PARAMETER ESTIMATION AND OTHER ASPECTS OF UNIVARIATE AND BIVARIATE GENERALIZED POISSON DISTRIBUTIONS

A study of univariate and bivariate generalized Poisson distributions, with special reference to parameter estimation, goodness-of-fit, and infinite divisibility

BY

JULIUS ABIODUN OLADIPO

M.Sc., Grad. Cert. Ed.

A Dissertation submitted to

the University of Bradford, Bradford, England

for the degree of Ph.D.

1983

PREFACE

Interest in discrete distributions has increased markedly over the past two decades or so. These distributions are generally studied in the context of the specific chance mechanisms in which they take their genesis or in relation to classes defined by specific mathematical or probabilistic structures. One of the major classes of discrete distributions comprises generalized Poisson distributions (GPDs).

It is well known that GPDs are applicable to diverse areas e.g. in accident studies and in bacteriology and entomology. However, relatively few discrete distributions are viewed as GPDs. It is desirable to know which of the other numerous distributions presented in the literature are GPDs. A clearer picture of the functional forms and properties of this important class of discrete distributions is invaluable.

This thesis discusses several aspects of generalized Poisson distributions. Special attention is given to features which have direct bearing on applications. Chapter 1 is a discussion of discrete distributions in general. In Chapter 2 generalized Poisson distributions are discussed extensively. By invoking a theorem which characterizes GPDs by the property of infinite divisibility, some well known, and a few apparently new, discrete distributions are shown to be GPDs.

Chapter 3 is about parameter estimation. The major methods for estimating the parameters of GPDs are discussed. Attention is then focussed on one particular GPD, the Poisson Λ inverse normal distribution. For this distribution, a parameterization comprising orthogonal parameters is identified. Maximum likelihood equations for estimating these parameters are presented. Some other estimators of the parameters are also given.

Chapter 4 discusses the problem of how to choose from several plausible GPDs the one which best fits a given data-set. In Chapter 5 the thesis is concluded with a discussion of the possible directions in which further work on the subject matter can be pursued.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor C.D. Kemp for all his efforts and assistance throughout the period spent on this research. I am also greatly indebted to Dr. A.W. Kemp for her helpful advice and encouragement.

Thanks are also due to Mrs. Valerie Hunter for the speedy typing of the manuscript.

The research was funded jointly by the Kwara State Government of Nigeria and the Ahmadu Bello University, Zaria, Nigeria.

In the characteristic African manner, every member of my wide family and each of my friends took a personal and sustained interest in the progress of the work and hence when success is attained it is rightly "our" success.

C O N T E N T S

		Page
Preface		(i)
Acknowledgements		(iii)
CHAPTER 1.	SUMMARY AND GENERAL BACKGROUND	1
1.1	Summary	1
1.2	Basic Definitions	3
	1.2.1 Introduction	3
	1.2.2 Random variable and distribution function	4
	1.2.3 Moments	6
	1.2.4 Generating functions	7
1.3	Some Mathematical Functions	9
	1.3.1 Introduction	9
	1.3.2 The gamma and factorial functions	10
	1.3.3 The beta function	11
	1.3.4 The generalized hypergeometric function	11
	1.3.5 The modified Bessel functions	12
1.4	Families of Discrete Distributions	14
	1.4.1 Introduction	14
	1.4.2 Power series and generalized power series distributions	14
	1.4.3 Differece-equation distributions	15
	1.4.4 Generalized hypergeometric distributions	16
	1.4.5 Compound or mixed distributions	20
	1.4.6 Generalized distributions	24
	1.4.7 The Lagrangian probability distributions	27
	1.4.8 Binomial generalization distributions	29
1.5	Chance Mechanisms Generating Discrete Distributions	30
	1.5.1 Introduction	30
	1.5.2 Mixing models	31

		Page
	1.5.3 Clustering models	34
	1.5.4 Branching models	35
	1.5.5 Birth and birth-death models	37
	1.5.6 Waiting-time distributions	37
	1.5.7 Weighted distributions and damaged models	38
	1.5.8 STER models	40
1.6	Computer Generation of (Univariate) Discrete Variables	41
	1.6.1 Introduction	41
	1.6.2 The inverse transformation method	42
	1.6.3 The acceptance-rejection method	43
	1.6.4 The composition method	45
	1.6.5 Specialized methods based on stochastic models	45
1.7	Bivariate Discrete Distributions	47
	1.7.1 Introduction	47
	1.7.2 The structure of bivariate distributions	47
	1.7.3 Marginality and conditionality	51
	1.7.4 Chance mechanisms generating bivariate discrete distributions	54
	1.7.5 Computer generation of bivariate discrete variables	54
1.8	Summary on "Discrete Distributions"	55
	CHAPTER 2. GENERALIZED POISSON DISTRIBUTIONS	57
2.1	Introduction	57
2.2	Properties of Generalized Poisson Distributions	59
	2.2.1 Necessary conditions for a GPD	59
	2.2.2 The GPD as a mixture of an n-fold convolution	60
	2.2.3 The infinite divisibility of GPDs	61
	2.2.4 The cluster-size distribution of a GPD	63
	2.2.5 Cumulants and factorial cumulants	65

		Page
	2.2.6 Khatri-Patel recurrence formula for the probabilities	66
	2.2.7 Examples of univariate GPDs	67
	2.2.8 Bivariate generalized Poisson distributions	71
2.3	Mixtures of the Poisson with Continuous Mixing Distributions	73
	2.3.1 Examples of Mixtures of the Poisson distribution	73
	2.3.2 The overdispersion of mixtures of the Poisson	79
	2.3.3 Factorial moments and factorial cumulants	82
2.4	Mixtures of the Poisson with Infinitely Divisible Continuous Mixing Distributions	83
	2.4.1 Maceda's theorem	83
	2.4.2 Infinitely divisible mixtures of the gamma	89
	2.4.3 The reciprocal-gamma-mixture of the Poisson	91
	2.4.4 The $f(x)$-mixture of the Poisson where $x \geq a > 0$	96
2.5	Other Classes of GPDs	97
	2.5.1 Mixtures of the geometric distribution	97
	2.5.2 Logconvex lattice distributions	99
	2.5.3 The geometric with-added-zeroes and related GPDs	101
	2.5.4 The geometric with-decreased-zeroes and related GPDs	106
2.6	Genesis and Applications of GPDs	109
	2.6.1 The clustering model	109
	2.6.2 The stuttering-Poisson model	110
2.7	Summary on "Generalized Poisson Distributions"	111
CHAPTER 3.	THE POISSON INVERSE NORMAL DISTRIBUTION	114
3.1	Introduction	114
3.2	Parameter Estimation Methods for GPDs	119
	3.2.1 Maximum likelihood estimation	119
	3.2.2 The moments method	127
	3.2.3 The method of first two frequencies and their ratio	129

		Page
	3.2.4 The even-points method	130
	3.2.5 The minimum-chi-square method	135
3.3	The P.I.N. Distribution: Maximum Likelihood Estimation	136
	3.3.1 Partial derivatives of $\log P_x$: Method I	136
	3.3.2 Partial derivatives of $\log P_x$: Method II	139
	3.3.3 The maximum likelihood equations	142
	3.3.4 The asymptotic correlation of the estimates	144
	3.3.5 Discussion of the results on maximum likelihood estimation	146
3.4	The P.I.N. Distribution: Even-Points Estimation	148
	3.4.1 The even-points estimation-method	148
	3.4.2 Three other types of explicit estimators	149
	3.4.3 Asymptotic efficiencies of the four sets of explicit estimators	151
3.5	Summary on "The P.I.N. Distribution"	153
	CHAPTER 4. THE PROBLEM OF DISTRIBUTION SELECTION	155
4.1	Introduction	155
4.2	Distribution Selection, Model Discrimination, and the Global Superiority of a Distribution	155
4.3	Gurland's Graphical Method	157
4.4	Ranking by Likelihood	158
4.5	Goodness-of-Fit Statistics and Criteria for Distribution Selection	159
	CHAPTER 5. POSTVIEW AND PROSPECTS	161
	TABLES	164
	REFERENCES	173

CHAPTER 1

Summary and General Background

CHAPTER 1

SUMMARY AND GENERAL BACKGROUND

1.1 Summary

Endowed with a wide and rapidly expanding literature, both in diverse fields of applied science and in statistics, the subject of Discrete Distributions offers a fascinating area for study and research. The subject is commonly discussed in one of two ways. One is to investigate in detail individual discrete distributions or specific families of distributions for their mathematical and statistical properties and for their natural applications. The second approach is to concentrate on a scientific problem that generates or requires random counts and then create and coordinate appropriate discrete models and methods so as to arrive at a meaningful solution of the scientific problem.

This thesis is about a specific family of discrete distributions, the generalized Poisson distributions (GPDs) - their structural properties, their applications, and also the methods for estimating their parameters. To place the study in context, a discussion of discrete distributions in general is given (Chapter 1). This includes a description of the major families or classes. Then generalized Poisson distributions are discussed extensively (Chapter 2). The derivation of one apparently new generalized Poisson distribution is also given. Infinite divisibility characterizes GPDs among lattice distributions supported on non-negative integers, and this property dominates the discussion. The study then narrows down to one particular GPD, the recently proposed Poisson Λ inverse normal (PIN) distribution (Chapter 3), for whose parameters a new set of estimators are presented, in addition to a re-appraisal of its structure.

An important aspect of GPDs which has received relatively little attention in the literature is how to choose from a given set of

KASHIM IBRAHIM LIBRARY
AHMADU BELLO UNIVERSITY

plausible GPDs the one most appropriate for the situation. This problem is briefly discussed (Chapter 4) in the context of future work. The last chapter (Chapter 5) is a recapitulation and an outline of some ensuing problems which are subjects for further research.

The concept of generalized Poisson distributions has featured in statistical literature over the years in one form or the other since the early part of this century. GPDs are widely applicable e.g. in the modelling of populations of plants and insects, and in accident studies. However, in practice, relatively few discrete distributions are viewed as GPDs and, due to various computational difficulties only a proportion of these have been adopted by practitioners. As has been noted by Fiacco and McCormick (1968, p.156), "a vital test and justification of any body of theory of how to solve problems is the feasibility of computational implementation and practical application". This forms the general motivation for the thesis.

The Neyman Type A distribution in its usual form is Poisson V Poisson, but can also be represented as Poisson V zero-truncated Poisson. The latter form is apparently not widely known. Similarly the Polya-Aeppli distribution, usually given as Poisson V shifted geometric, is in fact identical with Poisson V geometric.

All mixtures of the Poisson distribution are overdispersed; so are all generalized Poisson distributions, except the Poisson, which is trivially a GPD having the Bernoulli distribution as its cluster-size distribution. Hence overdispersion is a necessary (but certainly not sufficient) condition for a non-Poisson lattice distribution to be a GPD.

The quotient of the probability generating functions (pgfs) of two geometric distributions with parameters θ_1 and θ_2, $0 < \theta_2 < \theta_1 < 1$, gives the pgf of the geometric distribution-with-added-zeroes, a distribution which is infinitely divisible and hence a GPD. Some other infinitely divisible convolutions of pseudo-binomials are similarly re-appraised and recast into conventional forms.

All logconvex lattice distributions are infinitely divisible and are hence generalized Poisson distributions. A subset of generalized hypergeometric distributions is identified as being logconvex. Its members include the Waring, the shifted logarithmic, and the shifted lost-games distributions.

All mixtures of the Poisson with infinitely divisible mixing distributions are infinitely divisible. Some infinitely divisible continuous distributions with support on the positive half real line are listed. One of these is the reciprocal gamma distribution. Its mixture of the Poisson is derived and shown to be closely related to but simpler than the already known generalized-inverse-normal mixture of the Poisson. The special case analogous to the Poisson Λ inverse normal is discussed. Some other infinitely divisible mixtures of the Poisson are also derived.

The Poisson Λ inverse normal, although only recently proposed, has generated a considerable amount of literature. It is noted here that its cluster-size distribution is a lost-games distribution. Its two usual parameterizations are discussed and it is shown computationally and theoretically that the maximum likelihood estimates of the parameters in one of these two forms are asyptotically uncorrelated. For this form also, the sample mean is a maximum likelihood estimator of one of the parameters. A new estimator, the even-points estimator, is presented for the other parameter.

1.2 Basic Definitions

1.2.1 Introduction

This section comprises some of the basic definitions and the related notation connected with the characteristic properties of distributions. They are available in the literature but are set out here because they will be assumed in the rest of the thesis. It should

be noted at the outset that, in particular, discussions of features of most discrete distributions have been approached via their probability generating functions.

Well-established abbreviations such as cdf for 'cumulative distribution function', pdf for 'probability density function' (continuous), pf for 'probability function' (discrete), i.i.d. for 'independently and identically distributed', and inf div for 'infinitely divisible' will be used freely whenever their full meanings are unlikely to be in doubt. The definitions and notation adopted for "generalized" and "compound" distributions are those of Gurland (1957). All logarithms are natural logarithms except where otherwise indicated.

1.2.2 Random variable and distribution function

For a given probability space $(\Omega, A, Pr[\cdot])$ a random variable, denoted by X or $X(\cdot)$, is a function with domain Ω and counterdomain the real line. The function $X(\cdot)$ must be such that the set A_r, defined by $A_r = \{w: X(w) \leqslant r\}$ belongs to A for every real number r. A random variable X is said to be discrete if the range is countable; otherwise it is said to be continuous.

The cumulative distribution function (cdf), or simply the distribution function, of a random variable X denoted by $F_X(x)$, or $F(x)$, is defined to be that function with the domain the real line and counterdomain the interval $[0, 1]$ which, for every real number x, satisfies

$$\begin{aligned} F_X(x) &\equiv Pr[w: X(w) \leqslant x] \\ &\equiv Pr[X \leqslant x] \\ &\equiv \sum_{\alpha \in \beta} f(x), \ \beta \text{ being the range of } X. \end{aligned} \qquad (1.2.2.1)$$

The function $f(x)$, in the continuous case, is $F'(x)$, the derivative of $F(x)$ with respect to x. It is called the probability density function. In the discrete case, $f(x)$ can be regarded as the Radom-Nikodym

"derivative" of F(x) with respect to a counting measure. In this case, f(x) is usually referred to as the probability function, or sometimes, probability mass function, and alternatively denoted by P_x.

A discrete distribution is called a lattice distribution if the range of the random variable is not merely countable but in addition the intervals between those of its values for which there are non-zero probabilities are all integral multiples of one quantity, a quantity depending on the random variable. Points having the values of the random variable as coordinates then form a lattice. By an appropriate linear transformation it can be arranged that the random variable takes values which are integers. Most discrete distributions considered in this thesis are lattice distributions with the random variables taking non-negative integral values. They will be referred to simply as lattice distributions.

A discrete (lattice) distribution will be said to be finite, or terminating, if f(x) is non-zero for only a finite subset of the non-negative integers; otherwise it is said to be infinite, or non-terminating.

Unless otherwise indicated, capital letters (e.g. X, Y) denote random variables, while the corresponding small letters (x, y, etc.) represent the values taken by the random variables. Greek letters will usually be used to represent parameters, and in particular a parameter taking values between zero and one (inclusive if indicated accordingly) will be denoted by θ subscripted or unsubscripted. A specific distribution will sometimes be denoted in the common notation whereby, for example, Geometric (x; θ) implies that X is distributed as the geometric with the parameter θ and Neg binom (x; α, β) denotes that the distribution of X is negative binomial with parameters α and β where it would be clear from the context which of the two is the exponent parameter.

1.2.3 Moments

The rth power moment or simply the rth moment is the expected value of x^r and for a discrete distribution it is defined by

$$E[X^r] = \Sigma x^r f(x) \equiv \mu'_r, \quad r = 1, 2,... \qquad (1.2.3.1)$$

The mean, E[X], will usually be written as μ rather than as μ'_1.

The rth central moment, the expected value of $(x - \mu)^r$, of a discrete distribution, is defined by

$$E[(X - \mu)^r] = \Sigma(x - \mu)^r f(x) \equiv \mu_r, \quad r = 2, 3,... \qquad (1.2.3.2)$$

The variance, $E[(X - \mu)^2]$, will usually be referred to as var(X).

The rth descending factorial moment, often usually referred to simply as the rth factorial moment, is defined by

$$E[X_{(r)}] = \sum_{x=r}^{\infty} x_{(r)} f(x) \equiv \mu_{(r)}, \quad r = 1, 2,... \qquad (1.2.3.3)$$

where $x_{(r)} = x(x - 1) \ldots (x - r + 1)$, $r = 1, 2,...$, $x_{(0)} = 1$.

The ascending factorial moment is defined by

$$E[X^{(r)}] = \sum_{x=0}^{\infty} x^{(r)} f(x) \equiv \mu^{(r)}, \quad r = 1, 2,... \qquad (1.2.3.4)$$

where $x^{(r)} = x(x + 1) \ldots (x + r - 1)$, $r = 1, 2,...$, $x^{(0)} = 1$.

The summations given in these definitions are not necessarily convergent for every discrete distribution and hence their corresponding moments may, sometimes, not exist. Using the relations between the different types of moments (see e.g. Kendall and Stuart (1969), Chapter 4) it can be shown that the existence of one of μ'_r, μ_r, $\mu_{(r)}$, $\mu^{(r)}$ implies the existence of the others, and the existence of all moments of order less than r.

The rth inverse moment is defined by

$$E[1/X^r] = \sum_{x=0}^{\infty} \frac{1}{x^r} f(x) \equiv \mu'_{1/r}, \quad r = 1, 2,... \qquad (1.2.3.5)$$

A discrete distribution can possess inverse moments only if $f(0) = 0$.

It is possible to define other moments (e.g. incomplete moments and absolute moments) for discrete distributions.

1.2.4 Generating functions

The probability generating function (pgf) of a non-negative, integer-valued random variable X is defined by

$$G(z) = E(z^x) = \sum_{x=0}^{\infty} P_x z^x, \quad |z| \leq 1 \qquad (1.2.4.1)$$

The pgf always exists for every proper distribution of X and gives a complete characterization of the distribution. It provides a useful tool for the study of discrete distributions. The probabilities P_x can be obtained from the pgf G(z) using the relation

$$P_x = \frac{1}{x!}\left\{\left[\frac{d^x}{dz^x} G(z)\right]\Big|_{z=0}\right\}, \quad x = 0, 1,... \qquad (1.2.4.2)$$

Suppose $G_1(z)$ is the pgf of a given lattice distribution D_1. The distribution D_2 with the pgf $G_2(z) = z^k G_1(z)$, k integral, is said to be the shifted form of D_1. The distribution D_2 is said to have been shifted k units to the right or to the left according as to whether k is positive or negative. A reversed distribution has a pgf of the form $z^n G\left(\frac{1}{z}\right)$ where G(z) is the pgf of a distribution on the integers 0, 1,..., n.

Other generating functions can be defined for discrete distributions, in terms of the pgf. The moment generating function (mgf) can be defined as

$$G(e^z) = \sum_{x=0}^{\infty} e^{xz} P_x \qquad (1.2.4.3)$$

where the rth moment, if it exists, can be obtained as

$$\mu'_r = \left[\frac{d^r}{dz^r} G(e^z)\right]\Big|_{z=0}, \quad r = 1, 2,... \qquad (1.2.4.4)$$

In each expression, G(z) is the pgf.

The cumulant generating function (cgf) is defined as the logarithm of the moment generating function and hence is given by

$$\log G(e^z) = \log\left[\sum_{x=0}^{\infty} P_x e^{xz}\right]. \qquad (1.2.4.5)$$

The rth cumulant, k_r, if it exists, is defined as

$$k_r = \left[\frac{d^r}{dz^r} \log G(e^z)\right]\Bigg|_{z=0}, \quad r = 1, 2,... \qquad (1.2.4.6)$$

The cumulants are sometimes called the semi-invariants because the cumulants k_r, $r \geq 2$, of a random variable X are the same as those for X + A for any constant A. In particular, the cumulants, k_r, $r \geq 2$, of a lattice distribution are the same as those of its shifted forms.

The (descending) factorial moment generating function (fmgf) can be defined as

$$G(1 + z) = \sum_{x=0}^{\infty} (1 + z)^x P_x \qquad (1.2.4.7)$$

where $\mu_{(r)}$ can be obtained, if it exists, from

$$\mu_{(r)} = \left[\frac{d^r}{dz^r} G(1 + z)\right]\Bigg|_{z=0}, \quad r = 1, 2,... \qquad (1.2.4.8)$$

Note also that

$$\mu_{(r)} = \left[\frac{d^r}{dz^r} G(z)\right]\Bigg|_{z=1}, \quad r = 1, 2,... \qquad (1.2.4.9)$$

which implies that the pgf G(z) can also be regarded as a factorial moment generating function.

The ascending factorial moment generating function (afmgf) can be defined as

$$G\left\{(1 - z)^{-1}\right\} = \sum_{x=0}^{\infty} P_x/(1 - z)^x \qquad (1.2.4.10)$$

where $\mu^{(r)}$ can be obtained, if it exists, from

$$\mu^{(r)} = \left[\frac{d^r}{dz^r} G\left\{(1 - z)^{-1}\right\}\right]\Bigg|_{z = 0}, \quad r = 1, 2,... \qquad (1.2.4.11)$$

The (descending)factorial cumulant generating function can be defined as

$$\log G(1 + z) = \log \sum_{x=0}^{\infty} P_x (1 + z)^x \qquad (1.2.4.12)$$

and the factorial cumulant $k_{(r)}$ if it exists is given by

$$k_{(r)} = \left[\frac{d^r}{dz^r} \log G(1 + z)\right]\Bigg|_{z = 0}, \quad r = 1, 2,... \qquad (1.2.4.13)$$

The ascending factorial cumulant generating function (afcgf) can be defined as

$$\log\left[G\left\{(1 - z)^{-1}\right\}\right] = \log\left[\sum_{x=0}^{\infty} P_x/(1 - z)^x\right] \qquad (1.2.4.14)$$

and hence the rth ascending factorial cumulant, if it exists, is defined as

$$k^{(r)} = \left\{\frac{d^r}{dz^r} \ell n\left[G\left\{(1 - z)^{-1}\right\}\right]\right\}\Bigg|_{z = 0}, \quad r = 1, 2,... \qquad (1.2.4.15)$$

From the relationship between the cumulants and moments (see e.g. Kendall and Stuart (1969I, Chapter 4), Douglas (1971)), it can be shown that k_r, $k_{(r)}$, and $k^{(r)}$ exist if and only if μ'_r exists.

1.3 Some Mathematical Functions

1.3.1 Introduction

Some mathematical functions which have some key roles to play in the thesis are stated below, together with their associated notation. These are the gamma, beta, and factorial functions, the generalized hypergeometric function, and the modified Bessel functions.

1.3.2 The gamma and factorial functions

The gamma function is defined by Euler's integral as

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad Re(x) > 0 \qquad (1.3.2.1)$$

while the factorial function is defined by

$$x! = x(x-1) \ldots 1, \quad x = 1, 2, \ldots; \; 0! = 1. \qquad (1.3.2.2)$$

The recurrence relation

$$\Gamma(x+1) = x\Gamma(x), \quad x \neq 0, -1, -2, \ldots \qquad (1.3.2.3)$$

provides a link between the gamma and the factorial functions and can be used to extend the range of the latter, since

$$x! = \Gamma(x+1).$$

Ratios of gamma and factorial functions can be given meaning over the negative integers by defining (Kemp and Kemp (1956a))

$$\frac{x!}{(x+y)!} = \frac{(-1)^y(-x-y-1)!}{(-x-1)!} = \frac{(-1)^y\Gamma(-x-y)}{\Gamma(-x)} \qquad (1.3.2.4)$$

where $x < 0$, $y < 0$, and y is an integer.

Other related functions are the descending factorial, the ascending factorial, and the binomial coefficients. The descending factorial is defined by

$$x_{(r)} = x(x-1) \ldots (x-r+1) = \Gamma(x+1)/\Gamma(x-r+1) \qquad (1.3.2.5)$$

with $x_{(r)} = 0$ for $r > x$ if x is a positive integer. The ascending factorial is defined by

$$x^{(r)} = x(x+1) \ldots (x+r-1) = \Gamma(x+r)/\Gamma(x) \qquad (1.3.2.6)$$

with $x^{(r)} = 0$ for $r > |x|$ if x is an integer. The binomial coefficients are given by

$$\binom{x}{r} = \frac{x!}{r!(x-r)!} = \frac{\Gamma(x+1)}{\Gamma(r+1)\,\Gamma(x-r+1)}, \quad r = 0, 1, \ldots, x. \qquad (1.3.2.7)$$

Note that $\binom{x}{r} = x_{(r)}/r!$ and $\binom{x+r-1}{r-1} = x^{(r)}/x!$

Note also that $\binom{-x}{r} = (-1)^r \binom{r+x-1}{r} = (-1)^r \Gamma(x+r)/\Gamma(x)\,\Gamma(r+1)$.

1.3.3 The beta function

The beta function is defined by

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt, \quad x > 0,\ y > 0, \tag{1.3.3.1}$$

or, in terms of gamma functions, as

$$B(x, y) = \Gamma(x)\,\Gamma(y)/\Gamma(x+y). \tag{1.3.3.2}$$

By putting $t = u/(u+1)$, the function can also be represented by

$$B(x, y) = \int_0^\infty \frac{u^{x+1}}{(u+1)^{x+y}}\,du, \quad x > 0,\ y > 0. \tag{1.3.3.3}$$

1.3.4 The generalized hypergeometric function

The generalized hypergeometric function is defined by the series

$${}_pF_q(x) = {}_pF_q[(a);\,(b);\,x] = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n}\frac{z^n}{n!} \tag{1.3.4.1}$$

where the symbol (a) denotes the sequence of parameters $a_1, \ldots, a_p$ and (b) that of $b_1, \ldots, b_q$ while

$$(a)_x = \prod_{i=1}^{p} [\Gamma(x + a_i)/\Gamma(a_i)]$$

and

$$(b)_x = \prod_{j=1}^{q} [\Gamma(x + b_j)/\Gamma(b_j)],$$

with the conditions that $a_i \neq 0$ and $b_j > 0$.

The convergence of ${}_pF_q(x)$ is dependent on the following conditions:

(a) If $p < q$, the series converges for all finite values of its argument, real or complex.

(b) If $p = q + 1$, the series converges

(i) for $|x| < 1$ for all values of its parameters

(ii) for $x = 1$ provided $\mathrm{Re}(\Sigma b_j - \Sigma a_j) > 0$, and

(iii) for $x = -1$ provided $\mathrm{Re}(\Sigma b_j - \Sigma a_j) > -1$.

