identifiability of mixtures, *Annals of Mathematical Statistics*, **32**, 244–248. [8.2.2, 8.2.6, 8.3.1]

Teicher, H. (1963). Identifiability of finite mixtures, *Annals of Mathematical Statistics*, **34**, 1265–1269. [8.2.2, 8.2.6]

Thiele, T. N. (1889). *Theory of Observations* (reprinted in English in *Annals of Mathematical Statistics*, 1931, **2**, 165–308). [4.2.1]

Thomas, D. G., and Gart, J. J. (1971). Small sample performance of some estimators of the truncated binomial distribution, *Journal of the American Statistical Association*, **66**, 169–177. [3.11]

Thomas, H. A. (1948). Frequency of minor floods, *Journal of the Boston Society of Civil Engineering*, **35**, 425. (Graduate School of Engineering Publication No. 466, Harvard University, Cambridge, MA.) [6.2.2]

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology, *Biometrika*, **36**, 18–25. [9.10]

Thomas, M. A., and Taub, A. E. (1975). *Binomial trials with variable probabilities*, Technical Note DK-25/75, Dahlgren, VA: Naval Surface Weapons Center. [3.12.3]

Thomas, M. A., and Taub, A. E. (1982). Calculating binomial probabilities when the trial probabilities are unequal, *Journal of Statistical Computation and Simulation*, **14**, 125–131. [3.12.3]

Thompson, H. R. (1954). A note on contagious distributions, *Biometrika*, **41**, 268–271. [5.3, 9.3]

Thurston, S. W., Wand, M. P., and Wiencke, J. K. (2000). Negative binomial additive models, *Biometrics*, **56**, 139–144. [5.10]

Thyrion, P. (1960). Note sur les distributions "par grappes," *Association Royale des Actuaires Belges Bulletin*, **60**, 49–66. [4.11, 5.3, 9.3]

Tiago de Oliviera, J. (1965). Some elementary tests for mixtures of discrete distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 379–384. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [8.2.5]

Tiku, M. L. (1964). A note on the negative moments of a truncated Poisson variate, *Journal of the American Statistical Association*, **59**, 1220–1224. [4.10.1]

Tiku, M. L. (1989). Modified maximum likelihood estimation, *Encyclopedia of Statistical Sciences, Supplement*, S. Kotz, N. L. Johnson, and C. B. Read (editors), 98–103. New York: Wiley. [1.2.16]

Tippett, L. H. C. (1932). A modified method of counting particles, *Proceedings of the Royal Society of London, Series A*, **137**, 434–446. [4.10.2]

Titterington, D. M. (1990). Some recent research in the analysis of mixture distributions, *Statistics*, **21**, 619–641. [8.2.1, 8.2.2]

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Chichester: Wiley. [8.2.1, 8.2.2, 8.2.5]

Tiwari, R. C., Tripathi, R. C., and Gupta, R. C. (1992). Some weighted distributions of order *k*, *Communications in Statistics—Theory and Methods*, **21**, 411–422. [10.7.4]

Todhunter, I. (1865). *A History of the Mathematical Theory of Probability*, Cambridge: Macmillan (reprinted 1965, New York: Chelsea). [5.3, 6.2.1, 6.2.2, 10.7.1]

Tripathi, R. C. (1983). Kemp families of distributions, *Encyclopedia of Statistical Sciences*, Vol. 4, S. Kotz, N. L. Johnson, and C. B. Read (editors), 360–366. New York: Wiley. [6.11.1]

Tripathi, R. C., Gupta, P. L., and Gupta, R. C. (1986). Incomplete moments of modified power series distributions with applications, *Communications in Statistics—Theory and Methods*, **15**, 999–1015. [2.2.2, 7.2.8]

Tripathi, R. C., and Gupta, R. C. (1985). A generalization of the log-series distribution, *Communications in Statistics—Theory and Methods*, **14**, 1779–1799. [6.11.1, 7.1.11]

Tripathi, R. C., and Gupta, R. C. (1987). A comparison between the ordinary and the length-biased modified power series distributions with applications, *Communications in Statistics—Theory and Methods*, **16**, 1195–1206. [6.11.1]

Tripathi, R. C., and Gupta, R. C. (1988). Another generalization of the logarithmic series and the geometric distributions, *Communications in Statistics—Theory and Methods*, **17**, 1541–1547. [7.1.11]

Tripathi, R. C., Gupta, R. C., and Gurland, J. (1994). Estimation of parameters in the beta binomial model, *Annals of the Institute of Statistical Mathematics*, **46**, 317–331. [6.7.2]

Tripathi, R. C., Gupta, R. C., and White, T. J. (1987). Some generalizations of the geometric distribution, *Sankhyā, Series B*, **49**, 218–223. [6.11.1, 7.1.11]

Tripathi, R. C., and Gurland, J. (1977). A general family of discrete distributions with hypergeometric probabilities, *Journal of the Royal Statistical Society, Series B*, **39**, 349–356. [2.3.1, 3.4, 4.12.4, 6.11.1]

Tripathi, R. C., and Gurland, J. (1979). Some aspects of the Kemp families of distributions, *Communications in Statistics—Theory and Methods*, **8**, 855–869. [2.3.1, 2.4.1, 2.4.2, 4.12.4, 5.8.1, 6.11.3]

Tripathi, R. C., Gurland, J., and Bhalerao, N. R. (1986). A unified approach to estimating parameters in some generalized Poisson distributions, *Communications in Statistics—Theory and Methods*, **15**, 1017–1034. [9.7, 9.9]

Truesdell, C. (1945). On a function which occurs in the theory of the structure of polymers, *Annals of Mathematics*, **46**, 144–157. [11.2.20]