(c) If $p > q + 1$, the series only converges in the trivial case when $x = 0$ and is only a formal representation of the generalized hypergeometric function concerned. However, if one or more of the numerator parameters is a negative integer, then the series is a terminating series and the question of convergence does not arise.

Left-hand truncation (head truncation) of the first k terms yields

$$\frac{(a_1)_k \ldots (a_p)_k}{(b_1)_k \ldots (b_q)_k} \frac{z^k}{k!} {}_{p+1}F_{q+1}\left(1, (a_i + k); 1 + k, (b_j + k); z\right) \qquad (1.3.4.2)$$

The following integrals pertaining to generalized hypergeometric distributions are used extensively in the study of mixtures of distributions:

$${}_{p+1}F_{q+1}\left(c, (a_i); d, (b_j); z\right)$$

$$= \frac{\Gamma(d)}{\Gamma(c)\,\Gamma(c-d)} \int_0^1 t^{c-1}(1-t)^{d-c-1} {}_pF_q\left((a_i); (b_j); zt\right)dt, \qquad (1.3.4.3)$$

$${}_{p+1}F_q\left(c, (a_i); (b_j); z\right)$$

$$= \frac{1}{\Gamma(c)} \int_0^\infty e^{-y} y^{c-1} {}_pF_q\left((a_i); (b_j); zy\right)dy, \qquad (1.3.4.4)$$

and

$${}_pF_q\left((a_i); (b_j); \frac{z}{s}\right)$$

$$= \frac{s^c}{\Gamma(c)} \int_0^\infty e^{-sx} x^{c-1} {}_pF_q\left((a_i); c, (b_j); zx\right)dx \qquad (1.3.4.5)$$

(see Erdelyi (1954, pp.297-298, 337, 365)).

1.3.5 The modified Bessel functions

A solution of the differential equation, ε_1,

$$z^2 \frac{d^2y}{dz^2} + z \frac{dy}{dz} + (z^2 - \alpha^2)y = 0 \qquad (1.3.5.1)$$

is given by

$$J_\alpha(z) = \sum_{m=0}^{\infty} \left[(-1)^m \left(\frac{z}{2}\right)^{\alpha+2m}\right] \Big/ \left[m!\ \Gamma(\alpha + m + 1)\right], \quad z \neq 0 \qquad (1.3.5.2)$$

The differential equation ε_1 is called the Bessel equation and $J_\alpha(z)$ is commonly referred to as the Bessel function of order α, although other solutions of ε_1 exist (e.g. the Hankel function $Y_\alpha(z)$, sometimes regarded as the Solution II, and the Nielson functions $H_\alpha^{(1)}(z)$ and $H_\alpha^{(2)}(z)$ jointly regarded as the Solution III).

Another differential equation, ε_2 say, similar to ε_1 is

$$z^2 \frac{d^2y}{dz^2} + z \frac{dy}{dz} - (z^2 + \alpha^2)y = 0 \qquad (1.3.5.3)$$

One of its solutions is

$$I_\alpha(z) = \sum_{m=0}^{\infty} \left[\left(\frac{z}{2}\right)^{\alpha+2m}\right] \Big/ \left[m!\ \Gamma(\alpha + m + 1)\right] \qquad (1.3.5.4)$$

and is related to $J_\alpha(z)$ by

$$I_\alpha(z) = e^{-\frac{1}{2}\alpha\pi i} J_\alpha(iz) \qquad (1.3.5.5)$$

but $I_\alpha(z)$ is a real function of z. It is called the modified Bessel function of the first kind of order α and argument z.

Another solution of ε_2 is given by $I_{-\alpha}(z)$. Note however that $I_{-n}(z) = I_n(z)$ for integral n. In order to obtain a fundamental system of solutions for ε_2, the function $K_\alpha(z)$ has been defined for all α, as

$$K_\alpha(z) = [\pi/(2\sin\alpha\pi)][I_{-\alpha}(z) - I_\alpha(z)]. \qquad (1.3.5.6)$$

This is called the Macdonald function, or more commonly, the modified Bessel function of the second kind, of order α and argument z. Because

$$K_\alpha(z) = \tfrac{1}{2}\pi i e^{\frac{1}{2}\alpha\pi i} H_\alpha^{(1)}(iz), \qquad (1.3.5.7)$$

some authors call $K_\alpha(z)$ the modified Bessel function of the third kind, in the sense that it is a modification of the Nielson function $H_\alpha^{(1)}(z)$ which is one of the two forms of the Solution III to the Bessel equation.

1.4 Families of Discrete Distributions

1.4.1 Introduction

By a family or class we mean simply a collection of distributions each of which can be put in a given mathematical structure. The given mathematical structure delineates the family. However it is possible, and is indeed usual, for a particular distribution to belong to many families.

Increasing attention is being given to derivation of broad classes or families, each including as wide a variety of distributions as possible within as specific a formulation as possible. This removes the need for piecemeal derivations and represents considerable economy of effort. It is also resulting in the discovery of interesting relationships between distributions previously thought to be widely dissimilar.

The major classes of discrete distributions on $x = 0, 1, \ldots$ are briefly described below. In general, applications are left for the next section.

1.4.2 Power Series and Generalized Power Series Distributions

A large number of mathematical functions can be expanded in series having positive terms. By multiplying these terms by a suitable constant and provided that the series is convergent, one can make the terms sum to unity and thus produce a discrete distribution. For example, the Poisson distribution is obtained from the series expansion of $e^{\lambda x}$, normalized by $e^{-\lambda}$. This approach provides, in particular, the wide class of discrete distributions called the power-series distributions (PSDs). They are represented by the probability function

$$f(x) = a_x \beta^x / A(\beta), \qquad x = 0, 1, \ldots \tag{1.4.2.1}$$

where $a_x \equiv a(x) \geq 0$, $A(\beta) = \Sigma a_x \beta^x$ or, if there are other parameters $\underline{\alpha}$,

$a_x \equiv a(x, \underline{\alpha})$ and $A \equiv A(\beta, \underline{\alpha})$. Their pgf is given by

$$G(z) = A(\beta z)/A(\beta). \qquad (1.4.2.2)$$

The first discussion of the PSD class of discrete distributions seems to be that given by Noack (1950), although Tweedie (1947) gives it a mention. Noack obtains recurrence relations for the moments and the cumulants and also points out that the Poisson, negative binomial, logarithmic, and Polya distributions are PSDs. Khatri (1959) presents recurrence relations for the cumulants and the factorial cumulants and uses these to show that a single-parameter PSD is uniquely determined by its first two moments.

The concept of a PSD has been extended by Patil (1962) to a class allowing the support of the random variable to be any countable subset of the set of real numbers. Distributions in this broader class are called the generalized power-series distributions (GPSDs). The GPSD class is the discrete analogue of the single-parameter exponential family of continuous distributions. Properties of GPSDs are set out in Patil (1962). See also Ord (1972, pp.116-121).

1.4.3 Difference-equation distributions

The difference-equation distributions (DEDs) are discrete distributions whose successive probabilities satisfy relations of the form

$$\frac{P_{x+1}}{P_x} = \frac{u(x)}{v(x)} \qquad (1.4.3.1)$$

or, equivalently,

$$\frac{P_x}{P_{x-1}} = \frac{g(x)}{h(x)} \qquad (1.4.3.2)$$

i.e. $$\Delta P_x = \frac{g(x) - h(x)}{g(x)} P_x$$

where $u(x)$, $v(x)$, $g(x)$, and $h(x)$ are polynomials and ΔP_x $(\equiv P_x - P_{x-1})$

denotes the difference operation. Note that, in general, the support is not restricted to the set of non-negative integers (e.g. Ord (1968)); however, only the subclass taking values on $x = 0, 1, \ldots$ is of interest in the present study.

DEDs have been known in the literature for a very long time. For example, Pearson (1895) notes that the hypergeometric distribution is a DED and uses this as a starting point for obtaining, by a limiting process, the differential equation defining the Pearson system of continuous distributions. Guldberg (1931) points out that recurrence relations can be established between moments of any discrete distributions belonging to the system defined by

$$\frac{P_{x+1}}{P_x} = \frac{\lambda(x + \alpha_1) \ldots (x + \alpha_m)}{(x + \beta_1) \ldots (x + \beta_n)} . \qquad (1.4.3.3)$$

The generalized hypergeometric probability distributions (ghpd's) discussed in the next section (section 1.4.4) are all DEDs since they are represented by

$$\frac{P_{x+1}}{P_x} = \frac{\lambda}{x + 1} \frac{(x + a_1) \ldots (x + a_p)}{(x + b_1) \ldots (x + b_q)} \qquad (1.4.3.4)$$

$$\equiv u(x)/v(x)$$

with $u(x) = \lambda(x + a_1) \ldots (x + a_p)$ and $v(x) = (x + 1)(x + b_1) \ldots (x + b_q)$ which are polynomials.

1.4.4 Generalized hypergeometric distributions

The generalized hypergeometric series defined in Section 1.3.4 has positive coefficients. By normalizing these and outlining the convergence conditions, Kemp (1968a) has defined a wide class of lattice distributions which are known as generalized hypergeometric probability distributions (ghpd's). They have pgfs of the form

$$G(z) = {}_pF_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;\lambda z)/{}_pF_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;\lambda),$$
$$= {}_pF_q((\alpha_i);(\beta_j);\lambda z)/{}_pF_q((\alpha_i);(\beta_j);\lambda) \qquad \alpha_i \neq 0,\ \beta_j > 0$$
$$= C\,{}_pF_q((\alpha_i);(\beta_j);\lambda z) \qquad (1.4.4.1)$$

where the factor C here, and in the pgf of any ghpd's to be mentioned subsequently, denotes the appropriate normalising constant. It is obvious from the form of their pgfs that they are power series distributions, and also since the ratio of their successive probabilities is given by

$$\frac{P_{x+1}}{P_x} = \frac{\lambda}{x+1}\cdot\frac{(x+\alpha_1)\ldots(x+\alpha_p)}{(x+\beta_1)\ldots(x+\beta_q)},\quad x = 0, 1,\ldots, \qquad (1.4.4.2)$$

they are difference-equation distributions. Most discrete distributions can be represented in this form or in forms related to it, especially by truncation, shifting, reversing, or generalizing.

Kemp (1968a) also presents the differential equations from which recurrence relations for probabilities, moments, factorial moments, cumulants, and factorial cumulants can be obtained, and in addition considers urn models and limiting distributions. A further description of ghpd's has been given by Dacey (1972) who proposes a method for identifying the constituent members of the class and also lists the parameter ranges of the different subsystems with pgfs

$$G(z) = C_pF_q(\lambda z),\ p+q = 3,\ p, q = 0, 1, 2, 3. \qquad (1.4.4.3)$$

Guldberg's (1931) system of discrete distributions is represented by

$$\frac{P_{x+1}}{P_x} = \frac{(x+a_1)\ldots(x+a_m)}{(x+b_1)\ldots(x+b_n)}$$
$$= \frac{\lambda}{x+1}\,\frac{(x+1)(x+\alpha_2)\ldots(x+\alpha_p)}{(x+\beta_1)\ldots(x+\beta_q)}. \qquad (1.4.4.4)$$

Hence its subsystem supported on $x = 0, 1,\ldots$ with $\alpha_i \neq 0$, $\beta_j > 0$ is comprised of ghpd's having pgfs of the form

$$G(z) = {}_pC_qF_q(1, \alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q; \lambda z). \qquad (1.4.4.5)$$

Ord (1967a) discusses a system defined by a difference equation which can be equivalently written as

$$\frac{P_{x+1}}{P_x} = \frac{(x + a_1)(x + a_2)}{(x + b_1)(x + b_2)}$$

$$= \frac{1}{(x + 1)} \cdot \frac{(x + 1)(x + \alpha_2)(x + \alpha_3)}{(x + \beta_1)(x + \beta_2)} . \qquad (1.4.4.6)$$

Its lattice subsystem with $\alpha_2, \alpha_3 \neq 0$ has the pgf

$$G(z) = {}_3C_2F_2(1, \alpha_2, \alpha_3; \beta_1, \beta_2; z) \qquad (1.4.4.7)$$

if $\beta_1 + \beta_2 > \alpha_2 + \beta_3 + 1$. Ord presents a graphical method for distinguishing between the distributions in the system.

Katz (1948) refers to his work on the discrete distributions defined by

$$\frac{P_{x+1}}{P_x} = \frac{\lambda(x + a)}{(x + b_1)(x + b_2)}, \quad x = 0, 1, \ldots$$

$$= \frac{\lambda}{(x + 1)} \cdot \frac{(x + 1)(x + \alpha_2)}{(x + \beta_1)(x + \beta_2)} . \qquad (1.4.4.8)$$

The cases with $a \neq 0$ have pgfs given by

$$G(z) = {}_2C_2F_2(1, \alpha_2; \beta_1, \beta_2; \lambda z). \qquad (1.4.4.9)$$

In a study of birth-and-death processes, Hall (1956) obtains a system of lattice distributions defined by the pgf

$$G(z) = {}_2C_2F_1(\alpha_1, \alpha_2; \beta; \theta z), \qquad |\theta| \leq 1 \qquad (1.4.4.10)$$

The distributions in the system can be called the general Gaussian hypergeometric distributions. Tripathi and Gurland (1977) obtain minimum chi-square estimators for the parameters of the special cases having pgfs of the form

$$G(z) = {}_2C_2F_1(1, \alpha; \beta; \theta z), \qquad |\theta| \leq 1. \qquad (1.4.4.11)$$

Kemp and Kemp (1956a) have listed out the constituent members of the system having pgfs of the form

$$G(z) = {}_2F_1(\alpha_1, \alpha_2; \beta; z). \qquad (1.4.4.12)$$

These are called the Gaussian hypergeometric probability distributions (Ghpd's), a name obvious from the form of the pgfs.

A subsystem of the ghpd's comprising distributions which can be called the general confluent hypergeometric probability distributions is included among Hall's (1956) ghpd's arising from birth-and-death processes. The distributions in this subsystem have pgfs of the form

$$G(z) = {}_1F_1(\alpha; \beta; \lambda z). \qquad (1.4.4.13)$$

Bhattacharya (1966) discusses this subsystem and derives a mixture on the parameter λ with a continuous distribution called 'the general exponential' as the mixing distribution. Gurland and Tripathi (1975) present some minimum-chi-square estimators for parameters of the general confluent hypergeometric probability distributions. Bradwell and Crow (1964) have given the name 'hyper-Poisson distributions' to the special cases with the pgfs

$$G(z) = {}_1F_1(1; \beta; \lambda z). \qquad (1.4.4.14)$$

Those with $\lambda < 1$ are said to be 'sub-Poisson' and those with $\lambda > 1$ 'super-Poisson'; the case $\lambda = 1$ being equivalent to the Poisson distribution. This division corresponds to whether or not the variance exceeds the mean (Ord (1972, p.92)). Staff (1964) also independently discusses the hyper-Poisson distributions.

The system of lattice distributions represented by

$$\frac{P_{x+1}}{P_x} = \frac{\lambda}{(x+1)}(x+\alpha), \quad x = 0, 1, \ldots \qquad (1.4.4.15)$$

has engaged the attention of Katz (1946, 1948, 1965). The cases with $\alpha \neq 0$ have pgfs of the form

$$G(z) = {}_1F_0(\alpha; -; \theta z), \quad |\theta| < 1. \qquad (1.4.4.16)$$

They can be called the general binomial distributions in a wider sense than Khatri and Patel's (1961). They comprise the (positive) binomial,

the negative binomial, and Johnson and Kotz's (1969, p.41) set of terminating distributions akin to the (positive) binomial. Katz (1965) suggests using a plot of P_{x+1}/P_x against x to distinguish between the constituent members of his system.

Another class of generalized hypergeometric distributions has been defined by Kemp and Kemp (1974) who also give differential equations for obtaining recurrence formulae for probabilities, cumulants, and moments of the distributions in the class. The distributions have pgfs of the form

$$G(z) = {}_pF_q((\alpha_i); (\beta_j); \lambda(z-1)) \qquad (1.4.4.17)$$

and have been called the generalized hypergeometric factorial-moment distributions (ghfd's) because their factorial moments are given by the terms of the generalized hypergeometric function i.e.

$${}_pF_q((\alpha_i); (\beta_j); \lambda t) = G(1+t), \qquad (1.4.4.18)$$

G(1 + t) being the factorial-moment generating function. Some examples are given in an earlier paper of the same authors (Kemp and Kemp (1969)). Models giving rise to some ghfd's have been discussed by Kemp and Kemp (1975).

A broad class subsuming the ghpd's and the ghfd's is that of the generalized hypergeometric recast distributions (ghrd's) (Kemp (1974)) having the pgf

$$G(z) = {}_pF_q((\alpha_i); (\beta_j); \lambda z + \gamma)/{}_pF_q((\alpha_i); (\beta_j); \lambda + \gamma) \qquad (1.4.4.19)$$

Examples of ghrd's with $\gamma \neq 0$ are Kemp's (1968b) limited risk compound-Poisson-process distribution and Gurland and Tripathi's (1975) ECB distribution.

1.4.5 Compound or mixed distributions

The definition of a compound distribution is embodied in the following theorem which has been stated in many versions in the literature (e.g. Feller (1943), Teicher (1960), Blischke (1965)).

Theorem 1

Consider an arbitrary cdf $G(x; \alpha, \underline{\beta})$ depending on a parameter α ($\in \Lambda$) and possibly, on other parameter(s) $\underline{\beta}$. Suppose there is another distribution $H(\alpha; \underline{\gamma})$ supported on $\Lambda_0 \subseteq \Lambda$ and depending on the parameter(s) $\underline{\gamma}$. Then the function

$$F(x) \equiv F(x; \underline{\beta}, \underline{\gamma}) = \int_{\Lambda_0} G(x; \alpha, \underline{\beta})\, dH(\alpha; \underline{\gamma}) \qquad (1.4.5.1)$$

is again a cdf; provided that the integral is convergent. This integral is defined in the Stieltjes sense.

The distribution F(x) in the theorem is called a compound distribution, or more commonly, a mixture. Only the cases in which F(x) is a discrete distribution are of interest in this study. These arise when $G(x; \alpha, \underline{\beta})$ is a discrete distribution but with discrete or continous $H(\alpha; \underline{\gamma})$. They can be considered to be of three forms according as to whether $H(\alpha; \underline{\gamma})$ is a continuous distribution, a finite (or terminating) discrete distribution, or a (truncated or untruncated) non-terminating discrete distribution. The latter two types of F(x) are called finite mixtures and countable mixtures respectively. In either type, F(x) is given by

$$F(x) = \Sigma p_i\, G(x; \alpha_i) = \Sigma p_i\, G_i(x) \qquad (1.4.5.2)$$

where the $G_i(x)$ relate to the mixed distribution while the p_i, the $H(\alpha)$ probabilities, can be regarded as the proportions of the mixing.

Referring to the designations of the cdfs in the theorem, the distribution F(x) is called an H-mixture of G or a compound-G distribution while $H(y; \underline{\gamma})$ is the compounder or the mixing distribution. Gurland (1957) has proposed a notation which is now well-established in the literature. In this notation and with X_F, X_G, Y_H referring respectively to the rvs of the distributions $F(x; \underline{\beta}, \underline{\gamma})$, $G(x; \alpha, \underline{\beta})$, $H(y; \underline{\gamma})$, the

above relationship is represented as

$$X_F \equiv X_G(\alpha) \underset{\alpha}{\Lambda} Y_H \qquad (1.4.5.3)$$

where X_G is the mixed, Y_H is the mixer or compounder and α is the parameter of the distribution of X_G on which the mixing is performed.

One of the principal concepts upon which many applications of mixture is based is that of conditionality. In many problems the distribution of a random variable X is known conditionally, given the value of Y a related random variable, whereas the unconditional distribution of X is of interest. This unconditional distribution is simply a weighted average, i.e. a mixture, of the known conditional distribution.

Among discrete mixtures those of the Poisson distribution are the commonest, although mixtures of other classical discrete distributions have also been developed. It should be noted, however, that the issue of identifiability often severely restricts the number of distinct ones out of the conceptually possible mixtures of a given distribution.

The question of identifiability of mixtures concerns their unique characterization. Before estimation or hypothesis testing for a mixture can be meaningfully discussed, it must be known whether the mixture is uniquely determined; that is whether the parameters are 'identified'. Failing this, estimation may be impossible.

Identifiability for mixtures of the Poisson distribution has been shown by Feller (1943). It is also well known (e.g. Blischke (1964)) that mixing the binomial or the negative binomial distribution on the exponent parameter results in a class of mixtures which is identifiable. Note that in either case, if the mixing distribution is the Poisson, a generalized Poisson distribution results (Section 1.4.6 below). If, however, the binomial or the negative binomial is being mixed on the non-exponent parameter, only a restricted set of mixing distributions produce identifiable mixtures (Teicher (1960, 1961), Boes (1963), and Blischke (1964, 1965)).

Compound Poisson distributions are discussed further in Chapter Two, particularly in Sections 2.3 and 2.4. Besides compound Poisson distributions other discrete mixtures mentioned in the literature include compound hypergeometric distribution (Hald (1960)), the beta-binomial distribution which is Kemp and Kemp's Gaussian hypergeometric Type IIIA (Kemp and Kemp (1956a)) having the pgf

$$G(z) = {}_2F_1(-n, \alpha; -n - \beta + 1; z) \qquad (1.4.5.4)$$

and represented, in Gurland's notation, by

$$\text{Binomial}(x; n, \theta) \underset{\theta}{\wedge} \text{Beta-I}(\theta; \alpha, \beta), \qquad (1.4.5.5)$$

and the beta-Pascal distribution which is Kemp and Kemp's Type IV, otherwise known as the generalized Waring distribution (Kemp and Kemp (1956a), Irwin (1968)) with the pgf

$$G(z) = {}_2F_1(\alpha, \beta; \alpha + \beta + \gamma; z) \qquad (1.4.5.6)$$

and is the mixing on θ of the negative binomial $h_1(z) = [(1 - \theta)/(1 - \theta z)]^{\alpha}$ by the Beta-I(θ; α, β) or the negative binomial $h_2(z) = (1 + \lambda - \lambda z)^{-\alpha}$ by the Beta-II(λ; α, β). Another example of compound distributions is Subrahmaniam's (1966) distribution which is the mixing on θ of the Neyman Type A having the pgf $h(z) = \exp[\lambda\theta(e^{\gamma(z-1)} - 1)]$ by the Beta-I(θ; α, β) i.e.

$$\text{NTA}(x; \lambda\theta, \gamma) \underset{\theta}{\wedge} \text{Beta-I}(\theta; \alpha, \beta). \qquad (1.4.5.7)$$

It is important to note that when considering the mixing of a distribution, it is crucial to know which particular parameterization one is dealing with, because mixtures of different parameterizations of a given distribution (even a one-parameter distribution such as the geometric) are, in general, not identical. It seems that the introduction of an arbitrary scaling of the mixed parameter into the definition of a mixture (e.g. Feller (1943)) does not remove the need to clearly state the parameterization. Johnson and Kotz (1969) in their discussion of mixtures of discrete distributions give the following type of representations for compound hypergeometric distributions

$$\text{Hypergeom}(x;\, r,\, m,\, n) \underset{r}{\Lambda} \text{Binom}(r;\, m,\, \theta) \equiv \text{Binom}(x;\, n,\, \theta) \qquad (1.4.5.9)$$

$$\text{Hypergeom}(x;\, r,\, k,\, \ell) \underset{r}{\Lambda} \text{Hypergeom}(r;\, k,\, m,\, n) \equiv \text{Hypergeom}(x;\, \ell,\, m,\, n) \qquad (1.4.5.10)$$

but in the case of mixtures of the Poisson, the binomial, and the negative binomial they use the following types of representation:

$$\text{Poisson}(x;\, \alpha m) \underset{m}{\Lambda} D(m;\, \beta) \equiv \text{Poisson-D}(x;\, \alpha,\, \beta) \qquad (1.4.5.11)$$

$$\text{Binom}(x;\, mn,\, \theta) \underset{m}{\Lambda} D(m;\, \alpha) \equiv \text{Binom-D}(x;\, n,\, \alpha,\, \theta) \qquad (1.4.5.12)$$

and

$$\text{Neg Binom}(x;\, m\alpha,\, \beta) \underset{m}{\Lambda} D(m;\, \lambda) \equiv \text{Neg Binom-D}(x;\, \alpha,\, \beta,\, \lambda). \qquad (1.4.5.13)$$

It is also noted that some mixtures have been obtained by mixing unusual parameterizations of common distributions. For example, the well known Yule distribution, as originally obtained, is

$$\text{Geometric}\left(x;\, e^{-\alpha\beta}\right) \underset{\alpha}{\Lambda} \text{Exponential}(\alpha;\, \lambda) \equiv \text{Yule}(x;\, \beta\lambda) \qquad (1.4.5.14)$$

(see e.g. Johnson and Kotz (1969, pp.245-246)).

1.4.6 Generalized distributions

A substantial part of probability theory is connected with sums of independent random variables $S_n = X_1 + \ldots + X_n$, and the special case where the X_i's have a common distribution is of particular interest. In many situations the number, n, of terms in such sums is itself a random variable. This latter type of sum gives rise to the wide class of discrete distributions commonly called the generalized distributions (e.g. Feller (1943), Gurland (1957), Katti (1966)), otherwise known as the (true) contagious distributions (e.g. Sprott (1965), Ord (1972, p.66)), or randomly stopped sums (Douglas (1970, 1980)), or simply random sums (Chatfield and Theobald (1973)).

Formally, let $X_1, X_2, \ldots$ be a sequence of mutually independent lattice variables with a common distribution having the pgf u(z). We are interested

in the sums

$$S_N = X_1 + X_2 + \ldots + X_N \qquad (1.4.6.1)$$

where the number N of terms is a random variable independent of the X_i's and having the distribution with the pgf

$$v(z) = \Sigma p_k s^k. \qquad (1.4.6.2)$$

Note that the distribution of the conditional variable $S_N | N = n$ is an n-fold convolution of X the common distribution of the X_i and hence has the pgf $h(z) = [u(z)]^n$. Therefore the pgf of the unconditional random variable S_N is given by

$$\begin{aligned} G(z) &= \Sigma p_n [u(z)]^n \\ &= v\{u(z)\}. \end{aligned} \qquad (1.4.6.3)$$

A convenient notation has been proposed by Gurland (1957) to represent the relation between the generalized distribution and its components. Let the random variables X_1, X_2 have the pgfs $g_1(z)$, $g_2(z)$ respectively, and let the random variable with the pgf $g_1\{g_2(z)\}$ be denoted by $X_1 \vee X_2$. Then $X_1 \vee X_2$ is called a generalized-X_1 variable with respect to the generalizer X_2. Note that the generalizing operation "V" is not commutative, in general. In this notation the above defining expression can therefore be represented as

$$S_N \equiv N \vee X. \qquad (1.4.6.4)$$

In the generalizing operation, all the parameters of the generalizer and of the distribution being generalized are preserved. Therefore the number of the parameters of a generalized distribution is the sum of those of its components. This implies that a generalized distribution is, in general, a multi-parameter distribution. This feature is often a disadvantage when application is concerned since the fewer the parameters to be estimated in a distribution the better it is for most practical purposes. This difficulty of multiplicity of parameters is often tackled by having for either of the two component distributions a special case of

a given distribution. For example, in the application of the Poisson-binomial (i.e. Poisson(λ) V Binomial(n, θ)) it is usual to specify the value of n in advance.

Katti (1966) has given some relations between the factorial moments and the factorial cumulants of generalized distributions and those of their components. Khatri and Patel (1961) present recurrence relations for the probabilities of the generalized Poisson, the generalized binomial, the generalized negative binomial, and the generalized logarithmic series distributions respectively. They also give some relations between the factorial cumulants of these generalized distributions and those of their respective components.

Among generalized distributions, the generalized Poisson distributions (GPDs) are the commonest. They are the subject of this thesis and their main properties are discussed in the next chapter. Other generalized distributions specifically discussed in the literature include Khatri and Patel's (1961) General binomial V General Binomial where by a general binomial distribution they mean a distribution which is either a binomial distribution or a negative binomial distribution. Their distribution therefore comprises

Binomial V Binomial

Binomial V Negative binomial

Negative binomial V Binomial, and

Negative binomial V Negative binomial

each of which is, in general, a four-parameter distribution. Hence only special cases would normally find applications. Johnson and Kotz (1969) give brief descriptions of Binomial V Poisson, Negative binomial V Poisson, Binomial V log series, Negative binomial V log series, and Log Series V log series. Katti and Rao's (1970) log-zero-Poisson distribution is Log Series-with-added-zeroes V Poisson, while Subrahmaniam's (1966) distribution is obtained by generalizing a generalized hypergeometric

factorial-moment distribution with the pgf ${}_1F_1(\alpha;\ \alpha + \beta;\ \lambda(z - 1))$ by a Poisson distribution.

1.4.7 The Lagrangian probability distributions

By exploiting certain features of the well-known Lagrange's formula for series expansion, Consul and Shenton (1972) have defined two related classes of discrete distributions, the basic Lagrangian probability distributions (BLPDs) and the general Lagrangian probability distributions (GLPDs). Members of the first take values on $x = 1, 2,\ldots$ while those of the other are supported on $x = 0, 1,\ldots$ These new systems of distributions have produced a flourishing body of literature, as evidenced in Johnson and Kotz (1982).

Let $u(s)$ be any pgf defined on some or all non-negative integers such that $u(s) \neq 0$. Then the transformation

$$z = s/u(s) \qquad (1.4.7.1)$$

defines, for the numerically smallest nonzero root s, a new pgf $s \equiv g(z)$ whose expansion in powers of z is given by the Lagrange's formula as

$$s \equiv g(z) = \sum_{x=1}^{\infty} \frac{z^x}{x!} \frac{d^{x-1}}{ds^{x-1}} \{[u(s)]^x\}\Big|_{s=0}. \qquad (1.4.7.2)$$

Its probability function is given by

$$P_x = \frac{1}{x!} \frac{d^{x-1}}{ds^{x-1}} \{[u(s)]^x\}\Big|_{s=0}, \quad x = 1, 2,\ldots \qquad (1.4.7.3)$$

This distribution is called the basic Lagrangian probability distribution with respect to the distribution with the pgf $u(s)$ and can be denoted by $L_0\{u(s)\}$ e.g. $L_0(\text{Poisson}(\lambda)) \equiv \text{Borel}(\lambda)$. Precedence on this type of distribution actually goes to Dwass (1967) and Steutel (1970, p.13).

Suppose $u(s)$ and $v(s)$ are two given pgf's defined on some or all non-negative integers and such that $u(s) \neq 0$. Then the transformation

$$z = s/u(s) \qquad (1.4.7.4)$$

defines, for the smallest non-zero root s, a new pgf $v(s) \equiv h(z)$ whose Lagrangian expansion in powers of z is given by

$$h(z) = v(0) + \sum_{x=1}^{\infty} \frac{z^x}{x!} \frac{d^{x-1}}{ds^{x-1}} \{[u(s)]^x v'(s)\}\Big|_{s=0}. \qquad (1.4.7.4)$$

Its probability function is given by

$$P_0 = v(0)$$

$$P_x = \frac{1}{x!} \frac{d^{x-1}}{ds^{x-1}} \{[u(s)]^x v'(s)\}\Big|_{s=0}, \quad x = 1, 2,... \qquad (1.4.7.5)$$

The distribution is called the general Lagrangian probability distribution with respect to the ordered pair of pgfs $\{u(s), v(s)\}$ and is denoted by $L\{u(s), v(s)\}$ e.g. $L\{\text{Binomial}(m, \theta), \text{Binomial}(n, \theta)\}$ which is Jain and Consul's (1971) distribution with the probability function

$$P_x = \left\{\frac{n}{n + mx}\right\}\binom{n + mx}{x}\theta^x(1 - \theta)^{n+(m-1)x}, \quad x = 0, 1,... \qquad (1.4.7.7)$$

and $L\{\text{Poisson}(\theta), \text{Poisson}(\lambda)\}$ which is Consul and Jain's (1971) distribution having the probability function

$$P_x = \lambda(\lambda + \theta x)^{x-1} e^{-(\lambda+\theta x)}/x!, \quad x = 0, 1,... \qquad (1.4.7.8)$$

The latter can be regarded as a general Borel-Tanner distribution.
The distribution resulting from $L\{\text{Neg binomial}(\alpha, P), \text{Neg binomial}(\beta, P)\}$ has the pf

$$P_x = \frac{\beta}{\beta + (\alpha + 1)x} \frac{\Gamma((\alpha + 1)x + \beta + 1)}{x!\,\Gamma(\alpha x + \beta + 1)} \left\{\frac{P}{Q}\right\} Q^{-(\alpha x+\beta)} \qquad (1.4.7.9)$$

The structure and properties of the Lagrangian distributions are detailed in Consul and Shenton (1972), Jain (1974), and Consul (1975). Information on these distributions is, however, far from being complete. The defining conditions state that any pgf u(s) would yield a BLPD and any ordered pair $\{u(s), v(s)\}$ a GLPD. However Nelson (1975) has pointed out that there are cases when restrictions are required on the ranges of the parameters of the component distributions to obtain a Lagrangian distribution which is a proper distribution. Consul and Shenton (1975)

themselves note that L(Poisson(α), Poisson(β)) is a proper distribution only if $0 < \alpha < 1$. Furthermore, it may be observed that the successive probabilities of Jain and Consul's L(Binomial(m, θ), Binom(n, θ)) mentioned above is represented by

$$\frac{P_{x+1}}{P_x} = \frac{\lambda}{x+1} \frac{(x+1)(x+m)(x+\alpha_1) \ldots (x+\alpha_{m-1})}{(x+\beta_1) \ldots (x+\beta_{m-1})} \qquad (1.4.7.10)$$

where $\lambda \equiv (m\theta)\left[\left(\frac{m}{m-1}\right)(1-\theta)\right]^{m-1}$, $\alpha_i = (n+i)/m$, and $\beta_j = (n+j)/(m-1)$. This implies that, in generalized hypergeometric form, the pgf is represented by

$$G(z) = {}_{m+1}F_{m-1}(1, m, \alpha_1, \ldots, \alpha_{m-1}; \beta_1, \ldots, \beta_{m-1}; \lambda z) \qquad (1.4.7.11)$$

which is convergent only in restricted cases (see Section 1.3.4). It definitely cannot be non-terminating.

1.4.8 Binomial generalization distributions

A number of new distributions have been defined based on structures which are, or are related to, generalizations of the binomial expansion. These distributions can therefore be called binomial generalization distributions.

From certain urn models Consul (1974) has developed some distributions which are related to Abel's generalizations (e.g. Riordan (1979)) of the binomial expansion. These are called Abel distributions (Johnson and Kotz (1982)). They include the quasibinomial distribution-I with the pf

$$P_x = \frac{\alpha}{(\alpha+\beta+n\lambda)^n} \binom{n}{\alpha} (\alpha+\lambda x)^{x-1} [\beta+\lambda(n-x)]^{n-x}, \quad x = 0, 1, \ldots, n, \qquad (1.4.8.1)$$

the quasibinomial distribution-II with pf

$$P_x = \frac{\alpha\beta}{(\alpha+\beta)} \binom{n}{x} (\alpha+\lambda x)^{x-1} [\beta+\lambda(n-x)]^{n-x-1} / (\alpha+\beta+\lambda n)^{n-1}, \quad x = 0, 1, \ldots, n, \qquad (1.4.8.2)$$

the quasihypergeometric distribution with the probability function

$$P_x = \frac{(\alpha\beta)(\alpha + \beta + \lambda n)}{(\alpha + \beta)(\alpha + \lambda x)[\beta + \lambda(n - x)]} \binom{\alpha + \lambda x}{x} \binom{\beta + \lambda(n - x)}{n - x} \Big/ \binom{\alpha + \beta + \lambda n}{n},$$

$$x = 1, \ldots, n, \quad (1.4.8.3)$$

and the quasi-Polya distribution with the pf

$$P_x = \binom{n}{x} \alpha\{\alpha + \lambda(x - 1)\}^{[x-1]} \{\beta + \lambda(n - x)\}^{[n-x]} / (\alpha + \beta + \lambda n)^{[n]},$$

$$x = 0, 1, \ldots, n \quad (1.4.8.4)$$

where the parameters α, β, λ can be any positive number although proper distributions are also obtained with some other sets of values of these parameters.

Another system of binomial generalization distributions has been proposed by Dwass (1979) who observes that

$$(x + y)^{(n)} = \sum_{0 \leq i \leq n} \binom{n}{i} x^{(i)} y^{(n-i)} \quad (1.4.8.5)$$

where $u^{(r)} = u(u - 1) \ldots (c - \alpha(r - 1))$, α being any arbitrary real number satisfying $\alpha(n - 1) \leq u$. This expression provides a distribution defined by the probability function

$$P_x = \binom{n}{x} a^{(x)} b^{(n-x)} / (a + b)^{(n)}, \quad x = 0, 1, \ldots, n \quad (1.4.8.6)$$

provided that $a > 0$, $b > 0$ and $c^{(r)} = c(c - 1) \ldots (c - \alpha(r - 1))$ where α is an arbitrary real number satisfying $\alpha(n - 1) \leq (a + b)$. Dwass (1979) provides a computer program for computing the probabilities and also suggests an application in acceptance sampling.

1.5 Chance Mechanisms Generating Discrete Distributions

1.5.1 Introduction

A discussion of the stochastic method or methods generating a distribution provides information about the nature and scope of its applications.

It is usual, and indeed more common than often realised, for a given distribution to arise from more than one underlying chance

mechanism. The case of the ubiquitous negative binomial is a point in view (see e.g. Boswell and Patil (1970)). The long-drawn controversy over the term "contagion" stems from the exploration of certain distributions by different people using different stochastic models.

It is often rightly emphasized that a distribution should be chiefly a hypothesis-generating device rather than a calibration tool. However a remark commonly made as an addendum to this fact is disputable. This states that a distribution whose genesis is based on a multiplicity of models is of little applicational value since choice among alternative hypotheses becomes complicated. On the other hand, there is no doubt about the importance of the Gaussian distribution despite that it arises from diverse types of situations, a feature for which it earned the name the "normal" distribution.

The pragmatism of Feller (1968, pp.2-3) is possibly worth advocating: "In applications, the abstract mathematical models serve as tools, and different tools can describe the same empirical situation. The manner in which mathematical theories are applied does not depend on preconceived ideas; it is a purposeful technique depending on, and changing with, experience". It should be borne in mind that stochastic models, like all other mathematical models, rest on assumptions and consist of abstractions. It is generally on the validity of the assumptions that the success of applications depends rather than on the uniqueness or non-uniqueness of the models.

1.5.2 Mixing Models

A great number of mixing models are based on the concept of conditionality. Consider for example the spatial distributions of plants or insects. Suppose that the region over which the individuals are dispersed varies in its attractiveness. The probability distribution of the number of individuals per plot depends on a parameter λ corresponding

to the region in which it falls but because of this heterogeneity the parameter λ is itself a random variable, commonly a discrete variable. If the variation in this parameter is described by a distribution with the probability function $f(\lambda)$, we have the marginal distribution of the plot count X given by

$$\begin{aligned} Pr(X = x) &= \int Pr(X = x|\lambda)\ f(\lambda)d\lambda \qquad \text{for continuous } f(\lambda), \text{ or} \\ &= \Sigma Pr(X = x|\lambda = \lambda_i)\ f(\lambda_i) \quad \text{in the discrete case.} \end{aligned} \qquad (1.5.2.1)$$

For example, Neyman (1939) in his discussion of insect populations, suggests taking the distribution of $X|\lambda$ as Poisson$(x; \lambda)$ and that of λ as Poisson$(\lambda; \alpha)$ thus obtaining the distribution of X as Neyman Type A.

In accident studies a similar model is developed for investigating the phenomenon of accident proneness. It is postulated that the number X of accidents sustained by an individual in a given population of equally exposed people is distributed as Poisson$(x; \lambda)$ where λ represents a quantity regarded as a personal factor and generally called proneness, liability, etc., constant for the individual but which when considered over the whole population is a random variable whose distribution is specifiable. For a given individual the value of λ is unaffected by the occurrence of accidents in any previous time-period or by accidents sustained by other individuals. The distribution of accidents in the population is therefore a mixture of the Poisson (see Chapter Two). The proneness model has also been adopted for the study of industrial absenteeism (Philipson (1960), Froggatt (1968, 1970)) and of occasions of sickness resulting in consulting a doctor (Froggatt et al. (1969)).

Dacey (1969) develops Kemp and Kemp's Gaussian hypergeometric Type IV (generalized Waring distribution) as a mixture of the Poisson and discusses an application in a location problem in geographical research. The same author (Dacey (1969, 1970)) has also considered certain finite

KASHIM IBRAHIM LIBRARY
ABDUL SELLO UNIVERSITY
ZARIA, NIGERIA

mixtures of the Poisson distribution in other similar problems.

Compound Poisson processes are a common tool in collective risk theory. In the usual form, a compound Poisson process has probabilities of the form

$$P_n(t) = \int_0^\infty \frac{(xt)^n e^{-xt}}{n!}\, dH(x) \qquad (1.5.2.2)$$

where H(x) is the distribution function of the unconditional risk distribution and t is a time interval (e.g. Philipson (1960, 1961)). The assumption of unlimited risk is the basis for using an unconditional risk distribution H(x) concentrated on the right half real line. Kemp (1968b) has pointed out that the unlimited risk assumption may be unrealistic and has suggested using an upper-bounded H(x) instead, deriving as an example the limited risk compound Poisson process resulting from having a truncated gamma distribution as the mixing distribution H(x).

Market research is another area where the use of mixing models is common (e.g. Chatfield (1970)). The negative binomial has been suggested by Chatfield et al. (1966) for describing the distribution of purchases made by repeat buyers in a stationary market. This is based on the assumption that the purchases made by a particular consumer in successive equal time-periods (usually weeks) are independent and follow a Poisson distribution with a constant mean, and that this average long-run rate of buying will vary from consumer to consumer, following a gamma distribution. Sichel (1982b) has discussed some other mixing distributions.

Chatfield and Goodhardt (1970) have suggested using the beta-binomial distribution to describe the distribution of weeks for which a particular brand or a specific pack-size is bought. They postulate that the probability that a given customer will buy at least one unit of the brand (or pack-size) in a particular week is a constant, p, which is independent of previous purchases. Thus in a time-period of n weeks,

the number of weeks in which the consumer buys at least one unit will follow the binomial distribution with the parameters n and p. It is also assumed that p varies from consumer to consumer, following a beta distribution.

Chatfield (1970) considers consumer purchasing in terms of the time between purchases rather than amounts bought. Consider a person whose probability of buying a particular brand in one week is equal to θ. If this is independent of previous purchases, the probability that he makes his first in the xth week is given by the shifted geometric with pgf $b(z) = z(1 - \theta)/(1 - \theta z)$. If θ varies from consumer to consumer following a beta distribution then the proportion of consumers in the population who make their first purchase in the xth week is given by

$$P_x = \int_0^1 \theta(1 - \theta)^{x-1} \text{Beta}(\theta; \alpha, \beta)d\theta \qquad (1.5.2.3)$$

which is the probability function of the beta-geometric distribution.

1.5.3 Clustering models

Generalized distributions arise mainly from clustering models. Suppose that clumps or "clusters" of objects are observed, such as insects or plants in a field or houses in a study area. Then the models proposed envisage a lattice distribution of such clusters, each cluster containing one or more objects, the number within a cluster following the generalizer distribution, another lattice distribution. Clustering models are also called true contagion models. In so far as the existence of a cluster means that an object is more likely to have other similar objects nearby we may say that these models represent "true contagion". This is in constrast with what obtains in mixing models where the clumping feature is due only to the heterogeneity of the area in which the object resides. Hence mixing models are regarded as "apparent contagion" models.

Most commonly the distribution of the clusters is assumed to be random, that is, Poisson, whereby the overall distribution of the objects is a generalized Poisson distribution. Clusters could follow distributions other than Poisson. However if the distribution of the clusters is a generalized Poisson distribution the overall distribution of the units is itself a generalized Poisson distribution. Generalized Poisson distributions and their genesis are discussed further in Chapter Two.

Studying the application of stochastic models in epidemiology, Neyman and Scott (1964) obtain the equation

$$G(z|u) = g[zh(z|u)|u] \qquad (1.5.3.1)$$

with

$$h(z|u) = \int_H g[zh(z|x)|x]\ f(x|u)dx$$

where $G(z|u)$ is the pgf of the total individuals infected anywhere in habitat H during an epidemic started by a single infectious at u; $g(z|u)$ is the pgf of the number of susceptibles $v(u)$ who would be infected at point u if an infectious is there; $f(x|u)$ is the pdf of the location x of a single first generation infectious; $h(z|u)$ is the unconditional pgf of the total epidemic started by one individual who himself became infectious at u (see also Neyman (1965)). $G(z|u)$, obviously, is the pgf of a generalized distribution. Although this is primarily a two-stage branching model, it can be given a clustering interpretation. So can any other two-stage branching model.

1.5.4 Branching models

A branching process is a mathematical model for random "motion" of a family of "particles" each of which is in an isolated process of multiplication and death. Examples of such random motions are population growth, miosis of genes, growth of number of neutrons in an atomic chain reaction and cascade showers of cosmic rays.

Branching models give rise to a great number of interesting discrete distributions. Examples of these are provided by some special cases of the lost-games distribution and its shifted form. These examples have been discussed in detail by Kemp and Kemp (1969). The lost-games distribution (Kemp and Kemp (1968, 1969, 1971)) has the pgf

$$K_{m,m} \equiv G_m(z) = [G_1(z)]^m \qquad (1.5.4.1)$$

where

$$G_1(z) = \tfrac{1}{2}[1 - (1 - 4\theta(1 - \theta)z)^{1/2}]/(1 - \theta)$$

and follows from the shifted distribution with the pgf

$$K_{m,n}(z) = z^n p^m {}_2F_1(\tfrac{1}{2}m, \tfrac{1}{2}(m + 1), m + 1; 4\theta(1 - \theta)z).$$

Kemp and Kemp (1969) shows that $K_{2,1}(z)$ is equivalent to an example of a multiplicative process, and $K_{1,0}(z)$ and $K_{2,0}(z)$ to special cases of Neyman and Scott's epidemiological model. Also, $K_{a,0}(z)$ is shown to be the pgf of the distribution of the total number of descendants of the initial individual in a three-generation branching process, on the assumption of a binomial first-generation distribution and $K_{2,0}(z)$ second-generation distribution. If, alternatively, a negative binomial first-generation and a $K_{1,0}(z)$ second-generation distribution are assumed, the total descendants pgf is again $K_{a,0}(z)$.

Feller (1968, pp.295-296) provides another example. Suppose $G_n(z)$ denotes the pgf of the size of the nth generation. Then $G_n(z) = G_1(G_{n-1}(z))$. If the number of direct descendants is subject to the geometric distribution with $G(z) \equiv G_1(z) = (1 - \theta)/(1 - \theta z)$, $\theta \neq \frac{1}{2}$, then

$$P_n(z) = q[(p^n - q^n) - (p^{n-1} - q^{n-1})pz]/[(p^{n+1} - q^{n+1}) - (p^n - q^n)pz] \qquad (1.5.4.2)$$

where $p \equiv \theta$ and $q = 1 - \theta$.

1.5.5 Birth and birth-death models

Pure birth, birth-and-death, and birth-death-immigration models provide a basic means for studying the dynamics of a population. It should be noted that the terms "births", "deaths", etc., are general terms connoting any form of inputs, outputs, etc. e.g. in a queueing system. The distributions of population size and of the increase in the population size whether at any arbitrary time or at an equilibrium state are usually lattice distributions.

Hall (1956) obtains the four-parameter system of general Gaussian hypergeometric probability distributions as the distributions (at equilibrium) of the population size in some birth and death processes. Restricted conditions produce the general confluent hypergeometric probability distributions, the negative binomial and the Poisson distributions, respectively.

Boswell and Patil (1970) describe the genesis of the negative binomial from four growth processes. The distribution is shown to be the population size increase distribution in a Yule-Furry (pure birth) process, the population size distribution or the population size increase distribution in a growth-with-immigration model, a linear birth process distribution, or a linear birth-death process distribution. The same authors have developed the log series-with-added-zeroes from a queueing or non-linear birth-death process. Some other discrete distributions are similarly discussed in Boswell and Patil (1972).

1.5.6 Waiting-time distributions

Waiting-time distributions arise in occupancy models and from queues and other models of recurrent events. By the waiting time for the event A, we mean the number of trials up to and including the first occurrence of A. This is a random variable and its distribution is called the waiting-time distribution. For discrete events, the associated waiting-time will usually have a discrete distribution.

Waiting-time problems arise in occupancy models when we wish to consider the number of balls needed to satisfy specified occupancy conditions for the urns e.g. the number of assignments, one ball at a time, needed for all urns to be occupied, or at least r urns to have at least m balls each or alternatively, k specified urns to have at least m balls each, and so on. Sometimes waiting-time calculations are specified by using the (conditional) distributions of numbers of balls in the urns, given n, the number of balls thrown.

The negative binomial and the negative hypergeometric distributions are obtained from waiting-time problems in situations where a fixed number of observations would produce binomial and hypergeometric distributions respectively. Similarly, inverse Polya distributions arise in situations where fixed numbers of observations follow the Polya-type distributions.

The distribution of the duration of the game in the classical ruin problem and of the busy period in a queueing system are examples of waiting-time distributions. Detailed discussions on these are given by Feller (1968).

1.5.7 Weighted distributions and damaged models

When an investigation collects a sample of observations produced by nature according to a certain model, the original distribution may not be reproduced due to various reasons. The alteration on the original distribution may be due to non-ascertainability of certain observations or to partial destruction. Another way in which the original distribution is altered is due to giving to different observations unequal chances of being included in the sample. This feature is inherent in some convenient and natural methods of sampling. A great number of distributions can be constructed from these models representing various degrees and types of imperfection in sampling and recording data. Rao (1965) has given a good discussion of the models.

If sampling is not random but the probability of inclusion in the sample depends on the value of the variable to be measured, the result is a weighted distribution. From an original distribution with probabilities π_x, and a weighting function w(x) (the probability of inclusion in the sample given that X = x), we obtain the probability function of the weighted distribution as

$$P_x = w_x \pi_x / \Sigma w_i \pi_i . \qquad (1.5.7.1)$$

Similar definitions are available for continuous distributions. Since any lattice distributions can be derived in this way from any other with the same support, interest resides in particular simple cases.

In the simple case when a distribution D with probabilities P_x is modified by increasing P_r and compensating for this by multiplying the remaining P_x's by an appropriate constant the resultant distribution D*, with probabilities P_x^*'s, say, is a modified D-distribution. The P_x^*'s are given by the relation

$$P_r^* = (1 - \alpha) + \alpha P_r$$

$$P_x^* = \alpha P_x , \qquad x \neq r \qquad (1.5.7.2)$$

The pgf of D* is

$$G^*(z) = (1 - \alpha)z^r + \alpha G(z). \qquad (1.5.7.3)$$

If $0 < \alpha < 1$, D* is called an inflated D-distribution and if $1 < \alpha \leq (1 - P_r)^{-1}$, it is called a deflated D-distribution. When $\alpha = (1 - P_r)^{-1}$, $P_r^* = 0$.

The cases of modified distributions most commonly considered are those for which r = 0. If r = 0 and $P_r^* = 0$, then we have a zero-truncated D-distribution, a decapitated distribution as it is sometimes also called. If r = 0 and $0 < \alpha < 1$, D* is sometimes called a D-distribution with-added-zeroes.

A detailed review of weighted distributions has been given by Patil and Rao (1977, 1978) who also present several practical examples.

Kemp (1973) has pointed out that a type of weighting function of the form $w(x) \propto \alpha^x$ if used to modify any ghpd would yield a modified distribution which is also ghpd of the same form. This feature would make it impossible to detect unascertainment in any situations in which this model applies.

1.5.8 STER Models

The use of STER models arises in connection with an inventory decision problem. Imposing a decision function on a random variable called fraction of leadtime out of stock, in a general inventory problem, amounts to selecting that value of x for which the sum of the values of

$$u(x) = \sum_{y=x+1} (p_y^*/y), \qquad x = 0, 1,... \qquad (1.5.8.1)$$

first exceeds the desired fraction of leadtime not out of stock, where y is the actual demand during the leadtime occurring with probability p_y^*. Bissinger (1965) observes that

$$\sum_{x=0}^{\infty} u(x) = 1 - p_0^* \qquad (1.5.8.2)$$

and hence that

$$P_x = \frac{1}{1 - p_0^*} \sum_{y=x+1} (p_y^*/y), \quad x = 0, 1,... \quad (1.5.8.3)$$

is a valid pf for a discrete distribution, the probabilities of this distribution being Sums successively Truncated from the Expectation of the Reciprocal of the original random variable Y, the demand variable. Hence the distribution of X is called the STER distribution corresponding to the given random variable Y. It is usual to call Y the demand variable even when the discussion of STERing is not in the context of inventory decision.