Tsui, K.-W., and Press, S. J. (1982). *Simultaneous Estimation of Several Poisson Parameters under K-Normalized Squared Error Loss*, Technical Report No. 38, Riverside, CA: University of California. [4.7.2]

Tukey, J. W. (1949). Moments of random group size distributions, *Annals of Mathematical Statistics*, **20**, 523–539. [10.4.1]

Tukey, J. W. (1977). *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. [1.2.12]

Turner, D. W., Young, D. M., and Seaman, J. W. (1995). A Kolmogorov inequality for the sum of independent Bernoulli random variables with unequal means, *Statistics and Probability Letters*, **23**, 243–245. [3.12.3]

Tweedie, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions, *Proceedings of the Cambridge Philosophical Society*, **43**, 41–49. [2.2.1, 11.1.2]

Tweedie, M. C. K. (1965). Further results concerning expectation-inversion technique, *Classical and Contagious Discrete Distributions*, G. P. Patil (editor), 195–218. Calcutta: Statistical Publishing Society; Oxford: Pergamon. [2.2.1]

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families, *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, J. K. Ghosh and J. Roy (editors), 579–604. Calcutta: Indian Statistical Institute. [11.1.2]

Tweedie, M. C. K., and Veevers, A. (1968). The inversion of cumulant operators for power series distributions, and the approximate stabilization of variance, *Journal of the American Statistical Association*, **63**, 321–328. [2.2.1]

Uhlmann, W. (1966). Vergleich der Hypergeometrischen mit der Binomial-Verteilung, *Metrika*, **10**, 145–158. [3.4, 6.5]

Uppuluri, V. R. R., and Blot, W. J. (1970). A probability distribution arising in a riff-shuffle, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, G. P. Patil (editor), 23–46. University Park: Pennsylvania State University Press. [5.12.7]

Uppuluri, V. R. R., Feder, P. I., and Shenton, L. R. (1967). Random difference equations occurring in one-compartment models, *Mathematical Biosciences*, **1**, 143–171. [5.9.1]

Uppuluri, V. R. R., and Patil, G. P. (1983). Waiting times and generalized Fibonacci sequences, *Fibonacci Quarterly*, **21**, 342–349. [10.7.1]

Uspensky, J. V. (1937). On Ch. Jordan's series for probability, *Annals of Mathematics*, **32**, 306–312. [3.6.2]

van der Vaart, H. R. (1972). A note on a functional equation for the generating function of a Poisson distribution, *Sankhyā, Series A*, **34**, 191–193. [4.8]

van der Waerden, B. L. (1960). Sampling inspection as a minimum loss problem, *Annals of Mathematical Statistics*, **31**, 369–384. [4.9]

Van de Ven, R. (1993). Estimating the shape parameter for the negative binomial distribution, *Journal of Statistical Computation and Simulation*, **46**, 111–123. [5.8.5]

Van de Ven, R., and Weber, N. C. (1993). Bounds for the median of the negative binomial distribution, *Metrika*, **40**, 185–189. [5.5]

Varian, H. R. (1975). A Bayesian approach to real estate assessment, *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, S. E. Fienberg and A. Zellner (editors), 195–208. Amsterdam: North-Holland. [4.7.2]

Vellaisamy, P. (1996). On the number of successes in dependent trials, *Communications in Statistics—Theory and Methods*, **25**, 1745–1756. [3.12.6]

Vellaisamy, P., and Chaudhuri B. (1996). Poisson and compound Poisson approximations for random sums of random variables, *Journal of Applied Probability*, **33**, 127–137. [4.2.2]

Vernon, P. E. (1936). The matching method applied to investigations of personality, *Psychological Bulletin*, **33**, 149–177. [10.1, 10.3]

Verrall, R. J. (1989). The individual risk model: A compound distribution, *Journal of the Institute of Actuaries*, **116**, 101–107. [9.13]

Vervaat, W. (1969). Upper bounds for the distance in total variation between the binomial and negative binomial and the Poisson distributions, *Statistica Neerlandica*, **23**, 79–86. [3.6.1]

Vervaat, W. (1973). Limit theorems for records from discrete distributions, *Stochastic Processes and Their Applications*, **1**, 317–334. [10.5]

Vit, I. (1974). Testing for homogeneity: The geometric distribution, *Biometrika*, **61**, 565–568. [5.2]

Vogler, L. E. (1964). *Percentage Points of the Beta Distribution*, Technical Note 215, Washington, DC: National Bureau of Standards. [1.1.5]

Vollset, S. E., Hirji, K. F., and Elashoff, R. M. (1991). Fast computation of exact confidence limits for the common odds ratio in a series of 2×2 tables, *Journal of the American Statistical Association*, **86**, 404–409. [6.11.1]

Volodin, I. N. (1965). On distinguishing between Poisson and Pólya distributions on the basis of a large number of samples, *Theory of Probability and Its Applications*, **10**, 335–338. [4.8]

Von Mises, R. (1921). Das Problem der Iterationen, *Zeitschrift für angewandte Mathematik und Mechanik*, **1**, 297–307. [10.6.1]

Von Mises, R. (1939). Über Aufteilungs und Besetzungswahrscheinlichkeiten, *Revue of the Faculty of Science, University of Istanbul, NS*, **4**, 145–163 (reprinted in *Selected Papers of R. von Mises*, Vol. 2, 313–331. Providence, RI: American Mathematical Society). [10.4.1]

Wadley, F. M. (1950). Notes on the form of distribution of insect and plant populations, *Annals of the Entomological Society of America*, **43**, 581–586. [9.6.5]

Wald, A., and Wolfowitz, J. (1940). On a test whether two samples are from the same population, *Annals of Mathematical Statistics*, **11**, 147–162. [10.6.1]