STERing is a modification which can be carried out on any lattice distribution supported on x = 0, 1,..., just as weighting, shifting, or truncation, although even the simplest discrete distributions seem to lead to complicated STER distributions.

Suppose G(z) and H(z) denote the pgfs of the demand and the STER distributions respectively. Patil and Joshi (1968) observe that

$$H(z) = \frac{1}{(1 - z)(1 - p_0^*)} \int_z^1 \frac{G(z) - p_0^*}{z} dz \qquad (1.5.8.4)$$

where p_0^* is the probability of zero demand. Note that zero-modification of the demand variable does not affect the STER distribution. Kemp and Kemp (1969) show that if

$$G(z) = z\,{}_pF_q((a\,);\,(b\,);\,\lambda(z - 1)) \qquad (1.5.8.5)$$

then

$$H(z) = {}_{p+1}F_{q+1}(1,\,(a\,);\,2,\,(b\,);\,\lambda(z - 1)) \qquad (1.5.8.6)$$

(see also Kemp and Kemp (1975)).

1.6. Computer Generation of (Univariate) Discrete Variables

1.6.1 Introduction

Distribution sampling is often required when investigating the small-sample properties of various estimation procedures. It is also sometimes required in pilot studies preceding complex theoretical or empirical investigations.

In considering the features of a particular method for generating data from a given distribution it is important to distinguish between two modes of usage: that in which the parameters of the distribution are fixed and that in which they change at each call of the generator so that any constants have to be recalculated. While some methods, such as Fishman's (1976) method for generating a Poisson variable, can operate satisfactorily in the two modes with relatively little adjustment, many are grossly inefficient when variable parameters are involved.

Another important consideration in the selection of a data-generating method is the computing device to be employed. On some

devices, such as micro-computers, storage is a major constraint, and hence a slow but compact algorithm may be preferred to a fast one requiring appreciable storage. On the mainframe, speed rather than storage is usually the determining factor. Walker's (1974) alias method for generating variates from any specified distribution is a very fast method although it uses a relatively large amount of storage.

Here we briefly describe the major types of methods for obtaining discrete random variates on the computer.

1.6.2 The inverse transformation method

Suppose we can, conveniently, generate a random variable Y with distribution function $F_Y(y)$ and are interested in generating a random variable X with distribution function $F_x(x)$. Then we say the random variable X is generated by a transformation $x = h(y)$ if $h(Y)$ has the desired distribution $F_x(x)$.

Rigorously considered, the principle applies only to continuous distributions. However, a development based on the same idea is employed for discrete distributions. Let P_i, $i = 1, 2, \ldots, m$ be any discrete distribution (possibly truncated to m outcomes) on the sample space x_i, $i = 1, 2, \ldots, m$ where $P_i = Pr(X = x_i)$. If u is a uniform (0, 1) variate then X has the desired distribution where $X = x_i$ if

$$\sum_{j=0}^{i-1} P_j < u \leq \sum_{j=0}^{i} P_j, \qquad i = 1, \ldots, m.$$

Note that in this case, the inverse of F(x) is not well-defined, but the idea is the same as for the continuous, and the procedure is likewise called the inverse transformation method. The implementation of the method, briefly, is as follows.

The probabilities P_x are calculated and successively accumulated to obtain the distribution function $F(x) = \Sigma P_j$ summing over j from 0 to x for all x up to m where m is either the upper limit of the range

of the support or for a distribution with infinite support, an arbitrary truncation point such that 1 - F(m) is negligible. Using a standard pseudo-random number generator a u_i is obtained which can be regarded as a random value from the uniform distribution (0, 1). The number u_i is compared with successive F(x), x = 0, 1,... until the first F(x) say F(r) is reached which equals or exceeds u_i. Then $x_i = r$ is the required value. This implementation technique is called the sequential search algorithm.

Chen and Asau (1974) have suggested an important modification to the sequential search technique. This involves splitting the cumulative probability range 0 - 1 into k (say) equal intervals and noting the interval in which the cumulative probabilities F(x) lie. When a random uniform value u is presented for interpolation, only the interval in which it lies is searched. This technique is called the indexed-search algorithm. Kemp and Loukas (1978a, 1978b) have pointed out that ordering the probabilities before accumulating them represents yet a further improvement in speed.

Fishman's (1976) algorithm for sampling from the Poisson distribution is a variation of the inverse transformation method.

1.6.3 The acceptance-rejection method

The acceptance-rejection method, like the inverse transformation method, was primarily introduced for continuous distributions (von Neumann (1951)). However it has been successfully implemented for generating discrete variables.

Suppose it is desired to generate a random variable X with distribution $F_x(x)$ (and probability function $f_x(x)$) called the "target" distribution. It is assumed that a satisfactory method (preferably an inverse transformation method) exists for generating another random variable Y (discrete or continuous) with distribution $F_y(y)$ (and p.f.

$f_Y(y)$) called the "envelope" distribution. This method consists of two parts, generating an observation of Y and then deciding whether or not to set X equal to this sample value or to reject the value and generate a new value of Y.

To generate X from observations of Y by the acceptance-rejection method the ratio $L(x) \equiv f_X(x)/f_Y(x)$ must be bounded. This implies that the support of $f_X(x)$ must be a subset of the support of $f_Y(x)$.

Let θ_y ($\simeq \theta_x$) be the probability of accepting Y (i.e. setting X = y) when Y = y, and suppose A is the event that sampling stops on a given trial. It can be shown that

$$\theta_x = P_r(A) \cdot f_X(x)/f_Y(x) \qquad (< 1) \qquad (1.6.3.1)$$

i.e. that the acceptance probability is proportional to the likelihood ratio of observations. An optimal design is to choose $P_r(A)$ as large as possible so as to have large θ_x for all x.

Patil et al. (1975*, add. ref.) have pointed out that the method can be regarded as a method of "negative mixture". They show that with $\alpha \equiv \sup\{f_X(x)/f_Y(x)\}$ (> 1), the function $f_Z(x)$ given by

$$f_Z(x) = [\alpha f_Y(x) - f_X(x)]/(\alpha - 1) \qquad (1.6.3.2)$$

is a probability function, and that

$$f_Y(x) = \theta_x f_X(x) + (1 - \theta_x) f_Z(x) \qquad (1.6.3.3)$$

where $\theta_x = \frac{1}{\alpha}$. Hence $f_Y(x)$ is a mixture of $f_X(x)$ the desired distribution and $f_Z(x)$ another distribution. In this sense $f_X(x)$ can be regarded as a "negative mixture" or an "inverse mixture" of $f_Y(x)$.

Atkinson (1979) and Kemp et al. (1979) have used the method to generate data from the Poisson and the binomial (which are discrete target distributions) using as envelope the logistic, a continuous distribution. Fishman (1979) has generated a discrete target, the binomial, from a discrete envelope, the Poisson. Kemp (1980) has discussed some discrete distributions having simple analytically invertible cumulative

functions, and which can therefore be used as envelopes for generation of discrete target distributions, as illustrated by the generation of a logarithmic target using certain of these distributions (the geometric and some special cases of the generalized Waring) as envelopes.

1.6.4 The composition method

Another method of generating stochastic variates on a computer is the composition method, otherwise known as the method of mixtures (e.g. Patil et al. (1975)). By a mixture here is meant a countable (or finite) mixture $F_x(x) = \Sigma p_i F_i(x)$ of a family F_i, $i = 1, 2, \ldots$ of distribution functions where $p_i > 0$, $i = 1, 2, \ldots$ and $\Sigma p_i = 1$.

A random variable X with distribution F(x) can then be generated by choosing one of the $F_i(x)$ with probability p_i, generating X_i from the chosen distribution and setting $X = x_i$.

In some simulation problems the underlying model may be a contaminated population. Suppose X_1 has distribution $F_1(x)$ and X_2 has distribution $F_2(x)$ and we know that the desired random variable X comes from $F_1(x)$ a fraction θ of the time and from $F_2(x)$ a fraction $(1 - \theta)$ of the time. Then

$$F_x(x) = \theta F_1(x) + (1 - \theta) F_2(x) \qquad (1.6.4.1)$$

and we can generate X by using this mixture.

Walker's "alias" procedure (Walker (1974), Kronmal and Peterson (1979)) for generating random variables from any discrete distribution is a special case of the composition method.

1.6.5 Specialized methods based on stochastic models

In some cases a particular stochastic model giving rise to a specified distribution may indicate ways of generating variates from the distribution by synthesising variates from the component distributions.

One of the simplest and most popular algorithms for sampling from the Poisson distribution (with pgf $G(z) = e^{\lambda(z-1)}$) is based on the exponential gap property of the Poisson process. Naylor et al. (1966, pp.112-114) give a good description of this procedure and its stochastic derivation.

A method of generating a Hermite variate is based on viewing the Hermite distribution as the sum of two correlated Poisson variables. A bivariate Poisson variable (X_1, X_2) is constructed from three independent Poisson variables Y_1, Y_2, Y_3 where $X_1 = Y_1 + Y_3$ and $X_2 = Y_2 + Y_3$. The sum $X_1 + X_2$ (i.e. $Y_1 + Y_2 + 2Y_3$) then gives a Hermite variable.

Another method of sampling from the Hermite distribution is to consider the variable X as the sum of an ordinary Poisson variable X_1 and a Poisson "doublet" variable X_2. This is evident from the pgf

$$\begin{aligned} G_X(z) &= \exp\{\alpha_1(z-1) + \alpha_2(z^2-1)\} \\ &= e^{\alpha_1(z-1)} \cdot e^{\alpha_2(z^2-1)} \\ &= G_{X_1}(z) \cdot G_{X_2}(z) \end{aligned} \qquad (1.6.5.1)$$

which is equivalent to the pgf of $X = X_1 + X_2$. Now, if $Y \sim \text{Poisson}(\lambda)$, then $2Y \sim$ Poisson "doublet"(λ). Hence if we generate two independent Poisson variables $Y_i \sim \text{Poisson}(\alpha_i)$, $i = 1, 2$, then $X = Y_1 + 2Y_2 \sim \text{Hermite}(x; \alpha_1, \alpha_2)$.

Kemp's (1981) methods of generating logarithmic variables are based on special features of its probability function. So is Kemp's (1982) variable-parameter method for generating data from a Poisson distribution.

1.7 Bivariate Discrete Distributions

1.7.1 Introduction

Generally, a meaningful bivariate distribution is produced by extending some special features of the univariate probabilistic models. In some cases, the bivariate distributions arise naturally in statistical sampling problems. Apparently for historical reasons, the commonest application of bivariate discrete distributions seems to be in the area of model discrimination. Some phenomena are such that two equally plausible models cannot be distinguished on the basis of observations over a single time-period or from quadrats of a single-size, etc., and we must take observations over two or more time-intervals, or quadrat-sizes, as the case may be (Arbous and Kerrich (1951)).

Any reference to bivariate distributions here and subsequently is intended for bivariate distributions having marginals of the same form e.g. (Poisson, Poisson), (Negative binomial, Negative binomial) which will be called bivariate Poisson, bivariate negative binomial, respectively. This is in contrast with the wider sense employed by Leiter and Hamdan (1973) and Cacoullos and Papageorgiou (1980) whose discussions include (Poisson, Binomial), (Poisson, Negative binomial), etc. Bivariate distributions with one marginal discrete and the other continuous such as Park's (1970) are excluded. Furthermore, since the marginals of the bivariate distribution discussed by Patil and Bildikar (1967) are not the logarithmic but rather the modified logarithmic distribution, we will call it the bivariate modified logarithmic distribution.

1.7.2 The structure of bivariate distributions

The most basic way of deriving a bivariate distribution is to make the substitution $z = z_1^k z_2^m$ in the pgf $u(z)$ of a univariate distribution, thus obtaining

$$h(z_1, z_2) = u(z_1^k z_2^m). \qquad (1.7.2.1)$$

This, though trivial on its own, has been used by Ord (1975) to define a general bivariate Poisson distribution, A, having the pgf

$$G_A(z_1, z_2) = \prod_{i=1}^{r} \exp[\lambda_i (z_1^{k_i} z_2^{m_i} - 1)]$$

$$= \exp\left[\sum_{i=1}^{r} \lambda_i \left(z_1^{k_i} z_2^{m_i} - 1\right)\right] \qquad (1.7.2.2)$$

(with non-negative integral k_i, m_i and $r = 1, 2, \ldots$), which is a convolution of r bivariate Poisson distributions of the form

$$h_A(z_1, z_2) = \exp(\lambda z_1^k z_2^m - 1) \qquad (1.7.2.3)$$

Some caution is needed on the name of A since its marginals are not Poisson in general. It is probably better to call A a bivariate Poisson-type distribution. It has been noted by Ord (1975) that the various types of bivariate Hermite are special cases of A.

Proceeding in the same line, a bivariate binomial-type distribution, B, is given by

$$G_B(z_1, z_2) = \prod_{i=1}^{r} \left(q_i + p_i z_1^{k_i} z_2^{m_i}\right)^{n_i} \qquad (1.7.2.4)$$

Its marginals are members of Teicher's (1961) general binomial family having the pgf

$$u(z) = \prod_{i=1}^{s} (q_i + p_i z^{m_i}). \qquad (1.7.2.5)$$

A bivariate negative binomial-type distribution, C, can be similarly defined by the pgf

$$G_C(z_1, z_2) = \prod_{i=1}^{r} \left(1 + \beta_i - \beta_i z_1^{k_i} z_2^{m_i}\right)^{-\alpha_i} \qquad (1.7.2.6)$$

For $r = 1$, it is obvious that bivariate negative binomial distributions of Arbous and Kerrich (1951), Edwards and Gurland (1961) and Khatri (1971, pp.213-215) are special cases of C.

It is well known that the two marginals of a given bivariate distribution are uniquely determined but, theoretically, an infinite

number of bivariate distributions can be derived with any two given univariate distributions as their marginals. However only a few of the bivariate distributions obtainable from two given univariates would be of practical value.

If a bivariate D-distribution with D-distribution as marginals has a simple functional form compared with other bivariates with D-distribution marginals but is general enough to have most of the important properties, it may be referred to as the bivariate D-distribution. In any event, it seems that a particular bivariate would qualify as the standard in relation to others of its class only on the grounds of its pre-eminence in applications.

An elementary result in probability theory states that if a bivariate distribution $F(x_1, x_2)$ is derived from two univariates known to be independent, then $\text{Correlation}(X_1, X_2) = 0$, but the converse does not necessarily hold. The marginals of a bivariate having the property $\text{Corr}(X_1, X_2) = 0$ may in fact be dependent. Bivariate discrete distributions for which zero correlation implies independence are very important regarding applications.

Ahmed (1971) has defined such a class of bivariate discrete distributions. The class is characterized by the joint probability function

$$f(x_1, x_2) = f_A(x_1)\, f_B(x_2)[1 + \theta H_A(x_1)\, H_B(x_2)], \quad |\theta| \leq 1 \qquad (1.7.2.7)$$

with

$$H_A(x_1) = F_A(x_1) + F_A(x_1 - 1) - 1$$

and

$$H_B(x_2) = F_B(x_2) + F_B(x_2 - 1) - 1$$

where $F_A(x_1)$, $F_B(x_2)$ are the cumulative probability functions of X_1 and X_2 respectively. Since there are only a few discrete distributions (e.g. Kemp (1980)) having simple form of $F(x)$, the use of this approach

for deriving bivariate discrete distributions is limited. However Ahmed (1971) has given some properties of this class which provide useful information without much difficulty. In particular upper bounds on correlations are obtained for members of the class having as marginals the Poisson, binomial, geometric, and discrete uniform distributions respectively.

A bivariate D-distribution whose marginals are known to be independent is usually called a double D-distribution. This follows from the corresponding multivariate which is called a multiple D-distribution. A multivariate discrete distribution is said to be homogeneous if its pgf contains the generating variables only in the form of a linear function within a function of a function. In particular, a homogeneous bivariate discrete distribution is a distribution with pgf of the form

$$G(z_1, z_2) = H(\alpha z_1 + \beta z_2) \qquad (1.7.2.8)$$

An important subset is given by the joint distribution of X_1, X_2 when $X_1 | X_2$ is distributed as Binomial$(x_1; \theta, x_2)$. This has a pgf of the form

$$G(z_1, z_2) = H(\theta z_1 + (1 - \theta) z_2). \qquad (1.7.2.9)$$

The bivariate discrete distributions specifically discussed in the literature include:

the bivariate Poisson with the pgf

$$G(z_1, z_2) = \exp[\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1) + \alpha_3(z_1 z_2 - 1)]$$

(e.g. Paleheimo (1972));

the bivariate binomial having the pgf

$$G(z_1, z_2) = [1 + \theta_1(z_1 - 1) + \theta_2(z_2 - 1) + \theta_3(z_1 z_2 - 1)]^n \qquad (1.7.2.10)$$

(e.g. Kawamura (1973));

the bivariate negative binomial with pgf

$$G(z_1, z_2) = [1 - \beta_1(z_1 - 1) - \beta_2(z_2 - 1) - \beta_3(z_1 z_2 - 1)]^{-\alpha} \qquad (1.7.2.11)$$

(e.g. Edwards and Gurland (1961)), and its special case given by

$$G(z_1, z_2) = [1 - \beta_1(z_1 - 1) - \beta_2(z_2 - 1)]^{-\alpha} \qquad (1.7.2.12)$$

(e.g. Khatri (1971, pp.213-215));

the bivariate modified logarithmic with

$$G(z_1, z_2) = [\log(1 - \theta_1 z_1 - \theta_2 z_2 - \theta_3 z_1 z_2)]/\log(1 - \theta_1 - \theta_2 - \theta_3)$$

$$0 < \Sigma\theta_i < 1,\ 0 < \theta_i < 1,\ i = 1, 2, 3 \qquad (1.7.2.13)$$

(e.g. Patil and Joshi (1968)), and its special case given by

$$G(z_1, z_2) = [\log(1 - \theta_1 z_1 - \theta_2 z_2)]/\log(1 - \theta_1 - \theta_2) \qquad (1.7.2.14)$$

(e.g. Kemp (1981));

the bivariate hypergeometric having the pgf

$$G(z_1, z_2) = [M_{00}^{(n)}/N^{(n)}]\,{}_4F_1(-n, -M_{10} - M_{11}; M_{00} - n + 1; z_1, z_2, z_1 z_2) \qquad (1.7.2.15)$$

(e.g. Steyn (1957));

the bivariate Neyman Type A-I, -II, -III with pgf given respectively by

$$G(z_1, z_2) = \exp\{\lambda[\exp\{\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1) + \alpha_{12}(z_1 z_2 - 1)\}]\}$$

$$G(z_1, z_2) = \exp\{\lambda_1[\exp\{\alpha_1(z_1 - 1)\} - 1] + \lambda_2[\exp\alpha_2(z_2 - 1) - 1] + \lambda_{12}[\exp\{\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1)\} - 1]\}$$

$$G(z_1, z_2) = \exp\{\lambda_1[\exp\{\alpha_1(z_1 - 1)\} - 1] + \lambda_2[\exp\{\alpha_2(z_2 - 1) - 1] + \lambda_{12}[\exp\{\alpha_{12}(z_1 z_2 - 1)\} - 1]\}$$

(e.g. Holgate (1966)); and

the bivariate Hermite with

$$G(z_1, z_2) = \exp[\alpha_1(z_1 - 1) + \alpha_2(z_1^2 - 1) + \alpha_3(z_2 - 1) + \alpha_4(z_2^2 - 1) + \alpha_5(z_1 z_2 - 1)]$$

(e.g. Kemp and Papageorgiou (1982)).

1.7.3 Marginality and conditionality

For a bivariate discrete distribution of (X_1, X_2) having the pgf $G(z_1, z_2)$, the marginals are given by

$$G_{X_1}(z_1) = [G(z_1, z_2)]_{z_2=1} \qquad (1.7.3.1)$$

$$G_{X_2}(z_2) = [G(z_1, z_2)]_{z_1=1} \qquad (1.7.3.2)$$

$$G_{X_1+X_2}(z) = [G(z_1, z_2)]_{z_1=z_2=z} \qquad (1.7.3.3)$$

and

$$G_{X_1-X_2}(z_1) = [G(z_1, z_2)]_{z_2=\frac{1}{z_1}} \qquad (1.7.3.4)$$

Subrahmaniam (1966) has shown that the conditional distribution of X_1 given $X_2 = n$ is obtainable from

$$G_{X_1|X_2=n}(z_1) = \left\{\frac{\partial^n}{\partial z_2^n} G(z_1, z_2) \Big/ \left[\frac{\partial^n}{\partial z_2^n} G(z_1, z_2)\right]_{z_1=1}\right\}_{z_2=0} . \qquad (1.7.3.5)$$

Kemp (1981) points out that for a homogeneous bivariate $H(\alpha z_1 + \beta z_2)$, this reduces to

$$G_{X_1|X_2=n}(z_1) = H^{(n)}(\alpha z_1)/H^{(n)}(\alpha) \qquad (1.7.3.6)$$

which, interestingly, is independent of β.

It is easy to show that

$$G_{X_1|X_1+X_2=n}(z_1) = \left\{\frac{\partial^n}{\partial z_2^n} G(z_1 z_2, z_2) \Big/ \left[\frac{\partial^n}{\partial z_2^n} G(z_1 z_2, z_2)\right]_{z_1=1}\right\}_{z_2=0} \qquad (1.7.3.7)$$

and that

$$G_{X_1|X_1-X_2=n}(z_1) = \left\{\frac{\partial^n}{\partial z_2^n} G\left(z_1 z_2, \frac{1}{z_2}\right) \Big/ \left[\frac{\partial^n}{\partial z_2^n} G\left(z_1 z_2, \frac{1}{z_2}\right)\right]_{z_1=1}\right\}_{z_2=0} \qquad (1.7.3.8)$$

and

$$G_{X_1|X_2-X_1=n}(z_1) = \left\{\frac{\partial^n}{\partial z_2^n} G\left(\frac{z_1}{z_2}, z_2\right) \Big/ \left[\frac{\partial^n}{\partial z_2^n} G\left(\frac{z_1}{z_2}, z_2\right)\right]_{z_1=0}\right\}_{z_2=0} . \qquad (1.7.3.9)$$

From the well-known expressions

$$E(X_1) = \left[\frac{d}{dz_1} G_{X_1}(z_1)\right]_{z_1=1} \qquad (1.7.3.10)$$

$$E(X_2) = \left[\frac{d}{dz_2} G_{X_2}(z_2)\right]_{z_2=1}, \text{ and} \qquad (1.7.3.11)$$

$$E(X_1 X_2) = \left[\frac{\partial^2}{\partial z_1 \partial z_2} G(z_1, z_2)\right]_{z_1=z_2=1}, \qquad (1.7.3.12)$$

the covariance, $\mathrm{Cov}(X_1, X_2)$, is easily obtained by evaluating

$$\mathrm{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1) \cdot E(X_2) \qquad (1.7.3.13)$$

It is, of course, clear that

$$\mathrm{Var}(X_1) = \left[\frac{\partial^2}{\partial z_1^2} \log G(e^{z_1}, e^{z_2})\right]_{z_1 = 0,\ z_2 = 0} \qquad (1.7.3.14)$$

and

$$\mathrm{Var}(X_2) = \left[\frac{\partial^2}{\partial z_2^2} \log G(e^{z_1}, e^{z_2})\right]_{z_1 = 0,\ z_2 = 0} \qquad (1.7.3.15)$$

Correlation and similar properties of bivariate distributions can be easily studied by expansion of the bivariate frequency function in canonical form. The canonical form of the bivariate Poisson has been derived by Campbell (1934). Those for bivariate binomial and bivariate hypergeometric have been given by Aitken and Gonin (1935). Hamdan and Albayyati (1971) have derived the canonical form of the bivariate negative binomial. Lancaster (1958) gives the general theory of canonical expansion of bivariate distributions and Eagleson (1964) applies this to bivariate distributions derived by the additive property of the corresponding univariate distribution e.g. Poisson, binomial, and negative binomial.

The regression of X_1 on X_2 is another property commonly of interest. When the regression is linear, analysis is comparatively easy. Kenney (1939) has pointed out that linearity of regression can be characterized by the differential equation

$$\left[\frac{\partial}{\partial z_1} G(z_1, z_2)\right]_{z_1 = 0} = c\, \frac{d}{dz_2} G(0, z_2). \qquad (1.7.3.16)$$

It has also been shown by Eagleson and Lancaster (1967) that the regression of X_1 on X_2 is linear if and only if X_1 and X_2 have additive random elements in common i.e. iff X_1 and X_2 are such that $X_1 = Y_1 + Y_3$ and $X_2 = Y_2 + Y_3$ where Y_1, Y_2, Y_3 are mutually independent and belong to the same additive family.

1.7.4 Chance mechanisms generating bivariate discrete distributions

As in the univariate case, most bivariate discrete distributions arise from either the mixing or the generalising (clustering) models. Khatri (1971) has given a good discussion of these models for multivariate discrete distributions of which the bivariates are a special case.

Hamdan and Tsokos (1971) derive the bivariate Poisson by mixing on n the bivariate binomial with the pgf

$$h(z_1, z_2) = [1 + \theta_1(z_1 - 1) + \theta_2(z_2 - 1) + \theta_3(z_1 z_2 - 1)]^n .$$

where $0 < \theta_i < 1$ and $\Sigma\theta_i \leqslant 1$. Arbous and Kerrich (1951) and Edwards and Gurland (1961) obtain their respective forms of the bivariate negative binomial distribution by mixing on λ the bivariate Poisson

$$h_1(z_1, z_2) = \exp\{\lambda[\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1)]\} \qquad (1.7.4.1)$$

and

$$h_2(z_1, z_2) = \exp\{\lambda[\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1) + \alpha_3(z_1 z_2 - 1)]\} \qquad (1.7.4.2)$$

respectively.

Holgate (1969) derives three types of bivariate Neyman Type A from different generalising (or clustering) models. Kemp and Papageorgiou (1982) have discussed the genesis of the bivariate Hermite distribution in terms of both the mixing and the generalising models.