Walker, A. J. (1974). New fast method for generating discrete random numbers with arbitrary frequency distributions, *Electronics Letters*, **10**, 127–128. [1.2.17]

Walker, A. J. (1977). An efficient method for generating discrete random variables, *ACM Transactions on Mathematical Software*, **3**, 253–256. [1.2.17, 6.6]

Wallenius, K. T. (1963). *Biased Sampling: The Noncentral Hypergeometric Probability Distribution*, PhD Thesis, Stanford, CA: Stanford University. [6.11.4]

Wallis, W. A., and Moore, G. H. (1941). *A Significance Test for Time Series*, Technical Paper **1**. New York: National Bureau of Economic Research. [10.6.2]

Walsh, J. E. (1955). The Poisson distribution as a limit for dependent binomial events with unequal probabilities, *Operations Research*, **3**, 198–209. [4.9]

Walter, G. G., and Hamedani, G. G. (1987). Empirical Bayes estimation of binomial probability, *Communications in Statistics—Theory and Methods*, **16**, 559–577. [3.8.2]

Wang, P. C. C. (1970). *A Characterization of the Poisson Distribution Based on Random Splitting and Random Expanding*, Technical Report 1160, Stanford, CA: Stanford University Department of Statistics. [4.8]

Wang, Y. (1996). Estimation problems for the two-parameter negative binomial distribution, *Statistics and Probability Letters*, **26**, 113–114. [5.8.5]

Wang, Y., Young, L. J., and Johnson, D. E. (2001). A UMPU test for comparing means of two negative binomial distributions, *Communications in Statistics—Simulation and Computation*, **30**, 1053–1075. [5.8.5]

Wang, Y. H. (1972). On characterization of certain probability distributions, *Proceedings of the Cambridge Philosophical Society*, **71**, 347–352. [4.8]

Wani, J. K. (1967). Moment relations for some discrete distributions, *Skandinavisk Aktuarietidskrift*, **50**, 50–55. [7.1.8]

Wani, J. K. (1978). Measuring diversity in biological populations with logarithmic abundance distributions, *Canadian Journal of Statistics*, **6**, 219–228. [7.1.1]

Wani, J. K., and Lo, H.-P. (1975a). Large sample interval estimation for the logarithmic series distribution, *Canadian Journal of Statistics*, **3**, 277–284. [7.1.7]

Wani, J. K., and Lo, H.-P. (1975b). Clopper-Pearson system of confidence intervals for the logarithmic series distribution, *Biometrics*, **31**, 771–775. [7.1.7]

Wani, J. K., and Lo, H.-P. (1977). Comparing confidence intervals for the logarithmic series distribution, *Canadian Journal of Statistics*, **5**, 153–158. [7.1.7]

Warde, W. D., and Katti, S. K. (1971). Infinite divisibility of discrete distributions II, *Annals of Mathematical Statistics*, **42**, 1088–1090. [9.3]

Watanabe, H. (1956). On the Poisson distribution, *Journal of the Mathematical Society of Japan*, **8**, 127–134. [4.2.1]

Watson, H. W., and Galton, F. (1874). On the probability of extinction of families, *Journal of the Anthropological Institute*, **4**, 138–144. [9.1]

Weems, K. S., and Smith, P. J. (2004). On robustness of maximum likelihood estimates for Poisson-lognormal models, *Statistics and Probability Letters*, **66**, 189–196. [7.1.12]

Wei, L. J. (1979). The generalized Polya's urn design for sequential medical trials, *Annals of Statistics*, **5**, 291–296. [11.2.19]

Weinberg, C. R., and Gladen, B. C. (1986). The beta-geometric distribution applied to comparative fecundability studies, *Biometrics*, **42**, 547–560. [5.12.2, 6.10.4, 10.8.2]

Weiss, G. H. (1965). A model for the spread of epidemics by carriers, *Biometrics*, **21**, 481–490. [10.4.1, 10.4.3]

Weiss, L. (1976). Two sample tests and tests of fit, *Communications in Statistics—Theory and Methods*, **5**, 1275–1285. [10.6.1]

Weiss, L. (1983). Generalized maximum likelihood estimation, *Encyclopedia of Statistical Sciences*, Vol. 3, S. Kotz, N. L. Johnson, and C. B. Read (editors), 348–352. New York: Wiley. [1.2.16]

Weiss, L. (1988). Runs, *Encyclopedia of Statistical Sciences*, Vol. 8, S. Kotz, N. L. Johnson, and C. B. Read (editors), 222–226. New York: Wiley. [10.1, 10.6.1]

Wen Chen (1980). On the weak form of Zipf's law, *Journal of Applied Probability*, **17**, 611–622. [11.2.20]

Wesolowski, J. (1988). A remark on a characterization of the Poisson process, *Demonstratio Mathematica*, **21**, 555–557. [4.8]

Wesolowski, J. (1989). Characterizations of some processes by properties of conditional moments, *Demonstratio Mathematica*, **22**, 537–556. [5.9.1]

Wetherill, G. B., and Köllerström, J. (1979). Sampling inspection simplified, *Journal of the Royal Statistical Society, Series A*, **142**, 1–32 (erratum 404). [3.6.1]

Whittaker, L. (1914). On the Poisson law of small numbers, *Biometrika*, **10**, 36–71. [4.2.2, 5.3]

Whitworth, W. A. (1878). *Choice and Chance*, Cambridge: Deighton Bell. [10.2]

Whitworth, W. A. (1948). *Choice and Chance: With One Thousand Exercises* (reprint of the fifth edition, 1901), New York: Hafner. [10.2]

Widdra, W. (1972). Eine Verallgemeinerung des "Gestzes seltener Ereignisse," *Metrika*, **19**, 68–71. [4.2.2]

Wilks, S. S. (1962). *Mathematical Statistics*, New York: Wiley. [11.2.12]

Williams, C. B. (1944). The number of publications written by biologists, *Annals of Eugenics*, **12**, 143–146. [7.1.9]

Williams, C. B. (1947). The logarithmic series and its application to biological problems, *Journal of Ecology*, **34**, 253–272. [7.1.2, 7.1.10]

Williams, C. B. (1964). *Patterns in the Balance of Nature*, London: Academic Press. [7.1.1, 7.1.2, 7.1.4, 7.1.6, 7.1.9, 11.2.20]

Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity, *Biometrics*, **31**, 949–952. [6.7.2, 6.9.2]

Williamson, E., and Bretherton, M. H. (1963). *Tables of the Negative Binomial Distribution*, New York: Wiley. [5.7]

Williamson, E., and Bretherton, M. H. (1964). Tables of the logarithmic series distribution, *Annals of Mathematical Statistics*, **35**, 284–297. [7.1.6, 7.1.7, 7.1.9]

Willmot, G. E. (1986). Mixed compound Poisson distributions, *ASTIN Bulletin*, **16**, S59–S79. [11.1.5]

Willmot, G. E. (1987a). On the probabilities of the log-zero-Poisson distribution, *Canadian Journal of Statistics*, **15**, 293–297. [8.2.4]

Willmot, G. E. (1987b). The Poisson–inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 113–127. [11.1.5]

Willmot, G. E. (1988a). Sundt and Jewell's family of discrete distributions, *ASTIN Bulletin*, **18**, 17–29. [2.3.2, 5.12.2]

Willmot, G. E. (1988b). Parameter orthogonality for a family of discrete distributions, *Journal of the American Statistical Association*, **83**, 517–521. [1.2.16]

Willmot, G. E. (1989a). Limiting tail behaviour of some discrete compound distributions, *Insurance: Mathematics and Economics*, **8**, 175–185. [5.12.5, 8.3.2]

Willmot, G. E. (1989b). A remark on the Poisson-Pascal and some other contagious distributions, *Statistics and Probability Letters*, **7**, 217–220. [5.12.5, 9.12]

Willmot, G. E. (1990). Asymptotic tail behaviour of Poisson mixtures with applications, *Advances in Applied Probability*, **22**, 147–159. [8.3.2]

Willmot, G. E., and Sundt, B. (1989). On evaluation of the Delaporte distribution and related distributions, *Scandinavian Actuarial Journal*, **1989**, 101–113. [5.12.5]

Willson, L. J., Folks, J. L., and Young, J. H. (1984). Multistage estimation compared with fixed-sample-size estimation of the negative binomial parameter k, *Biometrics*, **40**, 109–117. [5.8.3]

Wilson, L. T., and Room, P. M. (1983). Clumping patterns of fruit and arthropods in cotton, with implications for binomial sampling, *Environmental Entomology*, **12**, 50–54. [5.10]

Wimmer, G., and Altmann, G. (1995). Generalized Gegenbauer distribution revised, *Sankhyā, Series B*, **57**, 450–452. [11.2.7, 11.2.20]

Wimmer, G., and Altmann, G. (1996). The multiple Poisson distribution: Its characteristics and a variety of forms, *Biometrical Journal*, **38**, 995–1011. [9.12]

Wimmer, G., and Altmann, G. (1999). *Thesaurus of Univariate Discrete Probability Distributions*, Essen: Stamm Verlag. [1.1.1, 1.1.3, 6.9.1–6.9.3, 6.10.4, 6.11.2, 6.11.3, 11.2.15, 11.2.20]

Wimmer, G., and Altmann, G. (2001). A new type of partial-sums distributions, *Statistics and Probability Letters*, **52**, 359–364. [5.12.2, 11.2.13]

Winkelmann, R. (2000). *Econometric Analysis of Count Data* (third edition), Berlin: Springer-Verlag. [3.11, 7.1.9, 8.2.3, 11.1.1, 11.1.3]

Winkelmann, R., Signorino, C., and King, G. (1995). A correction for an underdispersed event count probability distribution, *Political Analysis*, **5**, 215–228. [3.11, 11.1.3]

Winkelmann, R., and Zimmerman, K. F. (1995). Recent developments in count data modeling: Theory and applications, *Journal of Economic Surveys*, **9**, 1–24. [11.1.1]

Wise, M. E. (1946). The use of the negative binomial distribution in an industrial sampling problem, *Journal of the Royal Statistical Society, Series B*, **8**, 202–211. [5.8.3]

Wise, M. E. (1954). A quickly convergent expansion for cumulative hypergeometric probabilities, direct and inverse, *Biometrika*, **41**, 317–329. [6.5]

Wishart, J., and Hirschfeld, H. O. (1936). A theorem concerning the distribution of joins between line segments, *Journal of the London Mathematical Society*, **11**, 227–235. [10.6.1]

Witherby, H. F., Jourdain, F. C. R., Tichihurst, N. F., and Tucker, B. W. (1941). *The Handbook of British Birds*, Vol. 5, London: H. T. and G. Witherby. [7.1.9]

Wittes, J. T. (1972). On the bias and estimated variance of Chapman's two-sample capture-recapture population estimate, *Biometrics*, **28**, 592–597. [6.7.1, 6.9.1]

Wood, C. L., and Altavela, M. M. (1978). Large sample results for Kolmogorov-Smirnov statistics for discrete distributions, *Biometrika*, **65**, 235–239. [4.7.4]

Woodbury, M. A. (1949). On a probability distribution, *Annals of Mathematical Statistics*, **20**, 311–313. [11.2.19]

Worsley, K. J. (1982). An improved Bonferroni inequality and applications, *Biometrika*, **69**, 297–302. [10.2]