1.7.5 Computer generation of bivariate discrete variables

The most basic way of generating data from a bivariate discrete distribution is to obtain $k = x_1 + (m + 1)x_2$, the values of k being the locations when the $\{P(x_1, x_2)\}$ array is stored column-by-column in one-dimensional array, with locations $0, 1, \ldots, (m + 1)(n + 1) - 1$. The $\{P(k)\}$ is now treated as a univariate distribution from which random observations are generated using a general univariate method such as Walker's alias method (Walker (1974), Kronmal and Peterson (1979)) or

the modified indexed-search method (Kemp and Loukas (1978a, 1978b)). Other general techniques for the bivariate are discussed in Kemp and Loukas (1981).

Sometimes, as in the univariate case, the structure of a bivariate discrete distribution may suggest a particular method based on the stochastic model giving rise to the distribution. Kemp and Loukas (1978a) discuss such methods for generating the bivariate Hermite variable.

Note that the fixed-parameter and the variable-parameter modes of usage also apply to the methods for generating bivariate data.

1.8 Summary on "Discrete Distributions"

As a general background to the discussion of generalized Poisson distributions, the whole subject of discrete distributions has been briefly considered in this chapter.

The major families of discrete distributions have been described. These are convenient ways of classifying discrete distributions although they overlap considerably. Each family or class is defined essentially along certain mathematical structures. Of all these classes, generalized hypergeometric distributions, compound distributions, and generalized distributions will feature most prominently in the remaining part of the thesis.

The various major chance mechanisms generating discrete distributions have also been discussed. Mixing models and clustering models will be discussed further in Chapter 2 in connection with generalized Poisson distributions.

Methods of generating discrete random variables on the computer have been considered very briefly. Distribution sampling is often important in the study of the small-sample behaviour of estimation methods. It is also frequently employed in pilot studies preceding complex investigations.

The structure and genesis of bivariate discrete distributions has been discussed. Generally, a meaningful bivariate distribution is produced by extending some special features of the univariate probabilistic models. Computer generation of bivariate discrete variates is also considered.

CHAPTER 2

GENERALIZED POISSON DISTRIBUTIONS

CHAPTER 2

GENERALIZED POISSON DISTRIBUTIONS

2.1 Introduction

A lattice distribution on $x = 0, 1, \ldots$ is said to be a generalized Poisson distribution (GPD) if and only if its pgf can be put in the form

$$G(z) = \exp\{\lambda[h(z) - 1]\} \qquad (2.1.0.1)$$

where $h(z)$ is a pgf. Given λ and $h(z)$, then $G(z)$ is unique. We shall sometimes denote this by $G(z) \equiv GPD(\lambda, h(z))$. Note however that

$$\begin{aligned} G(z) &= \exp\{\alpha[u(z) - 1]\} \\ &= \exp\left\{\left\{\frac{\alpha}{\theta}\right\}\left\{[(1 - \theta) + \theta u(z)] - 1\right\}\right\}, \quad 0 < \theta \leq 1 \\ &= \exp\left\{\beta\left\{[(1 - \theta) + \theta u(z)] - 1\right\}\right\} \end{aligned}$$

i.e.

$$\begin{aligned} G(z) &\equiv GPD\{\alpha, u(z)\} \\ &\equiv GPD\{\beta, (1 - \theta) + \theta u(z)\}, \qquad 0 < \theta \leq 1. \end{aligned} \qquad (2.1.0.2)$$

Hence, for a given $G(z)$, the pair $(\lambda, h(z))$ is not unique.

In this chapter, properties and applications of the generalized Poisson distributions are discussed. Results from Feller (1968), Maceda (1948), Warde and Katti (1971), Steutel (1970), and Harn (1978) are unified to give a clearer picture of the structure of this important class of discrete distributions.

The expression "classical and contagious discrete distributions" is common and is indeed the title of the 1963 international symposium at Montreal, Canada (Patil (editor, 1965)). However until now the distinction between classical and contagious is unclear. Further complication is introduced by the subdivision of the latter into the truly contagious and the apparently contagious groups. GPDs and all other generalized discrete distributions are referred to as the truly contagious distributions,

in contrast with mixtures or compound distributions which are regarded as exhibiting apparent contagion. However some discrete distributions e.g. negative binomial and Neyman Type A are known to belong simultaneously to both groups.

One may tend to inquire whether the terms "contagious" and "non-contagious" do not merely describe different parametric forms of distributions which are basically the same. In other words, is every lattice distribution necessarily a contagious distribution? And similarly, is every discrete mixture necessarily a generalized distribution, or vice versa? In particular, is every lattice distribution on x = 0, 1,... necessarily a GPD? And is every compound Poisson distribution necessarily a generalized Poisson distribution? These latter two are among the questions considered in this chapter.

Infinite divisibility characterizes the GPDs among lattice distributions on non-negative integers and hence this property dominates the discussion. Some systems of discrete distributions which are subsets of the GPD class are considered. Certain members of these systems are identified and the cluster-size distribution of some are determined. One of the systems discussed comprises of mixtures of the Poisson with inf div continuous mixing distributions supported on the positive half real line. One of such mixtures of the Poisson is that obtained with the reciprocal gamma distribution (Pearson's Type V). Its probability function is derived and some of its important properties are discussed.

Finally, the different stochastic models giving rise to GPDs are considered.

2.2 Properties of Generalized Poisson Distributions

2.2.1 Necessary conditions for a GPD

It is clear that a lattice distribution which is left-truncated or terminating cannot be a GPD. Hence a necessary condition for a discrete distribution to be a GPD is that it be a lattice distribution supported on $x = 0, 1, \ldots$

It is important to ask: is every lattice distribution on non-negative integers a GPD? In other words, is the above condition a sufficient condition? Suppose that a distribution on $x = 0, 1, \ldots$ has a pgf $G(z)$. Note that $G(z) = \exp(u(z))$ with $u(z) = \log G(z)$. Levy (1937, quoted by Lukacs (1970, pp.251-252)) has shown that it is possible for

$$G(z) = \exp[v(z) - v(1)] \qquad (2.2.1.1)$$

to be a pgf while $v(z)$ is a polynomial which is not a pgf. However $v(z)$ must be such that a term with negative coefficient is preceded by one term and followed by at least two terms with positive coefficients. An example is provided by

$$v(z) = \alpha z - \beta z^2 + \gamma z^3 + \delta z^4, \qquad \beta > \frac{\alpha^2}{2}. \qquad (2.2.1.2)$$

More recently, Kemp (1978) has shown that the shifted zero-truncated Poisson distribution having the pgf

$$G(z) = (e^{\lambda z} - 1)/[z(e^{\lambda} - 1)] \qquad (2.2.1.3)$$

is not a GPD. Clearly a lattice distribution on $x = 0, 1, \ldots$ is not necessarily a GPD.

Another necessary condition for a lattice distribution to be a GPD is

$$\mathrm{var}(X) \geq (P_1/P_0) \qquad (2.2.1.4)$$

This condition has been stated by Johnson and Kotz (1982) quoting from an unpublished work of I.R. James. It is easily shown as follows.

Suppose $G(z) = \Sigma P_x z^x$ is the pgf of the distribution of X and $h(z) = \Sigma p_y^* z^y$ is that of the distribution of Y where

$$G(z) = \exp\{\lambda[h(z) - 1]\}. \qquad (2.2.1.5)$$

Then

$$\lambda = P_1/(P_0 p_1^*) \qquad (2.2.1.6)$$

and

$$\begin{aligned} \text{var}(X) &= \lambda E_Y(Y^2) \\ &= \left(P_1/(P_0 p_1^*)\right) \Sigma y^2 p_y^* \\ &= (P_1/P_0)\left(1 + (1/p_1^*) \sum_{y=2} y^2 p_y^*\right) \\ &\geqslant (P_1/P_0) \end{aligned} \qquad (2.2.1.7)$$

Another necessary condition similar to James's concerns the property of overdispersion. For a lattice distribution to be a GPD, it is necessary that

$$\text{var}(X) \geqslant \text{mean}(X) \qquad (2.2.1.8)$$

This is shown in Section 2.3.2. It is, in fact, a more stringent condition than James's.

Steutel (1970, p.77) has given yet another necessary condition, but for only non-increasing lattice distributions. If a non-increasing lattice distribution has probabilities P_x, $x = 0, 1, \ldots$, then a necessary condition for it to be a GPD is

$$x(P_{x-1} - P_x) \leqslant P_0, \quad x = 1, 2, \ldots \qquad (2.2.1.9)$$

2.2.2 The GPD as a mixture of an n-fold convolution

In section 1.4.6 it was mentioned that any generalized D-distribution is a mixture, on n, of the n-fold convolution of the generalizer, with the D-distribution as the mixing distribution. It follows therefore that the generalized Poisson distribution $G(z) \equiv \text{GPD}(\lambda, h(z))$ is a mixture on n of the n-fold convolution of h(z) with a Poisson mixing distribution.

Let $h_x(z)$ denote the pgf of a lattice random variable X. Then the pgf of Y, the n-fold convolution of X is $[h_x(z)]^n$. Considering n as a parameter of the distribution of Y and mixing the distribution on this parameter with the Poisson (n; λ), we have a mixture with the pgf

$$\begin{aligned} G(z) &= \Sigma[h(z)]^n \frac{e^{-\lambda}\lambda^n}{n!} \\ &= e^{-\lambda}\Sigma \frac{[\lambda h(z)]^n}{n!} \\ &= e^{-\lambda}\exp[\lambda h(z)] \\ &= \exp\{\lambda[h(z) - 1]\}, \end{aligned} \tag{2.2.2.1}$$

which, by definition, is the pgf of the generalized Poisson distribution. Kemp (1970) notes that Lundberg (1909) is probably the first person to discuss the concept of generalized Poisson distributions in this sense.

2.2.3 The infinite divisibility of GPDs

A random variable X is said to be n-divisible if, for a given natural number (i.e. positive integer) n, there exists independent and identically distributed random variables $X_1, \ldots, X_n$ such that $X = X_1 + \ldots + X_n$ is distributed in the same form as the X_i's. A random variable X is said to be infinitely divisible if X is n-divisible for every positive integer n. In terms of characteristic function, a distribution with the characteristic function C(it) is infinitely divisible if and only if, for every $n \in N_+$, there exists a distribution with the characteristic function U(it) such that

$$C(it) = [U(it)]^n \tag{2.2.3.1}$$

or equivalently, if and only if, for every $n \in N_+$,

$$U(it) = [C(it)]^{\frac{1}{n}} \tag{2.2.3.2}$$

is a characteristic function. A lattice distribution having the pgf G(z) is infinitely divisible if and only if, for every $n \in N_+$,

$$h(z) = [G(z)]^{\frac{1}{n}} \qquad (2.2.3.3)$$

is a valid pgf.

That GPDs are infinitely divisible is obvious from the form of the pgf

$$G(z) = \exp\{\lambda[h(z) - 1]\}$$

$$= \{\exp[h(z) - 1]\}^{\lambda}. \qquad (2.2.3.4)$$

However, for GPDs, infinite divisibility is not just a property but a fundamental property, as underlined by the following theorem which has been stated and proved by Feller (1968, pp.288-290).

Theorem 2

A discrete distribution on x = 0, 1,... is inf div if and only if its pgf can be put in the form

$$G(z) = \exp\{\lambda[h(z) - 1]\} \qquad (2.2.3.5)$$

where h(z) is also a pgf; that is, if and only if it is a GPD.
This theorem means that infinite divisibility characterizes GPDs among lattice distributions.

Infinite divisibility however is a complicated concept. For example, while the convolution of two inf div pgfs is infinitely divisible, an inf div pgf may factorize into two pgfs only one of which is inf div (see e.g. Section 2.5.3). Also, mixing an inf div lattice distribution (e.g. the Poisson or the negative binomial) on even the exponent parameter may result in a distribution which is not infinitely divisible.

It is generally difficult to ascertain that a given discrete distribution is or is not inf div since most pgfs are not usually presented in a base-exponent form, i.e. in the form

$$G(z) = [h(z)]^{a}.$$

To reduce the difficulty, some sufficient conditions have recently been proposed by which some discrete distributions can be shown to be infinitely divisible. These are considered elsewhere, particularly in Sections 2.4 and 2.5 in this chapter.

2.2.4 The cluster-size distribution of a GPD

In the representation $G(z) \equiv GPD(\lambda, h(z))$, if $h(z)$ is supported on $x = 0, 1, \ldots$ then the pair $(\lambda, h(z))$ is not unique, as shown in Section 2.1, but the pair $(\lambda, h_+(z))$ in $G(z) \equiv GPD(\lambda, h_+(z))$ is unique where $h_+(z)$ is supported on $x = 1, 2, \ldots$ i.e. $h_+(0) = 0$. This latter form is called the Levy-Feller representation. Note that the former can always be recast into this representation, since

$$\begin{aligned} G(z) &= \exp\{\lambda[h(z) - 1]\} \\ &= \exp\left\{\gamma\left[\frac{h(z) - h(0)}{1 - h(0)} - 1\right]\right\} \\ &= \exp\{\gamma[h_+(z) - 1]\}, \qquad \gamma = \lambda[1 - h(0)] \end{aligned} \tag{2.2.4.1}$$

In particular if $h(z)$ is a generalized hypergeometric probability distribution (see Section 1.4.4), i.e.

$$h(z) = {}_pF_q\left[(a_i); (b_i); \lambda z\right] / {}_pF_q\left[(a_i); (b_i); \lambda\right] \tag{2.2.4.2}$$

then

$$\begin{aligned} h_+(z) &= \frac{{}_pF_q\left[(a_i); (b_j); \lambda z\right] - 1}{{}_pF_q\left[(a_i); (b_j); \lambda\right] - 1} \\ &= \frac{z\,{}_{p+1}F_{q+1}\left[1, (a_i + 1); 2, (b_j + 1); \lambda z\right]}{{}_{p+1}F_{q+1}\left[1, (a_i + 1); 2, (b_j + 1); \lambda\right]} \end{aligned} \tag{2.2.4.3}$$

For example,

Neyman Type A distribution

$\equiv$ Poisson(λ) V Poisson(α)

$\equiv G(z)$

$$\begin{aligned} &= \exp\left\{\lambda\left[e^{\alpha(z-1)} - 1\right]\right\} \\ &= \exp\left\{\lambda\left[\frac{{}_0F_0(-; -; \alpha z)}{{}_0F_0(-; -; \alpha)} - 1\right]\right\} \\ &= \exp\left\{\gamma\left[\frac{z\,{}_1F_1(1; 2; \lambda z)}{{}_1F_1(1; 2; \lambda)} - 1\right]\right\} \\ &= \exp\{\gamma[((e^{\alpha z} - 1)/(e^{\alpha} - 1)) - 1]\} \end{aligned}$$

$\equiv$ Poisson(γ) V Zero-truncated Poisson(α)

with $\gamma = \lambda(1 - e^{-\alpha})$ (2.2.4.4)

Boswell et al. (1979, p.109) remark that the use of the Neyman Type A distribution in studying clustering processes has been criticized on the grounds that it allows for zero cluster sizes. This criticism clearly stems from a lack of awareness that the distribution can be formulated alternatively as Poisson V Zero-truncated Poisson as shown in expression (2.2.4.4).

Another example which does not seem to be widely known is

Poisson(λ) V Geometric(θ), $0 < \theta < 1$

$$\equiv G(z)$$

$$= \exp\left\{\lambda\left[\left(\frac{1-\theta}{1-\theta z}\right) - 1\right]\right\}$$

$$= \exp\left\{\lambda\left[\frac{{}_1F_0(1;\,-;\,\theta z)}{{}_1F_0(1;\,-;\,\theta)} - 1\right]\right\}$$

$$= \exp\left\{\gamma\left[\frac{z\,{}_1F_0(1;\,-;\,\theta z)}{{}_1F_0(1;\,-;\,\theta)} - 1\right]\right\}$$

$$= \exp\left\{\gamma\left[z\left(\frac{1-\theta}{1-\theta z}\right) - 1\right]\right\}$$

$$\equiv \text{Polya-Aeppli}(\gamma, \theta),\ \gamma = \lambda\theta \qquad (2.2.4.5)$$

The Poisson V geometric and the Polya-Aeppli tend to be regarded as two distinct but similar distributions (e.g. Patil and Joshi (1968, p.49)), but they are in fact identical, as expression (2.2.4.5) shows. The geometric distribution has the interesting feature whereby zero-truncation gives the same result as right-shifting by one unit i.e. the zero-truncated-geometric distribution is identical with the shifted-geometric distribution.

The Levy-Feller representation shows that GPDs can arize from a Poisson distribution of clusters where cluster-size is a random variable on $x = 1, 2, \ldots$, having the pgf $h_+(z)$. Hence the distribution $h_+(z)$ is called the cluster-size distribution of the generalized Poisson distribution, $G(z) \equiv \text{GPD}(\lambda, h_+(z))$. For example, as is obvious from the above representation, the cluster-size distribution of the Neyman Type A is the zero-truncated Poisson with the pgf $h_+(z) = (e^{\alpha z} - 1)/(e^{\alpha} - 1)$.

While noting the definition given here, it is convenient to refer to any generalizing component, $h(z)$ or $h_+(z)$, of a GPD simply as its cluster-size distribution, as has been done by Boswell et al. (1979, p.110).

When a particular lattice distribution is shown to be inf div (and hence a GPD) it is often difficult to obtain the cluster-size distribution. Kemp (1978) has presented a method for obtaining the cluster-size probabilities. This method requires that the form of $H(z) = [G(z)]^{\alpha}$ be known explicitly or equivalently, that the probability function of $H(z)$ be available explicitly, where here $G(z)$ is the pgf of the GPD in question and α is an arbitrary positive real constant. The theorem on which this method is based also yields a necessary and sufficient condition for the infinite divisibility of a lattice distribution.

2.2.5 Cumulants and factorial cumulants

A useful relation exists between the cumulants of a GPD and the moments of its generalizing component, the cluster-size distribution. A similar relation connects the factorial cumulants of the former with the factorial moments of the latter.

Consider

$$G_X(z) = \exp\{\lambda[h_Y(z) - 1]\} \qquad (2.2.5.1)$$

where X and Y denote the random variables of the GPD and its generalizing component respectively. The cumulant generating function of X is given by

$$K(t) = \log G(e^t) = \lambda[h(e^t) - 1]. \qquad (2.2.5.2)$$

It follows therefore that the rth cumulant, ${}_xk_r$, is given by

$$\begin{aligned} {}_xk_r &= \lambda\left[\frac{d^r}{dt^r} h(e^t)\right]_{t=0} \\ &= \lambda({}_y\mu_r'). \end{aligned} \qquad (2.2.5.3)$$

Similarly, the factorial cumulant generating function of X is given by

$$K_F(t) = \log G(t+1) = \lambda[h(t+1) - 1]$$

from which the rth factorial cumulant is obtained as

$$\kappa^{k}_{(r)} = \lambda\left[\frac{d^r}{dt^r} h(t+1)\right]_{t=0}$$

$$= \lambda\left({}_{r}\mu'_{(r)}\right). \qquad (2.2.5.4)$$

2.2.6 Khatri-Patel recurrence formula for the probabilities

Khatri and Patel (1961) have given a recurrence formula for computing the probabilities of any GPD when the probability function of its cluster-size distribution is known. A simple derivation based on a technique discussed by Kemp (1967b) is as follows.

Suppose the pgf of a GPD is $G(z) = \Sigma P_x z^x$ and that of its cluster-size distribution is $h(z) = \Sigma p^*_x z^x$.

$$G(z) = \exp\{\lambda[h(z) - 1]\} \qquad (2.2.6.1)$$

$$G'(z) = \lambda h'(z)\, G(z) \qquad (2.2.6.2)$$

i.e. $$\Sigma x P_x z^{x-1} = \lambda(\Sigma x p^*_x z^{x-1})(\Sigma P_x z^x) \qquad (2.2.6.3)$$

$$\Sigma[(x+1)P_{x+1}]z^x = \lambda\Sigma(x+1)p^*_{x+1} z^x \,.\, \Sigma P_x z^x$$

$$= \Sigma\left[\lambda \sum_{r=0}^{x} (r+1)p^*_{r+1} P_{x-r}\right] z^x$$

i.e.

$$(x+1)P_{x+1} = \lambda \sum_{r=0}^{x} (r+1)p^*_{r+1} P_{x-r}. \qquad (2.2.6.4)$$

A simple recursive formula often also exists for $(r+1)p^*_{r+1}$, in which case the Khatri-Patel formula can be implemented in two tiers e.g. for the Neyman Type A with the pgf

$$G(z) = \exp\{\beta[e^{\alpha(z-1)} - 1]\} \qquad (2.2.6.5)$$

we have

$$p_0^* = e^{-\alpha}, \quad (r+1)p_{r+1}^* = \alpha p_r^*, \quad r = 0, 1, \ldots \tag{2.2.6.6}$$

and the above Khatri-Patel formula reduces to

$$a_0 = e^{-\alpha}, \quad P_0 = \exp[-\beta(1 - e^{-\alpha})]$$

$$a_{r+1} = \alpha a_r, \qquad r = 0, 1, \ldots$$

$$(x+1)P_{x+1} = \beta \sum_{r=0}^{x} a_r P_{x-r}, \qquad x = 0, 1, \ldots \tag{2.2.6.7}$$

2.2.7 Examples of univariate GPDs

The negative binomial distribution is the first-known and the best-known of all overdispersed discrete distributions, that is, distributions whose variances exceed their means. Anscombe (1950) recalls that in Todhunter's History this distribution is said to have been described by Montmort in 1714. Greenwood and Yule (1920), in their frequently cited work, discuss it as the gamma mixture of the Poisson distribution, but later Lüders (1934) and Quenouille (1949) show that it is not only a compound Poisson distribution but also a generalized Poisson distribution, the logseries being its cluster-size distribution.

Its probability generating function is

$$\begin{aligned} G(z) &= (1 + \beta - \beta z)^{-\alpha}, \quad \alpha > 0, \beta > 0 \\ &= [(1-\theta)/(1-\theta z)]^{\alpha}, \quad \theta = \beta/(\beta+1) \\ &= {}_1F_0(\alpha; -; \theta z)/{}_1F_0(\alpha; -; \theta) \\ &= \exp\{\lambda[{}_2F_1(1, 1; 2; \theta z)/{}_2F_1(1, 1; 2; \theta) - 1]\} \\ &= \exp\{\gamma[\log(1-\theta z)/\log(1-\theta) - 1]\}. \end{aligned} \tag{2.2.7.1}$$

A useful recurrence relation between its successive probabilities is given by

$$P_0 = (1-\theta)^{\alpha}$$

$$(x+1)P_{x+1} = \theta(x+\alpha)P_x, \quad x = 0, 1, 2, \ldots \tag{2.2.7.2}$$

The properties and applications of the negative binomial have been reviewed by Bartko (1961) and Johnson and Kotz (1969, pp.122-142).

Lüders (1934) considers the convolution of the Poisson and the negative binomial distributions, a distribution which is a GPD having the cluster-size distribution

$$h(z) = \theta_2 z + (1 - \theta_2) \log(1 - \theta_1 z)/\log(1 - \theta_1), \qquad (2.2.7.3)$$

$0 < \theta_1 < 1$, $0 < \theta_2 < 1$ (see also section 2.4.4).

The generalising of the Poisson distribution by another Poisson gives the Neyman Type A distribution having the pgf

$$G(z) = \exp\{\beta[e^{\alpha(z-1)} - 1]\}, \qquad \alpha > 0, \beta > 0. \qquad (2.2.7.4)$$

As has been mentioned before, this GPD is equivalent to Poisson V zero-truncated Poisson having the pgf

$$G(z) = \exp\{\lambda[(e^{\alpha z} - 1)/(e^{\alpha} - 1) - 1]\}. \qquad (2.2.7.5)$$

To compute the probabilities of the NTA distribution two formulae are available, each of which is a two-tier relation. The commoner is that given in Section 2.2.6 as a special case of Khatri and Patel's (1961) general recursive formula for the probabilities of a GPD. The other is in terms of the Stirling numbers $S_m(n)$ (Douglas (1970)) and is given by

$$S_0(k) = 0, \quad S_k(k) = 0, \quad k = 1, 2, \ldots$$
$$S_i(k + 1) = S_{i-1}(k) - kS_i(k),$$
$$P_x = C \frac{\alpha}{x!} \sum_{r=1}^{x} \lambda^r S_x(r), \quad r = 1, 2 \qquad (2.2.7.6)$$

with $P_0 = \exp[-\beta(1 - e^{-\alpha})]$, $\lambda = \beta e^{-\alpha}$, and $C = \exp(1 - \beta)$. Martin and Katti (1962) have given an approximation for the probability function.

Cresswell and Froggatt (1963) discuss the convolution of the Poisson and the Neyman Type A distributions and call it the "Short" distribution, the Neyman Type A being referred to as the "Long" distribution. The 'Short' distribution can be viewed as a GPD having the cluster-size distribution

$$h(z) = \theta z + (1 - \theta)e^{\alpha(z-1)}, \qquad \alpha > 0, 0 < \theta < 1 \qquad (2.2.7.7)$$

Properties of this distribution have been discussed by Kemp (1967a).

The Thomas distribution (Thomas (1949)) is related to the Neyman Type A distribution, being the generalizing of the Poisson by the shifted Poisson. Its pgf is

$$G(z) = \exp\{\beta[ze^{\alpha(z-1)} - 1]\}. \qquad (2.2.7.8)$$

Neyman's Type B and Type C distributions are rarely used. These together with the Type A are special cases of Gurland's (1958) distribution resulting from generalizing the Poisson(λ) with a generalized hypergeometric factorial moment distribution having the pgf

$$h(z) = {}_1F_1(\alpha;\ \alpha + \beta;\ \lambda_2(z - 1)), \quad \alpha > 0,\ \beta \geq 0. \qquad (2.2.7.9)$$

The subsystem of this GPD for which $\alpha = 1$ is equivalent to Beall and Reacia's (1953) distribution obtained as a generalization of Neyman's. Neyman's Types A, B, C respectively are the special cases $\beta = 0, 1, 2$ of Beall and Reacia's.

The Polya-Aeppli distribution (e.g. Evans (1953), Kemp (1967b)), considered as Poisson shifted-geometric, has the pgf

$$G(z) = \exp\left\{\lambda\left[z\left(\frac{1 - \theta}{1 - \theta z}\right) - 1\right]\right\}, \quad \lambda > 0,\ 0 < \theta < 1. \qquad (2.2.7.10)$$

It is in fact equivalent to the Poisson-geometric having the pgf

$$G(z) = \exp\left\{\alpha\left[\left(\frac{1 - \theta}{1 - \theta z}\right) - 1\right]\right\}, \quad 0 < \theta < 1,\ \alpha = \lambda\theta > 0 \qquad (2.2.7.11)$$

(see Section 2.2.4). A recurrence relation between its probabilities is given by

$$(x + 1)P_{x+1} = [2\theta x + \lambda(1 - \theta)]P_x - \theta^2(x - 1)P_{x-1}. \qquad (2.2.7.12)$$

Skellam (1952) and Katti and Gurland (1961) have discussed the Poisson-negative binomial distribution. This has the pgf

$$G(z) = \exp\{\lambda[(1 + \beta - \beta z)^{-\alpha} - 1]\}, \quad \alpha > 0,\ \beta > 0,\ \lambda > 0, \qquad (2.2.7.13)$$

the special case $\alpha = 1$ being the Poisson-geometric. Its probabilities can be computed with the two-tier formula

$$a_0 = (\beta + 1)^{-\alpha}, \quad P_0 = \exp\{-\lambda[1 - (\beta + 1)^{-\alpha}]\}$$

$$a_{r+1} = (\beta/(\beta + 1))(r + \alpha)a_r, \qquad r = 0, 1, \ldots$$

$$(x + 1)P_{x+1} = \sum_{r=0}^{x} a_r P_{x-r}, \qquad x = 0, 1, \ldots \qquad (2.2.7.14)$$

The Poisson-binomial distribution (Skellam (1952) and McGuire et al. (1957)) has the pgf

$$G(z) = \exp\{\lambda\,((1 - \theta) + \theta z)^n - 1\}, \qquad (2.2.7.15)$$

$0 < \theta < 1$, $\lambda > 0$, n a positive integer,

which alternatively, is

$$G(z) = \exp\left\{\beta\left[\left(\frac{\alpha + z}{\alpha + 1}\right)^n - 1\right]\right\}, \qquad (2.2.7.16)$$

$\alpha > 0$, $\beta > 0$, n a positive integer.

The special case n = 2 is equivalent to Kemp and Kemp's (1965) Hermite distribution, which is more conveniently represented (with $\nu_1 = 2\lambda\theta(1 - \theta)$, $\nu_2 = \lambda\theta^2$) as

$$G(z) = \exp\{\nu_1(z - 1) + \nu_2(z^2 - 1)\}, \ \nu_1 > 0, \ \nu_2 > 0 \qquad (2.2.7.17)$$

The probabilities of the Poisson-binomial can be obtained with the two-tier recursive formula

$$a_0 = (\alpha/(\alpha + 1))^n, \quad P_0 = \exp\{-\lambda[1 - (\alpha/(\alpha + 1))^n]\}$$

$$a_{r+1} = -a_r(x - n)/(\alpha + 1)$$

$$(x + 1)P_{x+1} = \lambda \sum_{r=0}^{x} a_r P_{x-r}. \qquad (2.2.7.18)$$

Shumway and Gurland (1960a, 1960b) aim at obtaining recursive formulae less susceptible to rounding errors than the above. However the two which they have presented are in no way more satisfactory. They regard the one using the Stirling numbers $S_m(n)$ as the better of the two, but this has been dismissed by Douglas (1970) as being "not a practical method of computation for other than limited cases".

Here we present a simple recurrence relation for probabilities of the Poisson-binomial distribution. This is an (n + 1)-term formula and since n is usually not greater than 4 (McGuire et al. (1957)), it is

substantially less susceptible to rounding errors then the above two-tier formula or any of Shumway and Gurland's.

From the pgf G(z), using the approach of Kemp (1967b), we obtain

$$G'(z) = \left[\left(\frac{n\lambda}{\alpha+1}\right)\left(\frac{\alpha}{\alpha+1}\right)^{n-1}\right]\sum_{r=0}^{n-1}\binom{n-1}{r}\left(\frac{z}{\alpha}\right)^r G(z)$$

$$= c\sum_{x=0}^{\infty}\sum_{r=0}^{n-1} b_r P_x z^{x+r} \qquad (2.2.7.19)$$

which gives

$$(x+1)P_{x+1} = c\sum_{r=0}^{n-1} b_r P_{x-r} \qquad (2.2.7.20)$$

with $c = n\lambda\alpha^{n-1}/(\alpha+1)^n$

and $b_r = \left(\frac{1}{\alpha}\right)^r\binom{n-1}{r}, \qquad r = 0, 1, \ldots, n-1.$

The case n = 2 of this formula is given by

$$(x+1)P_{x+1} = \nu_1 P_x + 2\nu_2 P_{x-1} \qquad (2.2.7.21)$$

with $\nu_1 = 2\lambda\alpha/(\alpha+1)^2, \quad \nu_2 = \lambda/(\alpha+1)^2$

which, as expected, is the recurrence relation given by Kemp and Kemp (1965) for the probabilities of the Hermite distribution.

2.2.8 Bivariate generalized Poisson distributions

As mentioned in section 1.7.1, in some studies it is the bivariate rather than the univariate form of a discrete distribution which is of interest. The commonest way of deriving a bivariate GPD is to generalize a Poisson distribution with a (usually but not necessarily, correlated) bivariate lattice distribution with the pgf $g(z_1, z_2)$. The pgf of the resulting bivariate GPD then is

$$G(z_1, z_2) = \exp\{\lambda[g(z_1, z_2) - 1]\}. \qquad (2.2.8.1)$$

A bivariate GPD of the same form is obtained if a bivariate Poisson$(x_1, x_2; \lambda\alpha_1, \lambda\alpha_2, \lambda\alpha_{12})$ is mixed on λ by an inf div continuous distribution supported on the positive half real line or by a univariate

GPD. An example of this type of bivariate GPD is the bivariate Neyman Type A-I (Holgate, 1966) having the pgf

$$G(z_1, z_2) = \exp\{\lambda[\exp\{\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1) + \alpha_{12}(z_1 z_2 - 1)\}]\} \quad (2.2.8.2)$$

which is obtained by generalizing the univariate Poisson(x; λ) with the bivariate Poisson

$$g(z_1, z_2) = \exp[\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1) + \alpha_{12}(z_1 z_2 - 1)]. \quad (2.2.8.3)$$

Two other methods of forming a bivariate GPD consist of generalizing a correlated bivariate Poisson with two generalizers, $g_1(z_1)$ and $g_2(z_2)$, in the first case and three, $g_1(z_1)$, $g_2(z_2)$ and $g_{12}(z_1 z_2)$, in the second. The resulting pgfs, respectively are

$$G(z_1, z_2) = \exp\{\lambda_1[g_1(z_1) - 1] + \lambda_2[g_2(z_2) - 1] + \lambda_{12}[g_1(z_1)\, g_2(z_2) - 1]\} \quad (2.2.8.4)$$

and

$$G(z_1, z_2) = \exp\{\lambda_1[g_1(z_1) - 1] + \lambda_2[g_2(z_2) - 1] + \lambda_{12}[g_{12}(z_1 z_2) - 1]\} \quad (2.2.8.5)$$

The latter is equivalent to the joint distribution of the correlated variables $X_1 = Y_1 + Y_3$, $X_2 = Y_2 + Y_3$ where Y_1, Y_2, Y_3 are three independent univariate GPD variables of the generalizers. Bivariate Neyman Type A-II and bivariate Neyman Type A-III of Holgate (1966) are examples of the two methods. These have, respectively, the pgfs

$$G(z_1, z_2) = \exp\{\lambda_1[\exp\{\alpha_1(z_1 - 1)\} - 1] + \lambda_2[\exp\{\alpha_2(z_2 - 1)\} - 1] + \lambda_{12}[\exp\{\alpha_1(z_1 - 1) + \alpha_2(z_2 - 1)\} - 1]\} \quad (2.2.8.6)$$

and

$$G(z_1, z_2) = \exp\{\lambda_1[\exp\{\alpha_1(z_1 - 1)\} - 1] + \lambda_2[\exp\{\alpha_2(z_2 - 1)\} - 1] + \lambda_{12}[\exp\{\alpha_{12}(z_1 z_2 - 1)\} - 1]\}. \quad (2.2.8.7)$$

2.3 Mixtures of the Poisson with Continuous Mixing Distributions

2.3.1 Examples of mixtures of the Poisson

The "spells" model and the "proneness" model are the commonest of the tools proposed for investigating the tendency of a vehicle driver to sustain traffic accidents (e.g. Kemp (1970)). The former is a clustering model while the latter is a mixing model. On the assumption of randomness for the occurrence of accidents, generalized Poisson distributions (GPDs) are therefore used in the "spells" model and compound Poisson distributions (CPDs), i.e. Poisson mixtures, in the proneness model. However, most of the distributions employed are known to belong simultaneously to both the GPD and the CPD classes. Hence some discussion of CPDs is necessary in this thesis. The concept of mixture in general has been introduced in section 1.5.2. Here we are concerned only about mixtures of the Poisson distribution and only to the extent of their relationship with GPDs.

Mixtures of the Poisson obtained with the following continuous distributions supported on (the whole or part of) the positive half real line, i.e. on $A \subseteq R_+$, have been discussed in the literature, most of them in connection with the concept of accident proneness.

(a) The gamma distribution defined by

$$f(x) = (\beta^{\alpha}/\Gamma(\alpha))x^{\alpha-1}e^{-\beta x}, \quad x > 0, \alpha > 0, \beta > 0 \qquad (2.3.1.1)$$

mixes the Poisson distribution to give the well known negative binomial with the pgf

$$\begin{aligned} G(z) &= (1 + \beta - \beta z)^{-\alpha} \\ &= [(1 - \theta)/(1 - \theta z)]^{\alpha}, \quad \theta = 1/(\beta + 1) \\ &= {}_1F_0(\alpha; -; \theta z)/{}_1F_0(\alpha; -; \theta) \end{aligned} \qquad (2.3.1.2)$$

This CPD is employed in the commonly cited work of Greenwood and Yule (1920) on accident proneness. The special case $\alpha = 1$, the exponential mixture of the Poisson distribution, is the geometric distribution, an important distribution in its own right.

(b) The rectangular distribution having the density

$$f(x) = \frac{1}{\beta - \alpha}, \qquad 0 < \alpha < x < \beta, \tag{2.3.1.3}$$

as the mixing distribution for a mixture of the Poisson, gives the CPD with the probability function

$$P_x = \frac{1}{\beta - \alpha}\left[e^{-\alpha}\sum_{r=0}^{x}\frac{\alpha^r}{r!} - e^{-\beta}\sum_{r=0}^{x}\frac{\beta^r}{r!}\right]$$

$$= \frac{1}{\beta - \alpha}\frac{1}{x!}\left[\Gamma_*(x + 1, \beta) - \Gamma_*(x + 1, \alpha)\right] \tag{2.3.1.4}$$

where $\Gamma_*(c, u)$, the incomplete gamma function, is given by

$$\Gamma_*(c, u) = \int_0^u t^{c-1}e^{-t}\,dt.$$

This mixture of the Poisson has been discussed by Bhattacharya and Holla (1965) who suggest its application in the accident proneness problem.

(c) The Beta-I family is defined by the density

$$f(x) = \frac{1}{\beta(\alpha, \beta)}\frac{1}{\gamma^{(\alpha+\beta)-1}}x^{\alpha-1}(\gamma - x)^{\beta-1} \tag{2.3.1.5}$$

$0 < x < \gamma$, $\alpha > 0$, $\beta > 0$. With this as the mixing distribution on the Poisson parameter, the resulting CPD has the pgf

$$G(z) = {}_1F_1(\alpha; \alpha + \beta; \gamma(z - 1)), \tag{2.3.1.6}$$

a member of Kemp and Kemp's (1974) system of generalized factorial moment distributions ((1.4.4.17)). The standard form is the case $\alpha = 1$. Holla and Bhattacharya (1965) have specifically discussed its application in the study of accident proneness. The special case $\beta = 1$ of the Beta-I family is known as the power function. The mixture of the Poisson obtained with it has been applied to accident proneness by Rai (1971).

(d) The Beta-II family is represented by the pdf

$$f(x) = \frac{\gamma^{\beta}}{\beta(\alpha, \beta)}\frac{x^{\alpha-1}}{(x + \gamma)^{\alpha+\beta}}, \quad x > 0, \alpha > 0, \beta > 0, \gamma > 0. \tag{2.3.1.7}$$

Note that it is a gamma mixture of the gamma: $\text{Gamma}(x; \alpha, u) \underset{u}{\wedge} \text{Gamma}(u; \beta, \gamma)$

(see also Johnson and Kotz (1970, p.195)). The pgf of the mixture of the Poisson with (2.3.1.7) as the mixer is given by

$$G(z) = [\Gamma(\alpha + \beta)/\Gamma(\beta)]_1F_1(\alpha; 1 - \beta; -\gamma(z - 1)) \qquad (2.3.1.8)$$

(cf Gradshetyn et al. (1968, p.1058, entry no. 9.211.2)), which is a valid pgf only for $0 < \beta < 1$. The standard form of the Beta-II family is the case $\gamma = 1$, and its associated CPD has been discussed by Holla and Bhattacharya (1965) who mention accident proneness studies as one of its possible areas of application. It appears however that they are unaware that the range of the valid values of the parameters is severely limited.

(e) Wani (1968) has defined a system of continuous distributions called the linear exponential family. This is represented by the density

$$f(x) = \frac{1}{u(\beta)} e^{-\beta x} h(x). \qquad (2.3.1.9)$$

where $h(x) \geq 0$ depending only on x, ranges over an interval A on R, the real line, and $u(\beta)$ a finite and differentiable function of β is given by

$$u(\beta) = \int e^{-\beta x} h(x)dx. \qquad (2.3.1.10)$$

It is clear that this distribution has finite moments and its mgf is given by

$$M(t) = \frac{u(\beta + t)}{u(\beta)}. \qquad (2.3.1.11)$$

Since the Poisson parameter takes only positive real values, we are interested only in f(x) supported on $A \subseteq R_+$.

In particular if h(x) is a function of $x > 0$ and has the Laplace transform (LT) L(s), i.e.

$$L(s) = \int_0^\infty e^{-sx} h(x)dx, \qquad s > 0, \qquad (2.3.1.12)$$

obviously

$$f(x) = \frac{1}{L(\beta)} e^{-\beta x} h(x) \qquad (2.3.1.13)$$

is a pdf. Similarly if $h(x)$, $x \in A \subseteq R_+$ is a pdf and has a moment generating function, M(t), i.e.

$$M(t) = \int_A e^{tx} h(x)dx, \qquad (2.3.1.14)$$

then obviously

$$f(x) = \frac{1}{M(\beta)} e^{\beta x} h(x), \qquad \beta > 0 \qquad (2.3.1.15)$$

is a pdf and has finite moments.

The linear-exponential mixture of the Poisson distribution has the pgf given by

$$G(z) = U(\beta + z - 1)/U(\beta) \qquad (2.3.1.16)$$

and has been discussed by Sankaran (1970a).

It is obvious that the gamma distribution is a member of the linear exponential family. Other members include Lindley's (1958) distribution and Bhattacharya's (1966) distribution. Both are among the distributions considered in section 2.5.3 below.

(f) The distribution resulting from truncating that part of the normal distribution supported on R_-, the negative half real line has the density

$$f(x) = \frac{1}{C_o\left(-\frac{\mu}{\sigma}\right)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x > 0 \qquad (2.3.1.17)$$

where $$C_o(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-t^2/2} dt.$$

Patil (1964) has discussed the mixture of the Poisson obtained with the distribution. An explicit form of the probability function of the mixture is not yet known. However Kemp and Kemp (1967) have provided a recursive formula for its probabilities. They also show that it is a special case of the mixture of the Poisson discussed by Fisher (1931), that is, that obtained with the three-parameter distribution given by

$$f(x) = \frac{1}{C_m(a)} \frac{x^m}{m!\sigma^{m+1}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma} + a\right)^2\right], \quad x > 0 \qquad (2.3.1.18)$$

where $$C_m(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty \frac{(t-u)^m}{m!} e^{-t^2/2} dt.$$

(g) The density function of the lognormal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left\{ - \frac{1}{2\sigma^2} (\log x - \mu)^2 \right\} \qquad (2.3.1.19)$$

$x > 0$, $\sigma > 0$, $-\infty < \mu < \infty$. The CPD obtained from this distribution has been of considerable interest in the study of species aboundance since its use was suggested by Preston (1948), despite the fact that no closed form of its probability function is yet available. Other discussions of its application in this context include those of Anscombe (1950), Grundy (1951), Cassie (1962), Bliss (1963), Brown and Holgate (1971), Bulmer (1974) and Reid (1981). It is the zero-truncated form which was used by Preston and Bliss. Grundy, and Brown and Holgate present tables of the probabilities.

(h) The Laplace transform of the function

$$u(x) = x^{\gamma-1} e^{-\frac{\beta}{x}} \qquad (2.3.1.20)$$

is given by

$$L(s) = 2\left(\frac{\beta}{s}\right)^{\frac{\gamma}{2}} K_\gamma(2\sqrt{\beta s}), \qquad s > 0 \qquad (2.3.1.21)$$

(e.g. Erdelyi (1954, p.146)). Hence the function

$$f(x) = e^{-\alpha x} u(x)/L(\alpha), \qquad \alpha > 0$$

$$= \frac{1}{2}\left[\left(\frac{\alpha}{\beta}\right)^{\frac{\gamma}{2}} / K_\gamma(2\sqrt{2\alpha\beta})\right] x^{\gamma-1} e^{-(\alpha x + \frac{\beta}{x})}, \qquad (2.3.1.22)$$

$x > 0$, $\alpha > 0$, $\beta > 0$, $-\infty < \gamma < \infty$, is a pdf. This is the generalized inverse normal distribution (e.g. Jorgensen (1982)). The mixture of the Poisson obtained from this distribution has engaged the attention of Sichel (1971, 1973a, 1973b, 1974, 1975, 1982a, 1982b). The pgf of the mixture is

$$G(z) = \frac{1}{K_\gamma(2\sqrt{\alpha\beta})} \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-\frac{\gamma}{2}} K_\gamma\left(2\sqrt{\alpha\beta}\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{\frac{1}{2}}\right) \qquad (2.3.1.23)$$

Its probability function is

$$P_x = \left[\left(\frac{\alpha}{\alpha + 1}\right)^{\frac{\gamma}{2}} / K_\gamma(2\sqrt{\alpha\beta})\right] \frac{[\beta/(\alpha + 1)]^{\frac{\alpha}{2}}}{x!} K_{x+\gamma}(2\sqrt{\beta(\alpha + 1)}) \qquad (2.3.1.24)$$

A recurrence relation exists between the probabilities. This is

$$P_{x+1} = \left(\frac{\alpha}{\alpha + 1}\right)\left(\frac{x + \gamma}{x + 1}\right)P_x + \left(\frac{\beta\alpha^2}{\alpha + 1}\right)\frac{1}{x(x + 1)}P_{x-1} \qquad (2.3.1.25$$

It is the special case $\gamma = -\frac{1}{2}$ which has received the greatest attention. This is the inverse-normal mixture of the Poisson distribution. Holla's (1966) seems to be the first discussion of the distribution but appears to be unknown to Sichel. Others include Sankaran (1968) and Shaban (1981, 1982). Atkinson and Yeh (1982) consider the subsystem for which $\gamma = n + \frac{1}{2}$ where n is an integer. The generalized-inverse-normal mixture of the Poisson distribution is discussed further in sections 2.4.3 and 2.5.2.

(i) The Laplace transform (LT) of the function $u_0(x) = x^{b-1}{}_1F_1(a; c; \gamma x)$, $x > 0$, $b > 0$ is

$$L_0(s) = \frac{\Gamma(b)}{s^b}{}_2F_1\left(a, b; c; \frac{\gamma}{s}\right), \quad s > 0, s > |\gamma|. \qquad (2.3.1.26$$

If $b = c$ and $\gamma = 1$, $u_0(x)$ becomes

$$u(x) = x^{b-1}{}_1F_1(a; b; x)$$

and the corresponding LT reduces to

$$L(s) = \frac{\Gamma(b)}{s^b}{}_1F_0\left(a; -; \frac{1}{s}\right)$$

$$= \frac{\Gamma(b)}{s^b}\left(\frac{s}{s - 1}\right)^a$$

$$= \Gamma(b)s^{a-b}/(s - 1)^a. \qquad (2.3.1.27$$

Hence

$$f(x) = e^{-\lambda x}u(x)/L(\lambda) \quad \text{(see (e) above)}$$

$$= \frac{\lambda^{(b-a)}}{\Gamma(b)}(\lambda - 1)^a x^{b-1}e^{-\lambda x}{}_1F_1(a; b; x) \quad x > 0, \lambda > 1$$

$$= \frac{\alpha^a}{\Gamma(b)}(\alpha + 1)^{b-a}x^{b-1}e^{-(\alpha+1)x}{}_1F_1(a; b; x)$$

$$x > 0, \alpha = \lambda - 1 > 0 \qquad (2.3.1.28$$

is a pdf. This is the distribution discussed by Bhattacharya (1966)

who also obtains its associated CPD and considers the application of the latter on the accident proneness problem. The mixture of the Poisson has the pgf

$$G(z) = \left(1 + \frac{1}{\alpha + 1} - \frac{z}{\alpha + 1}\right)^{a-b} / \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{a}. \qquad (2.3.1.29)$$

This mixture will be considered further in section 2.5. The special case $a = 2$, $b = 1$ is the Poisson-Lindley distribution discussed by Sankaran (1970b).

2.3.2 The overdispersion of mixtures of the Poisson

As has been noted in section 1.5.3, a common area of application of generalized distributions is in the study of spatial distributions of the populations of plants, bacteria, and insects. In this area, the populations are classified into three types: the regular, the random, and the over-dispersed, according to whether the variance is less than, equal to, or greater than the mean respectively. Overdispersed populations seem to be the commonest.

Generalized Poisson distributions are the most widely used generalized distributions in investigating spatial pattern. Here we show that it is not only GPDs but also CPDs which are overdispersed. This feature poses major difficulties in data analysis.

Suppose X is a generalized Poisson variable resulting from the generalizing of the Poisson distribution by the distribution of a (discrete) random variable Y, the latter having the pgf $h_Y(z)$. Then the pgf of X is given by

$$G_X(z) = \exp\{\lambda[h_Y(z) - 1]\}. \qquad (2.3.2.1)$$

From this we have

$$G_X'(z) = \lambda h_Y'(z)\, G_X(z) \qquad (2.3.2.2)$$

$$\begin{aligned} G_X''(z) &= \lambda h_Y'(z)\, G_X'(z) + \lambda h_Y''(z)\, G_X(z) \\ &= [\lambda^2 (h_Y'(z))^2 + \lambda h_Y''(z)] G_X(z) \end{aligned} \qquad (2.3.2.3)$$

$$\text{Mean}(X) = G_X'(1)$$

$$= \lambda h_Y'(1)$$

$$= \lambda \text{Mean}(Y) \qquad (2.3.2.4)$$

$$E_X(X^2) - E_X(X)$$

$$= G_X''(1)$$

$$= \lambda^2 (h_Y'(1))^2 + \lambda h_Y''(1)$$

$$= (\text{Mean}(X))^2 + \lambda h_Y''(1)$$

i.e.
$$E_X(X^2) = \text{Mean}(X) + (\text{Mean}(X))^2 + \lambda h_Y''(1) \qquad (2.3.2.5)$$

$$\text{Var}(X) = E_X(X^2) - (\text{Mean}(X))^2$$

$$= \text{Mean}(X) + \lambda h_Y''(1)$$

$$= \lambda [h'(1) + h''(1)]$$

$$= \lambda E_Y(Y^2)$$

$$= \frac{\text{Mean}(X)}{\text{Mean}(Y)} \cdot E_Y(Y^2) \qquad (2.3.2.6)$$

i.e.

$$\frac{\text{Var}(X)}{\text{Mean}(X)} = \frac{E_Y(Y^2)}{E_Y(Y)}$$

$$> 1 \text{ if } h''(1) = E_Y(Y^2) - E(Y) > 0 \qquad (2.3.2.7)$$

Note that

$$h''(1) = \sum_{y \geq 2} y(y-1) p_y^* \qquad (2.3.2.8)$$

which is greater than zero if $p_y^* > 0$ for any (integer) $y > 1$. The cluster size distribution $h(z) = q + pz$ gives rise to a Poisson distribution as a trivial GPD. Excluding this and $h(z) = 1$, every other choice of $h(z)$ gives $h''(1) > 0$ and hence $\text{Var}(X) > \text{Mean}(X)$. Hence it follows that if we exclude consideration of the Poisson as a GPD, all GPDs are overdispersed.

This implies that a necessary condition for a lattice distribution, other than the Poisson, to be a GPD is that it be overdispersed i.e. that its variance exceed its mean. This condition is however not a sufficient condition. For example, Kemp (1978) has shown that the shifted zero-truncated Poisson distribution is not a GPD. Its pgf is

$$G(z) = (e^{\lambda z} - 1)/[z(e^{\lambda} - 1)]$$
$$= {}_1F_1(1, 2; \lambda z)/{}_1F_1(1; 2; \lambda). \qquad (2.3.2.9)$$

It is obvious that it is a special case of Bardwell and Crow's (1964) hyper-Poisson distribution represented by the pgf

$$G(z) = {}_1F_1(1; \alpha; \lambda z)/{}_1F_1(1; \alpha; \lambda), \quad \alpha > 0. \qquad (2.3.2.10)$$

However, as mentioned in section 1.4.4, when $\alpha > 1$ the distribution is said to be super-Poisson since in that case its variance exceeds its mean, that is, the distribution is overdispersed. Hence the shifted zero-truncated Poisson distribution is overdispersed despite that it is not a GPD.

To show the overdispersion of mixtures of the Poisson, suppose that a random variable X is a compound Poisson variable arising from mixing the Poisson parameter with a discrete or continuous mixing variable Y. Then its pgf is given by

$$G_X(z) = \int_S e^{cy(z-1)} f(y)dy$$
$$= E_Y(e^{\{c(z-1)\}y})$$
$$= M_Y(c(z - 1)) \qquad (2.3.2.11)$$

where c is an arbitrary constant, $M_y(t)$ is the mgf of Y, and the integral symbol is to be interpreted in the general Stieltjes sense.

$$\text{Mean}(X) = G_X'(1)$$
$$= cM_Y'(c(z - 1))\Big|_{z=1}$$
$$= cM_Y'(0)$$
$$= c\text{Mean}(Y) \qquad (2.3.2.12)$$

$$E_X(X^2) - E_X(X)$$
$$= G_X''(1)$$
$$= c^2E_Y(Y^2) \qquad (2.3.2.13)$$

i.e.

$$E_X(X^2) = \text{Mean}(X) + c^2E_Y(Y^2) \qquad (2.3.2.14)$$

$$\begin{aligned} \mathrm{Var}(X) &= E_x(X^2) - (E_x(X))^2 \\ &= \mathrm{Mean}(X) + c^2[E_Y(Y^2) - (E_Y(Y))^2] \\ &= \mathrm{Mean}(X) + c^2\,\mathrm{Var}(Y) \end{aligned} \qquad (2.3.2.15)$$

i.e.

$$\mathrm{Var}(X) > \mathrm{Mean}(X) \qquad (2.3.2.16)$$

Hence all compound Poisson distributions are overdispersed.