Wyshak, G. (1974). Algorithm AS68: A program for estimating the parameters of the truncated negative binomial distribution, *Applied Statistics*, **23**, 87–91. [5.11]

Xekalaki, E. (1981). Chance mechanisms for the generalized Waring distribution and related characterizations, *Statistical Distributions in Scientific Work*, Vol. 4: *Models, Structures, and Characterizations*, C. Taillie, G. P. Patil, and B. A. Baldessari (editors), 157–172. Dordrecht: Reidel. [6.2.3, 6.8, 6.9.3]

Xekalaki, E. (1983a). Infinite divisibility, completeness and regression properties of the generalized Waring distribution, *Annals of the Institute of Statistical Mathematics, Tokyo*, **35**, 279–289. [6.2.3, 6.9.3]

Xekalaki, E. (1983b). The univariate generalized Waring distribution in relation to accident theory: Proneness, spells or contagion? *Biometrics*, **39**, 887–895. [6.2.3, 6.9.3]

Xekalaki, E. (1983c). A property of the Yule distribution and its applications, *Communications in Statistics—Theory and Methods*, **12**, 1181–1189. [6.2.3, 6.9.3, 6.10.3, 11.2.13, 11.2.15]

Xekalaki, E. (1983d). Hazard functions and life distributions in discrete time, *Communications in Statistics—Theory and Methods*, **12**, 2503–2509. [6.2.3, 6.10.4, 11.2.15]

Xekalaki, E. (1984). Linear regression and the Yule distribution, *Journal of Econometrics*, **24**, 397–403. [6.10.3]

Xekalaki, E. (1985). Some identifiability problems involving generalized Waring distributions, *Publicationes Mathematicae*, **32**, 75–84. [6.2.3, 6.9.3]

Xekalaki, E., and Dimaki, C. (2005). Identifying the Pareto and Yule distributions by properties of their reliability measures, *Journal of Statistical Planning and Inference*, in press. [6.10.3]

Xekalaki, E., and Panaretos, J. (1988). On the association of the Pareto and the Yule distribution, *Teoriya Veroyatnostei i ee Primeneniya*, **33**, 206–210. [6.10.3]

Xekalaki, E., and Panaretos, J. (1989). On some distributions arising in inverse cluster sampling, *Communications in Statistics—Theory and Methods*, **18**, 355–366. [7.1.11, 10.7.6]

Xekalaki, E., Panaretos, J., and Philippou, A. N. (1987). On some mixtures of distributions of order *k*, *Fibonacci Quarterly*, **25**, 151–160. [10.7.6]

Yakowicz, S. J., and Spragins, J. D. (1968). On the identifiability of finite mixtures, *Annals of Mathematical Statistics*, **39**, 209–214. [8.2.2]

Yanagimoto, T. (1988). The conditional MLE in the two-parameter geometric distribution and its competitors, *Communications in Statistics—Theory and Methods*, **17**, 2779–2787. [5.2]

Yanagimoto, T. (1989). The inverse binomial distribution as a statistical model, *Communications in Statistics—Theory and Methods*, **18**, 3625–3633. [11.2.10]

Yang, M.-C. (1990). Compromise between generalized Bayes and Bayes estimators of Poisson means under entropy loss, *Communications in Statistics—Theory and Methods*, **19**, 935–951. [4.7.2]

Yannaros, N. (1991). Poisson approximation for random sums of Bernoulli random variables, *Statistics and Probability Letters*, **11**, 161–165. [3.12.3, 4.2.2]

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test, *Supplement to the Journal of the Royal Statistical Society*, **1**, 217–235. [6.9.1]

Yeh, H.-C. (2002). Six multivariate Zipf distributions and their related properties, *Statistics and Probability Letters*, **56**, 131–141. [11.2.20]

Yip, P. (1988). Inference about the mean of a Poisson distribution in the presence of a nuisance parameter, *Australian Journal of Statistics*, **30**, 299–306. [4.7.2]

Yoneda, K. (1962). Estimations in some modified Poisson distributions, *Yokohama Mathematical Journal*, **10**, 73–96. [8.2.4]

Young, D. H. (1970). The order statistics of the negative binomial distribution, *Biometrika*, **57**, 180–181. [5.5]

Young, T. (1819). Remarks on the probabilities of error in physical observations, and on the density of the earth, considered, especially with regard to the reduction of experiments on the pendulum. In a letter to Capt HENRY KATER, F.R.S., *By* Thomas Young, M.D., *For. Sec.* R. S., *Philosophical Transactions of the Royal Society of London*, **1819**, 70–95. [10.3]

Yousry, M. A., and Srivastava, R. C. (1987). The hyper-negative binomial distribution, *Biometrical Journal*, **29**, 875–884. [2.3.1, 5.12.9, 6.11.2]

Yu, C., Waller, L. A., and Zelterman, D. (1998). Discrete distributions for use in twin studies, *Biometrics*, **54**, 546–557. [6.9.1]

Yu, C., and Zelterman, D. (2002). Sums of dependent Bernoulli random variables and disease clustering, *Statistical and Probability Letters*, **57**, 363–373. [3.12.6]

Yuan, L., and Kalbfleisch, J. D. (2000). On the Bessel distribution and related problems, *Annals of the Institute of Statistical Mathematics, Tokyo*, **52**, 438–447. [11.2.4]

Yule, G. U. (1925). A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F.R.S., *Philosophical Transactions of the Royal Society of London, Series B*, **213**, 21–87. [5.3, 6.10.3, 11.2.20]

Zabell, S. (1983). Lexis, Wilhelm, *Encyclopedia of Statistical Sciences*, Vol. 4, S. Kotz, N. L. Johnson, and C. B. Read (editors), 624–625. New York: Wiley. [3.12.3]

Zacks, S., and Goldfarb, D. (1966). Survival probabilities in crossing a field containing absorption points, *Naval Research Logistics Quarterly*, **13**, 35–48. [10.8.1]

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions, *Journal of the American Statistical Association*, **81**, 446–451. [4.7.2]

Zhang, Z., Burtness, B. A., and Zelterman, D. (2000). The maximum negative binomial distribution, *Journal of Statistical Planning and Inference*, **87**, 1–19. [5.12.7]

Zheng, Q. (1997). On a birth-and-death process induced distribution, *Biometrical Journal*, **39**, 699–705. [3.12.2]

Zheng, X., and Loh, W.-Y. (1995). Bootstrapping binomial confidence intervals, *Journal of Statistical Planning and Inference*, **43**, 355–380. [3.8.3]

Zhou, X.-H., Tsao, M., and Qin, G. (2004). New intervals for the difference between two independent binomial proportions, *Journal of Statistical Planning and Inference*, **123**, 97–115. [3.8.3]

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*, Cambridge, MA: Addison-Wesley. [11.2.20]

Zörnig, P., and Altmann, G. (1995). Unified representation of Zipf distributions, *Computational Statistics and Data Analysis*, **19**, 461–473. [11.2.20]

Zou, G., and Wan, A. T. K. (2003). Admissible and minimax estimation of the parameter n in the binomial distribution, *Journal of Statistical Planning and Inference*, **113**, 451–466. [3.8.2]

Abbreviations

cdf	cumulative distribution function
cf	characteristic function
cgf	cumulate generating function
DFR	decreasing failure rate
DFRA	decreasing failure rate on average
fcgf	factorial cumulate generating function
fmgf	factorial moment generating function
FR	failure rate
FSD	factorial series distribution
GHFD	generalized hypergeometric factorial distribution
GHPD	generalized hypergeometric probability distribution
GHRD	generalized hypergeometric recast distribution
GPSD	generalized power series distribution
IFR	increasing failure rate
IFRA	increasing failure rate on average
MED	maximum entropy distribution
mgf	moment generating function
MLE	maximum likelihood estimator
MLP	maximum likelihood program
MPSD	modified power series distribution
MVUE	minimum variance unbiased estimator
NBU	new better than used
NBUE	new better than used in expectation
NWU	new worse than used
NWUE	new worse than used in expectation
pdf	probability density function
pgf	probability generating function
pmf	probability mass function
Pr(E)	probability of event E

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

PSD	power series distribution
rv	random variable
STER	Sums of Truncated forms of the Expected value of the Reciprocal
UMVUE	uniformly minimum variance unbiased estimator

Index

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

WILEY SERIES IN PROBABILITY AND STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

† ABRAHAM and LEDOLTER · Statistical Methods for Forecasting
AGRESTI · Analysis of Ordinal Categorical Data
AGRESTI · An Introduction to Categorical Data Analysis
AGRESTI · Categorical Data Analysis, *Second Edition*
ALTMAN, GILL, and McDONALD · Numerical Issues in Statistical Computing for the Social Scientist
AMARATUNGA and CABRERA · Exploration and Analysis of DNA Microarray and Protein Array Data
ANDĚL · Mathematics of Chance
ANDERSON · An Introduction to Multivariate Statistical Analysis, *Third Edition*
* ANDERSON · The Statistical Analysis of Time Series
ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies
ANDERSON and LOYNES · The Teaching of Practical Statistics
ARMITAGE and DAVID (editors) · Advances in Biometry
ARNOLD, BALAKRISHNAN, and NAGARAJA · Records
* ARTHANARI and DODGE · Mathematical Programming in Statistics
* BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BALAKRISHNAN and KOUTRAS · Runs and Scans with Applications
BARNETT · Comparative Statistical Inference, *Third Edition*
BARNETT and LEWIS · Outliers in Statistical Data, *Third Edition*
BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
BASILEVSKY · Statistical Factor Analysis and Related Methods: Theory and Applications
BASU and RIGDON · Statistical Methods for the Reliability of Repairable Systems
BATES and WATTS · Nonlinear Regression Analysis and Its Applications
BECHHOFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons
BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

Univariate Discrete Distributions, Third Edition.
By Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz

† BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
BENDAT and PIERSOL · Random Data: Analysis and Measurement Procedures, *Third Edition*
BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
BERNARDO and SMITH · Bayesian Theory
BHAT and MILLER · Elements of Applied Stochastic Processes, *Third Edition*
BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
† BIEMER, GROVES, LYBERG, MATHIOWETZ, and SUDMAN · Measurement Errors in Surveys
BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
BILLINGSLEY · Probability and Measure, *Third Edition*
BIRKES and DODGE · Alternative Methods of Regression
BLISCHKE AND MURTHY (editors) · Case Studies in Reliability and Maintenance
BLISCHKE AND MURTHY · Reliability: Modeling, Prediction, and Optimization
BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
BOLLEN · Structural Equations with Latent Variables
BOROVKOV · Ergodicity and Stability of Stochastic Processes
BOULEAU · Numerical Methods for Stochastic Processes
BOX · Bayesian Inference in Statistical Analysis
BOX · R. A. Fisher, the Life of a Scientist
BOX and DRAPER · Empirical Model-Building and Response Surfaces
* BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
BOX, HUNTER, and HUNTER · Statistics for Experimenters: Design, Innovation, and Discovery, *Second Editon*
BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
BRANDIMARTE · Numerical Methods in Finance: A MATLAB-Based Introduction
BROWN and HOLLANDER · Statistics: A Biomedical Introduction
BRUNNER, DOMHOF, and LANGER · Nonparametric Analysis of Longitudinal Data in Factorial Experiments
BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
CAIROLI and DALANG · Sequential Stochastic Optimization
CASTILLO, HADI, BALAKRISHNAN, and SARABIA · Extreme Value and Related Models with Applications in Engineering and Science
CHAN · Time Series: Applications to Finance
CHARALAMBIDES · Combinatorial Methods in Discrete Distributions
CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*
CHERNICK · Bootstrap Methods: A Practitioner's Guide
CHERNICK and FRIIS · Introductory Biostatistics for the Health Sciences
CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies, *Second Edition*
CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
* COCHRAN and COX · Experimental Designs, *Second Edition*
CONGDON · Applied Bayesian Modelling
CONGDON · Bayesian Statistical Modelling
CONOVER · Practical Nonparametric Statistics, *Third Edition*
COOK · Regression Graphics
COOK and WEISBERG · Applied Regression Including Computing and Graphics