2.3.3 Factorial moments and factorial cumulants

A simple relation exists between the factorial moments of a compound Poisson distribution and the moments of the compounder, and a similar one between the factorial cumulants of the former and the cumulants of the latter.

Let X be a compound Poisson variable and Y the (discrete or continuous) mixing variable

$$\begin{aligned} G_x(z) &= \int_s e^{cy(z-1)} f(y)dy \\ &= M_Y(c(z-1)) \end{aligned} \qquad (2.3.3.1)$$

The factorial moment generating function of X is given by

$$G_x(z+1) = M_Y(cz) \qquad (2.3.3.2)$$

Note that

$$\frac{d^r}{dz^r} G_x(z+1) = c^2\left[\frac{d^r}{dz^r} M_Y(cz)\right] \qquad (2.3.3.3)$$

The rth factorial moment of X is

$$\begin{aligned} {}_x\mu_{(r)} &= \left[\frac{d^r}{dz^r} G_x(z+1)\right]_{z=0} \\ &= c^r\left[\frac{d^r}{dz^r} M_Y(cz)\right]_{z=0} \\ &= c^r({}_Y\mu_r') \end{aligned}$$

i.e.

$${}_x\mu_{(r)} = c^2({}_Y\mu_r'). \qquad (2.3.3.4)$$

If Y is a continuous variable, it is usual to have c = 1. In this case

$$x\mu_{(r)} = {}_Y\mu'_r. \tag{2.3.3.5}$$

Similarly, the factorial cumulant generating function of X is given by

$$\log G_X(z + 1) = \log M_Y(cz) \tag{2.3.3.6}$$

from which we have

$$\frac{d^r}{dz^r} \log G_X(z + 1) = c^r \left[\frac{d^r}{dz^r} \log M_Y(cz)\right] \tag{2.3.3.7}$$

The rth factorial cumulant of X is obtained as

$$x\kappa_{(r)} = \left[\frac{d^r}{dz^r} \log G_X(z + 1)\right]_{z=0}$$

$$= c^r \left[\frac{d^r}{dz^r} \log M_Y(cz)\right]_{z=0}$$

$$= c^r ({}_Y\kappa_r)$$

i.e.

$$x\kappa_{(r)} = c^r ({}_Y\kappa_r). \tag{2.3.3.8}$$

2.4 Mixtures of the Poisson with Infinitely Divisible Continuous Mixing Distributions

2.4.1 Maceda's theorem

It has been mentioned already that a mixture of an inf div distribution is not necessarily infinitely divisible. It follows therefore that the inf divisibility of the Poisson distribution does not necessarily extend to every mixture of the Poisson. However an important type of mixture of the Poisson is inf div, as stated in the following theorem which is from the literature (e.g. Maceda (1948)).

Theorem 3

A mixture of the Poisson distribution produced by a discrete or continuous inf div mixer is also infinitely divisible. In other words every compound Poisson distribution arising from an inf div mixing distribution is also a generalized Poisson distribution.

It is obvious that if the (inf div) mixing distribution is a discrete distribution then its mixture of the Poisson is a GPD. Generalizing a GPD yields another GPD. Mixing the Poisson by a discrete distribution, D, is equivalent to mixing, on n, the n-fold convolution of the Poisson and hence the resulting mixture is a generalized D-distribution (see Section 2.2). That is, the D-mixture of the Poisson is a generalized Poisson distribution. If D is continous and inf div, the infinite divisibility of the D-mixture of the Poisson is not quite obvious but has been shown in the literature (e.g. Maceda (1948)).

This theorem is applied here in showing the infinite divisibility of certain mixtures of the Poisson distribution.

Recently a great number of continuous distributions supported on $A \subseteq R_+$ have been shown to be infinitely divisible. However because the relevant integrals are not yet obtainable, the derivation of mixtures of the Poisson with these distributions as mixers has not received attention.

Bondesson (1978) has shown that the distributions of powers of a gamma variable are inf divisible. These include members of Stacy's (1962) system represented by

$$f(x) = \frac{\gamma}{\beta^{\alpha\gamma}\Gamma(\alpha)} x^{\alpha\gamma-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\gamma}\right], \qquad (2.4.1.1)$$

$x > 0$, $\alpha > 0$, $\beta > 0$, $\gamma > 0$ as well as the reciprocal gamma distribution with the density

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x} \qquad (2.4.1.2)$$

$x > 0$, $\alpha > 0$, $\beta > 0$. The latter will be discussed further in Section 2.4.3.

Steutel (1970, pp.41-42) has shown that all completely monotone densities on $(0, \infty)$ (and all their mixtures) are infinitely divisible (see also Feller (1971, pp.416-417)). Harn (1978, p.26) states that completely monotone densities are equivalent to mixtures of the exponential distribution i.e. a pdf on $(0, \infty)$ is completely monotone if and only if it is a mixture of the exponential distribution. Steutel (1968) (also Goldie (1967)) separately proves that all mixtures of the exponential are inf divisible. Note that a pdf $f(x)$ $x > a \geq 0$ is said to be completely monotone if it possesses derivatives of all order with

$$(-1)^n \frac{d^n}{dx^n} f(x) \geq 0, \qquad n \in N_+. \qquad (2.4.1.3)$$

Examples of such densities are

$$f(x) = \alpha/(x + 1)^{\alpha+1}, \qquad x > 0, \alpha > 0, \qquad (2.4.1.4)$$

$$f(x) = \alpha/x^{\alpha+1}, \qquad x > 1, \alpha > 0, \qquad (2.4.1.5)$$

$$f(x) = \alpha\beta x^{\beta-1} \exp(-\alpha x^{\beta}), \quad x > 0, 0 < \alpha \leq 1, 0 < \beta \leq 1, \qquad (2.4.1.6)$$

and $$f(x) = \frac{1}{e - 1} \cdot \exp(e^{-x} - x), \; x > 0. \qquad (2.4.1.7)$$

Note also that if $u(x)$ is a completely monotone density so are $u(\alpha x)$, $u(x + \alpha)$ and $u(x) - u(x + \alpha)$, and if the densities $u(x)$, $v(x)$ are completely monotone so are their sum and product $u(x) + v(x)$, $u(x)\, v(x)$. With these properties the number of available infinitely divisible continuous distributions $f(x)$, $x \in A \subseteq R_+$ can be considerably extended. Of course, a continuous distribution $f(x)$ may be inf div while an $f(x)$-mixture of the Poisson does not exist. For example, expression (2.4.1.4) gives an infinite divisible special case of the Beta-II distribution. The existence of a mixture of the Poisson with this distribution as the mixer depends on the finiteness, or otherwise, of

$$P_x = \frac{\alpha}{x!} \int_0^\infty e^{-\lambda} \frac{\lambda^x d\lambda}{(\lambda + 1)^{\alpha+1}}$$

$$= \frac{\alpha}{x!} I_0 \qquad (2.4.1.8)$$

i.e. on the convergence of the integral

$$I_0 = \int_0^\infty e^{-\lambda} \frac{\lambda^x d\lambda}{(\lambda + 1)^{\alpha+1}} . \qquad (2.4.1.9)$$

This integral is not convergent as x tends to infinity.

Steutel (1970) has shown further that mixtures of certain special cases of the gamma distribution are infinitely divisible. These will be discussed in section 2.4.2 below.

The same author (1970, p.88) remarks that if f(x) is inf divisible, then

$$h(x) = Ce^{\beta x} f(x) \qquad (2.4.1.10)$$

is an inf div density, where

$$C = \int e^{\beta x} f(x) dx, \qquad (2.4.1.11)$$

provided that C is finite (see section 2.3.1(e)).

The Laplace transforms or moment generating functions of some densities on $A \subseteq R_+$ are known in a base-exponent form with an arbitrary positive number as the exponent parameter. Certain Bessel-type densities are of this form.

Feller (1971, pp.503 and 60, 437-438) discusses a Bessel-type density

$$f_1(x) = \frac{\lambda}{x} e^{-x} I_\lambda(x), \qquad x > 0, \lambda > 0 \qquad (2.4.1.12)$$

$$= \frac{\lambda}{2^\lambda \Gamma(\lambda + 1)} x^{\lambda-1} e^{-2x} {}_1F_1((\lambda + \tfrac{1}{2}); (\lambda + \tfrac{1}{2}) + (\lambda + \tfrac{1}{2}); 2x) \qquad (2.4.1.13)$$

having the Laplace transform

$$L_1(s) = [(1 + s) - \sqrt{(1 + s)^2 - 1}\,]^\lambda. \qquad (2.4.1.14)$$

The pgf of the $f_1(x)$-mixture of the Poisson obviously is

$$G_1(z) = [(2 - z) - \sqrt{(2 - z)^2 - 1}\,]^\lambda. \qquad (2.4.1.15)$$

A two-parameter variant of $f_1(x)$ is easily obtained by using Feller's (1971, p.436) result and thereby having

$$f_2(x) = \frac{\lambda}{x} e^{-\frac{x}{\alpha}} I_\lambda\left\{\frac{x}{\alpha}\right\}, \qquad x > 0,\ \lambda > 0,\ \alpha > 0 \tag{2.4.1.16}$$

with its Laplace transform as

$$L_2(s) = [(1 + \alpha s) - \sqrt{(1 + \alpha s)^2 - 1}\,]^{\lambda}. \tag{2.4.1.17}$$

The pgf of the mixture of the Poisson obtained with this distribution is

$$G_2(z) = [(1 + \alpha - \alpha z) - \sqrt{(1 + \alpha - \alpha s)^2 - 1}\,]^{\lambda}. \tag{2.4.1.18}$$

The same author (1971, pp.59, 177, 437-438) gives another Bessel-type density. It is

$$f_3(x) = \left\{\lambda^{-\alpha/2} e^{-\lambda}\right\} x^{\alpha/2} e^{-x} I_\lambda(2\sqrt{\lambda x}), \quad x > 0,\ \alpha > 0,\ \lambda > 0 \tag{2.4.1.19}$$

Its LT is given by

$$L_3(s) = (1 + s)^{-\alpha} \exp\{\lambda[(1 + s)^{-1} - 1]\} \tag{2.4.1.20}$$

Hence the pgf of its mixture of the Poisson is

$$G_3(z) = (2 - z)^{-\alpha} \exp\{\lambda[(2 - z)^{-1} - 1]\} \tag{2.4.1.21}$$

$$= h_\gamma(z)\, H(h_1(z)) \tag{2.4.1.22}$$

where

$H(z) = e^{\lambda(z-1)}$, the Poisson

$h_1(z) = (2 - z)^{-1}$, a geometric distribution

$h_\gamma(z) = (2 - z)^{-\gamma}$, a negative binomial

Thelander (1979) has discussed a distribution with the density

$$f_4(x) = \left[\Gamma(\lambda + 1)/\left\{e^{\lambda}\lambda^{\frac{\lambda}{2}}\right\}\right] x^{-\frac{\lambda}{2}} I_\lambda(2\sqrt{\lambda x}) \tag{2.4.1.23}$$

which is the Poisson-mixture of a gamma, given by

$$\text{Gamma}(x;\, u,\, 1) \underset{u}{\wedge} \text{Poisson}(u;\, \lambda) \tag{2.4.1.24}$$

and has the Laplace transform

$$L_4(s) = \exp\{\lambda[(1 + s)^{-1} - 1]\} \tag{2.4.1.25}$$

The pgf of the $f_4(x)$-mixture of the Poisson is

$$G_4(z) = \exp\{\lambda[(2 - z)^{-1} - 1]\} \qquad (2.4.1.26)$$

which is a Poisson-geometric (i.e. a Polya-Aeppli). This relationship of $f_4(x)$ and $G_4(z)$ indicates that, in a cyclic manner, many, and possibly all, generalized Poisson can be construed as mixtures of the Poisson distribution with continuous mixers, these continuous mixing distributions being Poisson-mixtures. Note that the Polya-Aeppli is usually given as the typical example for true contagion, but $G_4(\ )$ is a Polya-Aeppli which is a mixture of the Poisson and hence can also be viewed as an apparently contagious distribution.

Two other Bessel-type densities are given in Johnson and Kotz (1970, pp.28-30). They are the three-parameter McKay's distributions. Here we consider only two-parameter special cases. A two-parameter form of one of the densities is

$$f_5(x) = \left\{\left[\frac{(\alpha + 1)^2 - 1}{2(\alpha + 1)^2}\right]^{\lambda} \sqrt{\frac{2}{\pi}} \Big/ \Gamma(\lambda)\right\} \cdot x^{\lambda - \frac{1}{2}} e^{-x/(\alpha+1)} K_{\lambda - \frac{1}{2}}(x)$$

$$x > 0,\ \alpha > 0,\ \lambda > 0, \qquad (2.4.1.27)$$

with the moment generating function

$$M_5(t) = \{[(\alpha + 1)^2 - 1]/[(\alpha + 1)^2 - (1 - (\alpha + 1)t)^2]\}^{+\lambda} \qquad (2.4.1.28)$$

and an associated mixture of the Poisson having the pgf

$$G_5(z) = \{[(\alpha + 1)^2 - 1]/[(\alpha + 1)^2 - ((\alpha + 2) - (\alpha + 1)z)^2]\}^{\lambda} \qquad (2.4.1.29)$$

A similar two-parameter special case of the other McKay's distribution has the density

$$f_6(x) = \left\{\left[\left(\frac{1}{2}\right)((\alpha + 1)^2 - 1)\right]^{\lambda} \sqrt{\frac{2}{\pi}} \Big/ \Gamma(\lambda)\right\} \cdot x^{\lambda - \frac{1}{2}} e^{-(\alpha+1)x} I_{\lambda - \frac{1}{2}}(x) \qquad (2.4.1.30)$$

with the moment generating function

$$M_6(t) = \{[(\alpha + 1)^2 - 1)]/[((\alpha + 1)^2 - t)^2 - 1]\}^{\lambda}. \qquad (2.4.1.31)$$

Hence the pgf of its mixture of the Poisson is

$$G_6(z) = \{[(\alpha + 1)^2 - 1]/[((\alpha + 1)^2 + 1) - z)^2 - 1]\}^{\lambda} \qquad (2.4.1.32)$$

2.4.2 Infinitely divisible gamma mixtures

Steutel (1970, p.25) has shown that all mixtures, on β, of the gamma distribution

$$f(x) = \frac{\beta^{\theta}}{\Gamma(\theta)} x^{\theta-1} e^{-\beta x}, \quad x > 0,\ 0 < \theta \leq 1,\ \beta > 0 \qquad (2.4.2.1)$$

are infinitely divisible. The case $\theta = 1$ is the exponential distribution. Steutel (1967) (also Goldie (1967)) separately proves that all mixtures of the exponential distribution are infinitely divisible.

The Laplace transform (LT) of $f(x)$ is

$$L_0(s) = \left(\frac{\beta}{\beta + s}\right)^{\theta} \qquad (2.4.2.2)$$

and the LT of its mixture $g(x)$ arising from a mixing distribution $v(x)$, $x \in A \subseteq R_+$ is given by

$$L_1(s) = \int_A \left(\frac{\beta}{\beta + s}\right)^{\theta} v(\beta)d\beta. \qquad (2.4.2.3)$$

It is clear that the pgf of the $g(x)$-mixture of the Poisson is given by

$$G(z) = \int_A \left(\frac{\beta}{\beta + 1 - z}\right)^{\theta} v(\beta)d\beta \qquad (2.4.2.4)$$

which obviously is the pgf of the mixture, on β, of the negative binomial

$$h(z) = \left(1 + \frac{1}{\beta} - \frac{z}{\beta}\right)^{-\theta}, \quad \beta > 0,\ 0 < \theta \leq 1 \qquad (2.4.2.5)$$

with $v(x)$ as the mixing distribution. This implies that all mixtures, on β, of this negative binomial are inf div i.e. they are GPDs. The case $\theta = 1$ in $h(z)$ is the geometric distribution. Steutel (1967, and 1970, pp.60-61, 78) separately proves that all mixtures of the geometric distribution

$$h_1(z) = (1 - \theta_1)/(1 - \theta_1 z), \quad 0 < \theta_1 < 1 \qquad (2.4.2.6)$$

are inf div. These are discussed further in section 2.5.

The following is an example of inf div mixtures of the gamma distribution. A detailed discussion of the properties of this distribution is not of interest here.

Mixing, on u, the gamma distribution

$$f(x) = \frac{u^{\gamma}}{\Gamma(\gamma)} x^{\gamma-1} e^{-ux}, \qquad x > 0,\ u > 0,\ \gamma > 0 \qquad (2.4.2.7)$$

with the general Beta-I distribution

$$v(u) = \lambda^{-(\alpha+\beta-1)} \frac{1}{B(\alpha, \beta)} u^{\alpha-1} (\lambda - u)^{\beta-1}, \qquad 0 < u < \lambda \qquad (2.4.2.8)$$

gives the distribution with the density

$$f(x) = \lambda^{\gamma} \frac{\Gamma(\alpha + \beta)}{B(\alpha, \gamma)\, \Gamma(\alpha + \beta + \gamma)} \cdot x^{\gamma-1} {}_1F_1(\alpha + \gamma; \alpha + \beta + \gamma; -\lambda x)$$

$$= Cx^{\gamma-1} e^{-\lambda x} {}_1F_1(\beta; \alpha + \beta + \gamma; \lambda x) \qquad (2.4.2.9)$$

where $C = \lambda^{\gamma}/[\Gamma(\gamma)\ F(\beta, \gamma; \alpha + \beta + \gamma; 1)]$.

The Laplace transform of f(x) is obtained as

$$L(s) = C\int_0^{\infty} e^{-(\lambda+s)} x^{\gamma-1} {}_1F_1(\beta; \alpha + \beta + \gamma; \lambda x)dx$$

$$= C \frac{\Gamma(\gamma)}{(\lambda + s)^{\gamma}} {}_2F_1\left(\gamma, \beta; \alpha + \beta + \gamma; \frac{\lambda}{\lambda + s}\right)$$

$$= \left(\frac{\lambda}{\lambda + s}\right)^{\gamma} \frac{{}_2F_1\left(\gamma, \beta; \alpha + \beta + \gamma; \frac{\lambda}{\lambda + s}\right)}{{}_2F_1(\gamma, \beta; \alpha + \beta + \gamma; 1)}, \quad s > 0 \qquad (2.4.2.10)$$

From this the pgf of the f(x)-mixture of the Poisson is obtained as

$$G(z) = \left(\frac{\lambda}{\lambda + 1 - z}\right)^{\gamma} {}_2F_1\left(\gamma, \beta; \alpha + \beta + \gamma; \frac{\lambda}{\lambda + 1 - z}\right) / {}_2F_1(\gamma, \beta; \alpha + \beta + \gamma; 1)$$

$$= h_{\gamma}(z)\ H(h_1(z)) \qquad (2.4.2.11)$$

where

$$H(z) = {}_2F_1(\beta, \gamma; \alpha + \beta + \gamma; z) / {}_2F_1(\beta, \gamma; \alpha + \beta + \gamma; 1) \qquad (2.4.2.12)$$

i.e. the generalized Waring distribution,

$$h_{\gamma}(z) = \left(1 + \frac{1}{\lambda} - \frac{z}{\lambda}\right)^{-\gamma}, \qquad (2.4.2.13)$$

a negative binomial, and

$$h_1(z) = \left(1 + \frac{1}{\lambda} - \frac{z}{\lambda}\right)^{-1}, \qquad (2.4.2.14)$$

a geometric distribution.

pp.89, 176)). The infinite divisibility of f(x) has been shown by Grosswald (1976) and Bondesson (1978). Its rth moment is given by

$$\mu_r' = \beta^r \Gamma(\alpha - r)/\Gamma(\alpha). \qquad (2.4.3.3)$$

By the above result of Steutel, G(z) is a GPD at least for the cases $0 < \gamma \leq 1$. The infinite divisibility of G(z) for $0 < \gamma \leq 1$ can be shown in another way. Note that $h_1(z)$ and $h_\gamma(z)$ are GPDs. This mixture of the Poisson will be a GPD for the cases of H(z) which are GPDs. When $0 < \gamma \leq 1$, H(z) is logconvex (see section 2.5) and is therefore inf div. Hence G(z) is a GPD at least for the special cases $0 < \gamma \leq 1$.

The probability function of G(z) is given by

$$P_x = \frac{1}{\Gamma(\gamma)} \left(\frac{\lambda}{\lambda+1}\right)^{\gamma} \left(\frac{1}{\lambda+1}\right)^{x} \frac{\Gamma(x+\gamma)}{x!} \frac{{}_2F_1\left(x+\gamma, \beta; \alpha+\beta+\gamma; \frac{\lambda}{\lambda+1}\right)}{{}_2F_1(\gamma, \beta; \alpha+\beta+\gamma; 1)}$$

$$= \frac{\theta^{\gamma}}{\Gamma(\gamma)} (1-\theta)^{x} \frac{\Gamma(x+\gamma)}{\Gamma(x+1)} \frac{{}_2F_1(x+\gamma, \beta; \alpha+\beta+\gamma; \theta)}{{}_2F_1(\gamma, \beta; \alpha+\beta+\gamma; 1)} \qquad (2.4.2.15)$$

with $\theta = \lambda/(\lambda+1)$, $0 < \theta < 1$, a representation which, while in itself is not quite useful for computation, can be a basis for the derivation of a recursive formula.

2.4.3 The reciprocal-gamma mixture of the Poisson distribution

Here we present a compound Poisson distribution whose derivation is simple but which does not seem to have been discussed before.

Given that a random variable u follows the gamma distribution having the density

$$g(u) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} u^{\alpha-1} e^{-\beta u}, \quad u > 0, \alpha > 0, \beta > 0 \qquad (2.4.3.1)$$

then the density of the distribution of $x = \frac{1}{u}$ is given by

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{1}{x^{\alpha+1}} e^{-\beta/x}, \quad x > 0, \alpha > 0, \beta > 0 \qquad (2.4.3.2)$$

The distribution defined by f(x) is the reciprocal gamma distribution, or Pearson's Type V distribution (see e.g. Kendall and Stuart (1969, pp.89, 176)). The infinite divisibility of f(x) has been shown by Grosswald (1976) and Bondesson (1978). Its rth moment is given by

$$\mu_r' = \beta^r \Gamma(\alpha - r)/\Gamma(\alpha). \qquad (2.4.3.3)$$

The mixture of the Poisson with the reciprocal gamma as the mixing distribution has the probability function, P_x, obtained as follows.

$$P_x = \int_0^\infty \frac{e^{-u} u^x}{x!} f(u)du$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x!} \int_0^\infty u^{(x-\alpha)-1} e^{-(u+\beta/u)} du$$

$$= \frac{2\beta^{\alpha/2}}{\Gamma(\alpha)} \frac{\beta^{x/2}}{x!} K_{x-\alpha}(2\sqrt{\beta}), \qquad \alpha > 0, \beta > 0 \qquad (2.4.3.4)$$

$$= \frac{2b^a}{\Gamma(a)} \frac{b^x}{x!} K_{x-a}(2b), \quad a \equiv \alpha > 0, b \equiv \sqrt{\beta} > 0, x = 0, 1, \ldots \qquad (2.4.3.5)$$

It is well known that $K_\lambda(y)$ denoting the second-kind modified Bessel function is finite and positive for $y > 0$ and $-\infty < \lambda < \infty$. The definitional theorem for a mixture (section 2.1) implies that if the integral above is finite then the resulting P_x is a valid probability function.

Since the mixing distribution f(x) is inf div, so is its mixture of the Poisson distribution. In other words, the reciprocal-gamma mixture of the Poisson is a generalized Poisson distribution.

Computation of the probabilities is eased by use of the recursive formula:

$$P_{x+2} = \frac{x + 1 - a}{x + 2} P_{x+1} + \frac{b^2}{(x + 1)(x + 2)} P_x, \quad x = 0, 1, \ldots \qquad (2.4.3.6)$$

with $P_0 = 2b^a K_{-a}(2b)/\Gamma(a)$

and $P_1 = P_0 . bK_{1-a}(2b)/K_{-a}(2b)$

The rth factorial moment of the distribution is given by

$$\mu_{(r)} = b^{2r} \Gamma(a - r)/\Gamma(a) \qquad (2.4.3.7)$$

which is finite for all non-integral a, and all integral $a > r$. In particular the first two factorial moments are

$$\mu_{(1)} = b^2/(a - 1) \qquad (2.4.3.8)$$

and

$$\mu_{(2)} = \frac{b}{(a - 1)(a - 2)} \qquad (2.4.3.9)$$

from which the mean and variance are obtained respectively as

$$\mu = b^2/(a - 1), \qquad a > 1 \qquad (2.4.3.10)$$

$$\mu_2 = \mu\left(1 + \frac{\mu}{a - 2}\right) \qquad (2.4.3.11)$$

Note that some other important discrete distributions (e.g. the generalized Waring distribution) also have this feature of non-existence of certain moments for some parameter values.

The moments estimators of a and b respectively are

$$\hat{a} = 2 + \bar{x}^2/(s^2 - \bar{x}) \qquad (2.4.3.12)$$

and

$$\hat{b} = [\bar{x}\{1 + \bar{x}^2/(s^2 - \bar{x})\}]^{\frac{1}{2}} \qquad (2.4.3.13)$$

Since all non-trivial GPDs and all mixtures of the Poisson are overdispersed i.e. are such that in all cases $\sigma^2 > \mu$ (see section 2.3.2), presumably they would not be fitted to any data for which $s^2 \leqslant \bar{x}$. This implies that the estimators $\hat{a}$ and $\hat{b}$ are applicable in all the situations in which the fitting of the Poisson Λ reciprocal gamma distribution is particularly appropriate.

Note that this distribution is distinct from the generalized-inverse-normal mixture of the Poisson. Sichel (1973) remarks that the generalized inverse normal is intermediate to Pearson's Type III and Type V and tends to either only in the limit. The reciprocal gamma is a Type V. The mixture given here is simpler than Sichel's generalized-inverse-normal mixture of the Poisson, since it has only two parameters in contrast with Sichel's three. However there is an interesting relationship between the two distributions.