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

COOK and WEISBERG · An Introduction to Regression Graphics
CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Third Edition*
COVER and THOMAS · Elements of Information Theory
COX · A Handbook of Introductory Statistical Methods
* COX · Planning of Experiments
CRESSIE · Statistics for Spatial Data, *Revised Edition*
CSÖRGŐ and HORVÁTH · Limit Theorems in Change Point Analysis
DANIEL · Applications of Statistics to Industrial Experimentation
DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Eighth Edition*
* DANIEL · Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*
DASU and JOHNSON · Exploratory Data Mining and Data Cleaning
DAVID and NAGARAJA · Order Statistics, *Third Edition*
* DEGROOT, FIENBERG, and KADANE · Statistics and the Law
DEL CASTILLO · Statistical Process Adjustment for Quality Control
DEMARIS · Regression with Social Data: Modeling Continuous and Limited Response Variables
DEMIDENKO · Mixed Models: Theory and Applications
DENISON, HOLMES, MALLICK and SMITH · Bayesian Methods for Nonlinear Classification and Regression
DETTE and STUDDEN · The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis
DEY and MUKERJEE · Fractional Factorial Plans
DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
DODGE · Alternative Methods of Regression
* DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
* DOOB · Stochastic Processes
DOWDY, WEARDEN, and CHILKO · Statistics for Research, *Third Edition*
DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
DRYDEN and MARDIA · Statistical Shape Analysis
DUDEWICZ and MISHRA · Modern Mathematical Statistics
DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
* ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
ENDERS · Applied Econometric Time Series
† ETHIER and KURTZ · Markov Processes: Characterization and Convergence
EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition,* Revised; Volume II, *Second Edition*
FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
FITZMAURICE, LAIRD, and WARE · Applied Longitudinal Analysis
* FLEISS · The Design and Analysis of Clinical Experiments
FLEISS · Statistical Methods for Rates and Proportions, *Third Edition*
† FLEMING and HARRINGTON · Counting Processes and Survival Analysis
FULLER · Introduction to Statistical Time Series, *Second Edition*
FULLER · Measurement Error Models
GALLANT · Nonlinear Statistical Models
GEISSER · Modes of Parametric Statistical Inference
GEWEKE · Contemporary Bayesian Econometrics and Statistics
GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

GIESBRECHT and GUMPERTZ · Planning, Construction, and Statistical Analysis of Comparative Experiments
GIFI · Nonlinear Multivariate Analysis
GIVENS and HOETING · Computational Statistics
GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*
GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing
GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
† GROVES · Survey Errors and Survey Costs
* HAHN and SHAPIRO · Statistical Models in Engineering
HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
HALD · A History of Probability and Statistics and their Applications Before 1750
HALD · A History of Mathematical Statistics from 1750 to 1930
† HAMPEL · Robust Statistics: The Approach Based on Influence Functions
HANNAN and DEISTLER · The Statistical Theory of Linear Systems
HEIBERGER · Computation for the Analysis of Designed Experiments
HEDAYAT and SINHA · Design and Inference in Finite Population Sampling
HELLER · MACSYMA for Statisticians
HINKELMANN and KEMPTHORNE · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design
HINKELMANN and KEMPTHORNE · Design and Analysis of Experiments, Volume 2: Advanced Experimental Design
HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes
* HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
HOCHBERG and TAMHANE · Multiple Comparison Procedures
HOCKING · Methods and Applications of Linear Models: Regression and the Analysis of Variance, *Second Edition*
HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
HOGG and KLUGMAN · Loss Distributions
HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
† HUBER · Robust Statistics
HUBERTY · Applied Discriminant Analysis
HUNT and KENNEDY · Financial Derivatives in Theory and Practice
HUSKOVA, BERAN, and DUPAC · Collected Works of Jaroslav Hajek—with Commentary
HUZURBAZAR · Flowgraph Models for Multistate Time-to-Event Data
IMAN and CONOVER · A Modern Approach to Statistics
† JACKSON · A User's Guide to Principle Components
JOHN · Statistical Methods in Engineering and Quality Assurance
JOHNSON · Multivariate Statistical Simulation
JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
JOHNSON and BHATTACHARYYA · Statistics: Principles and Methods, *Fifth Edition*
JOHNSON and KOTZ · Distributions in Statistics

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 2, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Third Edition*
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
JUREČKOVÁ and SEN · Robust Statistical Procedures: Aymptotics and Interrelations
JUREK and MASON · Operator-Limit Distributions in Probability Theory
KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second Edition*
KASS and VOS · Geometrical Foundations of Asymptotic Inference
† KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
KEDEM and FOKIANOS · Regression Models for Time Series Analysis
KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models
* KISH · Statistical Design for Research
KLEIBER and KOTZ · Statistical Size Distributions in Economics and Actuarial Sciences
KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions, *Second Edition*
KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models: From Data to Decisions, *Second Edition*
KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions, Volume 1, *Second Edition*
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of Time-Dependent Systems with Practical Applications
LACHIN · Biostatistical Methods: The Assessment of Relative Risks
LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction
LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*
LAWSON · Statistical Methods in Spatial Epidemiology
LE · Applied Categorical Data Analysis
LE · Applied Survival Analysis
LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

LePAGE and BILLARD · Exploring the Limits of Bootstrap
LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics
LIAO · Statistical Group Comparison
LINDVALL · Lectures on the Coupling Method
LINHART and ZUCCHINI · Model Selection
LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*
LLOYD · The Statistical Analysis of Categorical Data
LOWEN and TEICH · Fractal-Based Point Processes
MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics, *Revised Edition*
MALLER and ZHOU · Survival Analysis with Long Term Survivors
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data
MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
MARCHETTE · Random Graphs for Statistical Pattern Recognition
MARDIA and JUPP · Directional Statistics
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science, *Second Edition*
McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
McFADDEN · Management of Data in Clinical Trials
* McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition
McLACHLAN, DO, and AMBROISE · Analyzing Microarray Gene Expression Data
McLACHLAN and KRISHNAN · The EM Algorithm and Extensions
McLACHLAN and PEEL · Finite Mixture Models
McNEIL · Epidemiological Research Methods
MEEKER and ESCOBAR · Statistical Methods for Reliability Data
MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice
MICKEY, DUNN, and CLARK · Applied Statistics: Analysis of Variance and Regression, *Third Edition*
* MILLER · Survival Analysis, *Second Edition*
MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis, *Third Edition*
MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
MUIRHEAD · Aspects of Multivariate Statistical Theory
MULLER and STOYAN · Comparison Methods for Stochastic Models and Risks
MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and Nonlinear Optimization
MURTHY, XIE, and JIANG · Weibull Models
MYERS and MONTGOMERY · Response Surface Methodology: Process and Product Optimization Using Designed Experiments, *Second Edition*
MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With Applications in Engineering and the Sciences
† NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
† NELSON · Applied Life Data Analysis
NEWMAN · Biostatistical Methods in Epidemiology
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
PALTA · Quantitative Methods in Population Health: Extensions of Ordinary Regressions

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
* PARZEN · Modern Probability Theory and Its Applications
PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
PIANTADOSI · Clinical Trials: A Methodologic Perspective
PORT · Theoretical Probability for Applications
POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
PRESS · Bayesian Statistics: Principles, Models, and Applications
PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
PUKELSHEIM · Optimal Experimental Design
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
† PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
QIU · Image Processing and Jump Regression Analysis
* RAO · Linear Statistical Inference and Its Applications, *Second Edition*
RAUSAND and HØYLAND · System Reliability Theory: Models, Statistical Methods, and Applications, *Second Edition*
RENCHER · Linear Models in Statistics
RENCHER · Methods of Multivariate Analysis, *Second Edition*
RENCHER · Multivariate Statistical Inference with Applications
* RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROBINSON · Practical Strategies for Experimenting
ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
ROSS · Introduction to Probability and Statistics for Engineers and Scientists
† ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
* RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Modern Regression Methods
RYAN · Statistical Methods for Quality Improvement, *Second Edition*
SALTELLI, CHAN, and SCOTT (editors) · Sensitivity Analysis
* SCHEFFE · The Analysis of Variance
SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
SCHOTT · Matrix Analysis for Statistics, *Second Edition*
SCHOUTENS · Levy Processes in Finance: Pricing Financial Derivatives
SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
* SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE · Matrix Algebra Useful for Statistics
SEARLE, CASELLA, and McCULLOCH · Variance Components
SEARLE and WILLETT · Matrix Algebra for Applied Economics
SEBER and LEE · Linear Regression Analysis, *Second Edition*
† SEBER · Multivariate Observations
† SEBER and WILD · Nonlinear Regression
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
* SERFLING · Approximation Theorems of Mathematical Statistics
SHAFER and VOVK · Probability and Finance: It's Only a Game!

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.

SILVAPULLE and SEN · Constrained Statistical Inference: Inequality, Order, and Shape Restrictions
SMALL and McLEISH · Hilbert Space Methods in Probability and Statistical Inference
SRIVASTAVA · Methods of Multivariate Statistics
STAPLETON · Linear Statistical Models
STAUDTE and SHEATHER · Robust Estimation and Testing
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
STYAN · The Collected Papers of T. W. Anderson: 1943–1985
SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
THOMPSON · Empirical Model Building
THOMPSON · Sampling, *Second Edition*
THOMPSON · Simulation: A Modeler's Approach
THOMPSON and SEBER · Adaptive Sampling
THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
TSAY · Analysis of Financial Time Series, *Second Edition*
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN BELLE · Statistical Rules of Thumb
VAN BELLE, FISHER, HEAGERTY, and LUMLEY · Biostatistics: A Methodology for the Health Sciences, *Second Edition*
VESTRUP · The Theory of Measures and Integration
VIDAKOVIC · Statistical Modeling by Wavelets
VINOD and REAGLE · Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments
WALLER and GOTWAY · Applied Spatial Statistics for Public Health Data
WEERAHANDI · Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models
WEISBERG · Applied Linear Regression, *Third Edition*
WELSH · Aspects of Statistical Inference
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for *p*-Value Adjustment
WHITTAKER · Graphical Models in Applied Multivariate Statistics
WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOODWORTH · Biostatistics: A Bayesian Introduction
WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
YANG · The Construction Theory of Denumerable Markov Processes
* ZELLNER · An Introduction to Bayesian Inference in Econometrics
ZHOU, OBUCHOWSKI, and McCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in a lower priced paperback edition in the Wiley–Interscience Paperback Series.