Steutel (1970, p.88) observes that if f(x) is an inf div density, then the function

$$f_{*}(x) = e^{cx} f(x) / \int e^{cx} f(x)dx \qquad (2.4.3.14)$$

is also an inf div density. This is the relationship between the density f(x) of the reciprocal gamma and the special case $\gamma = -a$, $a > 0$ of $f_{*}(x)$ the generalized inverse normal since

$$f_{*}(x) = Ce^{-\alpha x} \cdot x^{\gamma-1} e^{-\beta/x}, \qquad -\infty < \gamma < \infty \qquad (2.4.3.15)$$

The same author observed further that if P_x is the probability function of an inf div lattice distribution, then another inf div lattice distribution is defined by

$$P^{*}_{x} = \theta^{x} P_x / \Sigma \theta^{x} P_x, \qquad 0 < \theta < 1 \qquad (2.4.3.16)$$

This is the relationship between the distribution defined above and the subsystem $\gamma = -a$, $a > 0$ of Sichel's; P_x corresponding to the probability function of the former and P^{*}_{x} to that of the latter. Specifically,

$$P_x = 2b^{a}/\Gamma(a) \; \frac{b^{x}}{x!} K_{x-a}(2b), \qquad a > 0, b > 0 \qquad (2.4.3.17)$$

and

$$P^{*}_{x} = \left[(1-\theta)^{\frac{a}{2}} / K_{-a}(2b\sqrt{1-\theta})\right] \cdot \frac{(b\theta)^{x}}{x!} K_{x-a}(2b), \qquad (2.4.3.18)$$

$$a > 0, \; b > 0, \; 0 < \theta < 1.$$

The special case $\gamma = -\frac{1}{2}$ of the generalized inverse normal is the well-known inverse normal distribution (e.g. Tweedie (1957)) and, as mentioned earlier, its mixture of the Poisson has generated a corpus of literature. It is observed that the corresponding special case $a = \frac{1}{2}$ of the reciprocal gamma is also an important continuous distribution. It does not appear to have been given a specific name. We shall refer to it as Feller's stable distribution (Feller (1971, pp.173, 432, 436-437, 463)). A discussion of the mixture of the Poisson obtained with this distribution follows.

The density of Feller's stable distribution is given as

$$f(x) = \left[\frac{\alpha}{2}(\pi x^{3})^{-\frac{1}{2}}\right] \exp(-\alpha^{2}/4x), \quad x > 0, \; \alpha > 0 \qquad (2.4.3.19)$$

and using the relation

$$K_{-\frac{1}{2}}(u) = K_{\frac{1}{2}}(u) = \left(\frac{\pi}{2u}\right)^{\frac{1}{2}} e^{-u}, \qquad (2.4.3.20)$$

its Laplace transform is given by

$$L(s) = \exp(-\alpha\sqrt{s}) \qquad (2.4.3.21)$$

from which the pgf of its mixture of the Poisson is obtained as

$$G(z) = \exp\{-\alpha(1 - z)^{\frac{1}{2}}\} \qquad (2.4.3.22)$$

i.e.

$$G(z) = \exp\{\alpha([1 - (1 - z)^{\frac{1}{2}}] - 1)\} \qquad (2.4.3.23)$$

which is a generalized Poisson distribution if

$$h(z) = 1 - (1 - z)^{\frac{1}{2}} \qquad (2.4.3.24)$$

is a valid pgf.

Kemp and Kemp (1968) give the pgf of the lost-games distribution as

$$G(s) = (2q)^{-a}[1 - (1 - 4pqs)^{\frac{1}{2}}]^a \qquad (2.4.3.25)$$

$$= (ps)^a {}_2F_1(\tfrac{1}{2}a, \tfrac{1}{2}(a + 1); a + 1; 4pqs) \qquad (2.4.3.26)$$

This provides us with the relation

$$[1 - (1 - 4pqs)^{\frac{1}{2}}]^a = (\tfrac{1}{2})^a (4pqs)^a {}_2F_1(\tfrac{1}{2}a, \tfrac{1}{2}(a + 1); a + 1; 4pqs) \qquad (2.4.3.27)$$

On making the substitution $z = 4pqs$, we have

$$[1 - (1 - z)^{\frac{1}{2}}]^a = (\tfrac{1}{2})^a z^a {}_2F_1(\tfrac{1}{2}a, \tfrac{1}{2}a + \tfrac{1}{2}; a + 1; z) \qquad (2.4.3.28)$$

Putting $a = 1$ gives

$$1 - (1 - z)^{\frac{1}{2}} = (\tfrac{1}{2})z {}_2F_1(\tfrac{1}{2}, 1; 2; z)$$

$$= z {}_2F_1(\tfrac{1}{2}, 1; 2; z)/{}_2F_1(\tfrac{1}{2}, 1; 2; 1) \qquad (2.4.3.29)$$

which is the pgf of a shifted Waring distribution. Hence

$$h(z) = 1 - (1 - z)^{\frac{1}{2}} \qquad (2.4.3.30)$$

is not only a pgf; it is in fact the pgf of an important discrete distribution. This shows that like the Poisson-inverse normal (PIN) distribution, the mixture of the Poisson with Feller's distribution is

an important generalized Poisson distribution. This distribution has no moments. An estimator of α is $-\{\log(2\Sigma f_{2x}/N - 1)\}/\sqrt{2}$. (See section 3.2.4.)

2.4.4 The f(x)-mixture of the Poisson where $x \geq a > 0$

If f(x) is a pdf on $(0,\infty)$, the pdf $h(x) = f(x - \beta)$, $x \geq \beta > 0$ can be regarded as the shifted form of f(x). An example is the generalized gamma distribution

$$h(x) = [\gamma/(\beta^{\alpha\gamma}\Gamma(\alpha))](x - \lambda)^{\alpha\gamma-1}\exp\left[-\left(\frac{x - \lambda}{\beta}\right)^{\gamma}\right], \quad x \geq \lambda > 0 \qquad (2.4.4.1)$$

(Johnson and Kotz (1970, p.197)) which is the shifted form of Stacy's (1962)

$$f(x) = [\gamma/(\beta^{\alpha\gamma}\Gamma(\alpha))]x^{\alpha\gamma-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\gamma}\right], \quad x > 0. \qquad (2.4.4.2)$$

It is noted that if f(x) is inf div, so is h(x) (see e.g. Steutel (1970, p.42)). Infinitely divisible random variables bounded below have been discussed in detail by Baxter and Shapiro (1960) (see also Feller (1971, Chapter XVII) and Tucker (1961)).

Let $L_f(s)$ and $L_h(s)$ denote the Laplace transform of f(x) and the corresponding definite integral of h(x) respectively. The relationship between $L_f(s)$ and $L_h(s)$ is given by

$$L_f(s) = \int_0^{\infty} e^{-sx} f(x)dx \qquad (2.4.4.3)$$

$$= \int_{\beta}^{\infty} e^{-s(u-\beta)} f(u - \beta)du \qquad (2.4.4.4)$$

$$= e^{\beta s}\int_{\beta}^{\infty} e^{-su} h(x)du \qquad (2.4.4.5)$$

$$= e^{\beta s} L_h(s) \qquad (2.4.4.6)$$

i.e.
$$L_h(s) = e^{-\beta s} L_f(s) \qquad (2.4.4.7)$$

From this relation it is clear that the h(x)-mixture of the Poisson is the convolution of a Poisson distribution and the f(x)-mixture of the Poisson, provided that the latter exists.

If f(x) is inf div, so is its mixture of the Poisson and hence so is the h(x)-mixture of the Poisson its shifted form. For example, if h(x) is the three-parameter gamma distribution (Johnson and Kotz (1970, p.166))

$$h(x) = [\beta^{-\alpha}/\Gamma(\alpha)](x - \lambda)^{\alpha-1}\exp[-(x - \lambda)/\beta], \quad \alpha > 0, \beta > 0, x \geq \lambda > 0 \tag{2.4.4.8}$$

which is the shifted form of the standard gamma distribution, then the h(x)-mixture of the Poisson is the convolution of a Poisson distribution and a negative binomial distribution. This mixture has the pgf

$$G(z) = e^{\lambda(z-1)} \cdot (1 + \beta - \beta z)^{-\alpha} \tag{2.4.4.9}$$

This is the distribution discussed by Lüders (1934). It is a generalized Poisson distribution having the cluster-size distribution

$$h(z) = \theta_2 z + (1 - \theta_2)\log(1 - \theta_1 z)/\log(1 - \theta_1) \tag{2.4.4.10}$$

$0 < \theta_1 < 1$, $0 < \theta_2 < 1$ (see section 2.2.4).

2.5 Other Classes of GPDs

2.5.1 Mixtures of the geometric distribution

All generalized geometric distributions are GPDs, since generalizing a GPD gives another GPD. Steutel (1967, and 1970, pp.60-61, 78) has given an important result which states that all mixtures of (and all mixtures, on θ, of all distributions resulting from generalizing) the geometric distribution, G_θ, with the pgf

$$h(z) = (1 - \theta)/(1 - \theta z), \qquad 0 < \theta < 1 \tag{2.5.1.1}$$

are infinitely divisible. This means that it is not only all generalized G_θ-distributions but also all compound G_θ-distributions that are infinitely divisible. Hence this parametric form of the geometric distribution is

very important in a discussion of generalized Poisson distributions.

Harn (1978, p.20) has observed that a lattice distribution on non-negative integers is completely monotone if and only if it is a mixture of the parametric form G_θ of the geometric distribution. This system of lattice distributions is the main focus of Harn (1978). The following theorems about mixtures of this geometric distribution are due to Harn.

Theorem 4(a)

A lattice distribution, G(z), on x = 0, 1,... is a mixture of the parametric form G_θ of the geometric distribution if and only if

$$H(z) = \frac{2G(z) - 1}{2G(z)} \qquad (2.5.1.2)$$

is a pgf.

Theorem 4(b)

If H(z) is any pgf, then $G(z) = h_1(z) \cdot h_3(h_2(z)H(z))$ is a mixture of the geometric distribution G_θ, where

$$h_i(z) = (1 - \theta_i)/(1 - \theta_i z), \qquad (2.5.1.3)$$

$0 < \theta_i \leq 1$, i = 1, 2, 3, $\theta_1 \leq \theta_2$.

Theorem 4(c)(i)

If G(z) is the pgf of a mixture of the geometric distribution G_θ, then so are

$$H_1(z) = \frac{\alpha}{\alpha - \log G(z)}, \qquad \alpha > 0 \qquad (2.5.1.4)$$

$$H_2(z) = \frac{G(z)}{\alpha + (1 - \alpha)\, G(z)}, \qquad \alpha > 0 \qquad (2.5.1.5)$$

and

$$H_3(z) = \frac{\theta}{\theta + (1 - \theta)\, G(z)}, \qquad 0 < \theta \leq 1. \qquad (2.5.1.6)$$

Theorem 4(c)(ii)

If $G_1(z)$ and $G_2(z)$ are mixtures of the geometric distribution G_θ then so is

$$H(z) = \frac{G_1(z)\, G_2(z)}{G_1(z) + G_2(z) - G_1(z)\, G_2(z)}. \qquad (2.5.1.7)$$

An example of the mixtures of the geometric distribution G_θ is that with the standard Beta-I distribution as the mixing distribution. This mixture is usually called the beta-geometric (e.g. Chatfield (1970)) or the Waring distribution (Irwin (1965)). Its pgf is given by

$$G(z) = \int_0^1 \left(\frac{1-\theta}{1-\theta z}\right) \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \qquad (2.5.1.8)$$

$$= \frac{{}_2F_1(1, \alpha; \alpha+\beta+1; z)}{{}_2F_1(1, \alpha; \alpha+\beta+1; 1)} \qquad (2.5.1.9)$$

(see e.g. Kemp and Kemp (1956a)). Note that this distribution is logconvex (section 2.5.2) and hence satisfies another sufficiency condition for infinite divisibility.

2.5.2 Logconvex lattice distributions

A lattice distribution on $x = 0, 1, \ldots$ is said to be logconvex, if

$$P_x^2 \leq P_{x-1} P_{x+1}, \qquad x = 1, 2, \ldots \qquad (2.5.2.1)$$

i.e. if the ratios

$$R_x = \frac{P_{x+1}}{P_x}, \qquad x = 0, 1, 2, \ldots \qquad (2.5.2.2)$$

form a monotone increasing sequence. Warde and Katti (1971) have proved that logconvexity is a sufficient condition for infinite divisibility. This result is stated here as a theorem.

Theorem 5

A discrete distribution on $x = 0, 1,$ is infinitely divisible if it is logconvex.

That this condition is not a necessary condition for inf div is obvious with the Poisson distribution and the cases $\alpha > 1$ of the negative binomial

$$G(z) = (1 + \beta - \beta z)^{-\alpha} \qquad (2.5.2.3)$$

as counter-examples.

Harn (1978, p.21) also gives the following ordering of some subclasses of the class of generalized Poisson distributions.

Theorem 6

Suppose A denotes the set of all mixtures of the geometric distribution G_θ, B denotes the set of all logconvex lattice distribution, C the set of all generalized geometric distributions, and D the set of all generalized Poisson distributions. Then

$$A \subset B \subset C \subset D \qquad (2.5.2.4)$$

where all inclusions are strict.

The following are some examples of logconvex lattice distributions. With each is given the form of the ratio $R_x = P_{x+1}/P_x$ which determines logconvexity.

The special cases of the generalized hypergeometric distributions of the form

$$G_1(z) = \frac{{}_{m+1}F_m(\theta_1, (a_i); (a_i + b_i); z)}{{}_{m+1}F_m(\theta_1, (a_i); (a_i + b_i); 1)}, \quad 0 < \theta_1 \leq 1, \ \Sigma b_i > \theta_1 \qquad (2.5.2.5)$$

are logconvex. For these the ratio R is given by

$$R_x = \left(\frac{x + \theta_1}{x + 1}\right)\left(\frac{x + a_1}{x + a_1 + b_1}\right) \cdots \left(\frac{x + a_m}{x + a_m + b_m}\right), \qquad (2.5.2.6)$$

These include the Waring distribution with

$$G(z) = C_2 F_1(1, \alpha; \alpha + \beta + 1; z) \qquad (2.5.2.7)$$

and Kemp's (1980) convertible distributions

$$G(z) = C_3 F_2\left(\frac{1}{2}, \frac{1}{2}, 2; \frac{5}{2}, \frac{5}{2}; z\right) \qquad (2.5.2.8)$$

and

$$G(z) = C_4 F_3\left(1, 1, 1, \frac{5}{2}; \frac{3}{2}, 3, 3; z\right). \qquad (2.5.2.9)$$

Another set of logconvex special cases of the generalized hypergeometric distributions are those of the form

$$G_2(z) = \frac{{}_{m+1}F_m(\theta_1, (a_i); (a_i + b_i); \theta_2 z)}{{}_{m+1}F_m(\theta_1, (a_i); (a_i + b_i); \theta_2)}, \quad 0 < \theta_1 < 1, \ 0 < \theta_2 < 1, \qquad (2.5.2.10)$$

with

$$R_x \simeq \theta_2 \left(\frac{x + \theta_1}{x + 1}\right)\left(\frac{x + a_1}{x + a_1 + b_1}\right) \cdots \left(\frac{x + a_m}{x + a_m + b_m}\right). \quad (2.5.2.11)$$

Note that the relationship between $G_1(z)$ and $G_2(z)$ is of the type mentioned in section 2.4.3 (Stentel (1970, p.88)). Distributions of the form $G_2(z)$ include the shifted logarithmic distribution

$$G(z) = C_2F_1(1, 1, 2; z), \quad 0 < \theta < 1 \quad (2.5.2.12)$$

and the lost-games shifted-to-zero

$$G(z) = C_2F_1(\tfrac{1}{2}, 1; 2; 4(1 - \theta)z), \quad 0 < \theta < 1 \quad (2.5.2.13)$$

Note that if $G(z) = C_pF_q((\alpha_i); (\beta_j); z)$ is a logconvex pgf, so is $G(z) = C_pF_q((\alpha_i); (\beta_j); \theta z)$, $0 < \theta < 1$.

2.5.3 The geometric with-added-zeroes and related GPDs

The inf div lattice distributions discussed in this section and the next are all convolutions of pseudo-binomials. A function $\omega(z)$ is called a pseudo-binomial if it is in the form

$$\omega(z) = (q + pz)^\gamma, \quad \gamma \text{ non-integral}, \quad 0 < p < 1, \; q = 1 - p$$

or $$\omega(z) = (\beta + 1 - \beta z)^\alpha, \quad \alpha > 0, \; \beta > 0.$$

311117

Convolutions of pseudo-binomials have been discussed by Kemp (1979) who gives conditions under which they can be pgfs and presents examples, some of which are infinitely divisible e.g. the geometric with-added-zeroes and Bhattacharya's mixture of the Poisson. Some examples of pseudo-binomials have also been given by Feller (1968, pp.290-291). A pgf can, in general, be represented (in several ways) as a product of two purely mathematical functions. The objective here is to present some infinitely divisible pseudo-binomial convolutions as products of clearly recognizable inf div pgfs.

The expressions "with-added-zeroes" and "with-decreased-zeroes" are due to Johnson and Kotz (1969, pp.204-206) who give to a distribution with pgf of the form

$$G(z) = \theta + (1 - \theta)\, h(z), \quad |\theta| < 1 \qquad (2.5.3.1)$$

the name "h(z)-with-decreased-zeroes" if θ is negative and "h(z)-with-added-zeroes" if θ is positive.

The pgfs of two geometric distributions G_{θ_1}, G_{θ_2} with $0 < \theta_2 < \theta_1 < 1$ can be represented as

$$h_1(z) = (1 - \theta_1)/(1 - \theta_1 z) \qquad (2.5.3.2)$$

$$= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}, \quad \theta_1 = 1/(\alpha + 1),\ \alpha > 0 \qquad (2.5.3.3)$$

$$h_2(z) = (1 - \theta_2)/(1 - \theta_2 z) \qquad (2.5.3.4)$$

$$= \left(1 + \frac{1}{\alpha + \beta} - \frac{z}{\alpha + \beta}\right)^{-1}, \quad \theta_2 = 1/(\alpha + \beta + 1),\ \alpha > 0,\ \beta > 0. \qquad (2.5.3.5)$$

Their quotient is

$$G(z) = h_1(z)/h_2(z)$$

$$= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1} \Big/ \left(1 + \frac{1}{\alpha + \beta} - \frac{z}{\alpha + \beta}\right)^{-1}$$

$$= \left(1 + \frac{1}{\alpha + \beta} - \frac{z}{\alpha + \beta}\right) \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right) \qquad (2.5.3.6)$$

(a convolution of pseudo-binomials)

$$= \theta + (1 - \theta)\, h_1(z)$$

where $\theta = \alpha/(\alpha + \beta)$, $\alpha > 0$, $\beta > 0$. Hence G(z) is the pgf of the geometric distribution with-added-zeroes. Note that any geometric distribution can be put in the form $h_1(z)$.

Considering $h_1(z)$ in the form G_{θ_1} i.e. $h_1(z) = (1 - \theta_1)/(1 - \theta_1 z)$, it is obvious that G(z) is a mixture of $h_1(z)$ where the mixing distribution is the binomial with unit exponent parameter. Therefore since all mixtures of G_{θ_1} are infinitely divisible, G(z) is inf divisible. The infinite divisibility of G(z) can be shown in another way. The Levy-Feller form of G(z) is

KASHIM IBRAHIM LIBRARY
AHMADU BELLO UNIVERSITY

$$G(z) \simeq \exp\{\lambda[u(z) - 1]\} \qquad (2.5.3.7)$$

where $$\lambda \simeq -\log\left[1 - \frac{\beta}{\beta + \alpha(\alpha + \beta + 1)}\right] > 0 \qquad (2.5.3.8)$$

and $$u(z) = \sum_{x=1}^{\infty} p_x^* z^x \qquad (2.5.3.9)$$

with $$0 < p_x^* = \left(\frac{1}{x}\right)\left(\frac{1}{\alpha + 1}\right)^x \left[1 - \left(\frac{\alpha + 1}{\alpha + \beta + 1}\right)^x\right] < 1.$$

Since u(z) is a pgf and $\lambda > 0$, G(z) is a generalized Poisson distribution. In other words it is infinitely divisible.

The function

$$u(x) = x^{b-1}{}_1F_1(a; b; \beta x), \quad x > 0, b > 0 \qquad (2.5.3.10)$$

has the Laplace transform

$$L(s) = \frac{\Gamma(b)}{s^b}\left(\frac{s}{s - \beta}\right)^a, \qquad s > 0. \qquad (2.5.3.11)$$

This provides a pdf

$$f(x) = \frac{\gamma^b}{\Gamma(b)}\left(\frac{\gamma - \beta}{\gamma}\right)^a e^{-\gamma x} x^{b-1}{}_1F_1(a, b; \beta x), \quad x > 0, \gamma > \beta > 0$$

$$= \frac{1}{\Gamma(b)}\alpha^a(\alpha + \beta)^{b-a} e^{-(\alpha+\beta)x} x^{b-1}{}_1F_1(a; b; \beta x), \; x > 0, \alpha = \gamma - \beta > 0, \; \beta > 0 \qquad (2.5.3.12)$$

having the Laplace transform

$$L(s) = \frac{\alpha^a}{(\alpha + \beta)^{a-b}} \cdot \frac{(s + \alpha + \beta)^{a-b}}{(s + \alpha)^a} \qquad (2.5.3.13)$$

The f(x)-mixture of the Poisson has the pgf

$$G(z) = \frac{\alpha^a}{(\alpha + \beta)^{a-b}} \cdot \frac{(\alpha + \beta + 1 - z)^{a-b}}{(\alpha + 1 - z)^a}$$

$$= \left(1 + \frac{1}{\alpha + \beta} - \frac{z}{\alpha + \beta}\right)^{a-b} \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^a \qquad (2.5.3.14)$$

in its representation as a convolution of pseudo-binomials. The special case $\beta = 1$ of f(x) is Bhattacharya's (1966) continuous distribution and G(z), the f(x)-mixture of the Poisson derived by the same author (see section 2.3.1). The distributions f(x) and G(z) will therefore

be referred to as the extended versions of Bhattacharya's distributions.

We observe that

$$G(z) = \left(1 + \frac{1}{\alpha+\beta} - \frac{z}{\alpha+\beta}\right)^{a-b} \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{a}$$

$$= \left(1 + \frac{1}{\alpha+\beta} - \frac{z}{\alpha+\beta}\right)^{-b} \left[\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1} \Big/ \left(1 + \frac{1}{\alpha+\beta} + \frac{z}{\alpha+\beta}\right)^{-1}\right]^{a} \quad (2.5.3.-5)$$

$$= \left(1 + \frac{1}{\alpha+1} - \frac{z}{\alpha+1}\right)^{-b} \left[\theta + (1-\theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}\right]^{a}$$

$$= H_1(z)\, H_2(z) \quad (2.5.3.-6)$$

where

$$H_1(z) = \left(1 + \frac{1}{\alpha+1} - \frac{z}{\alpha+1}\right)^{-b} \quad (2.5.3.-7)$$

a negative binomial distribution, and

$$H_2(z) = \left[\theta + (1-\theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}\right]^{a}$$

$$= [u(z)]^{a} \quad (2.5.3.-8)$$

with u(z) being a geometric distribution with-added-zeroes, which is infinitely divisible. This representation exhibits clearly the infinite divisibility of G(z) by showing it as a convolution of two infinitely divisible distributions.

Note that if a > b we can have, alternatively,

$$G(z) = \left(1 + \frac{1}{\alpha+\beta} - \frac{z}{\alpha+\beta}\right)^{a-b} \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{a}$$

$$= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-b} \left[\left(1 + \frac{1}{\alpha+\beta} - \frac{z}{\alpha+\beta}\right) \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)\right]^{a-b}$$

$$= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-b} \left[\theta + (1-\theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}\right]^{(a-b)} = H_1(z)\, H_2(z) \quad (2.5.3.-9)$$

where, similarly, $H_1(z) = \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-b}$, a negative binomial

and $H_2(z) = [u(z)]^{(a-b)}$, $\quad a - b > 0$

Both $H_1(z)$ and $H_2(z)$ are again infinitely divisible.

The density

$$f(x) = \frac{\alpha^{3/2}}{(\alpha + \beta)\sqrt{\pi}} \frac{(1 + 2\beta x)e^{-\alpha x}}{\sqrt{x}}, \quad x > 0, \alpha > 0, \beta > 0 \qquad (2.5.3.20)$$

is the mixture of $\text{Gamma}\left(x; \frac{3}{2}, \alpha\right)$ and $\text{Gamma}\left(x; \frac{1}{2}, \alpha\right)$ and has the LT

$$L(s) = \left(\frac{\alpha + \beta + s}{\alpha + \beta}\right)\left(\frac{\alpha}{\alpha + s}\right)^{3/2} \qquad (2.5.3.21)$$

from which the pgf of its mixture of the Poisson is obtained as

$$G(z) = \left(1 + \frac{1}{\alpha + \beta} - \frac{z}{\alpha + \beta}\right) \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{3/2} \qquad (2.5.3.22)$$

a convolution of two pseudo-binomials. This distribution is actually the special case a = 3/2, b = 1/2 of the extended version of Bhattacharya's discussed above. Therefore following the procedure for the general form, in this case the latter alternative, the pgf G(z) in this special case can be represented as

$$\begin{aligned} G(z) &= \left(1 + \frac{1}{\alpha + \beta} - \frac{z}{\alpha + \beta}\right) \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{3/2} \\ &= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-\frac{1}{2}} \left[\theta + (1 - \theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}\right] \\ &= H_1(z)\, H_2(z) \qquad (2.5.3.23) \end{aligned}$$

where

$$H_1(z) = \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-\frac{1}{2}}, \text{ a negative binomial}$$

and $$H_2(z) = \theta + (1 - \theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}, \quad \theta = \alpha/(\alpha + \beta),$$

a geometric distribution with-added-zeroes.

Sankaran's (1970b) mixture of the Poisson by the Lindley distribution (i.e. by the mixture of Gamma(x; 2, α) and Gamma(x; 1, α)) with the density

$$f(x) = \frac{\alpha^2}{\alpha + 1}(x + 1)e^{-\alpha x} \qquad (2.5.3.24)$$

$$= \left(\frac{1}{\alpha + 1}\right)\left[\frac{\alpha^2}{\Gamma(2)} xe^{-\alpha x}\right] + \left(\frac{\alpha}{\alpha + 1}\right)\left[\alpha e^{-\alpha x}\right] \qquad (2.5.3.25)$$

has the pgf

$$G(z) = \left(1 + \frac{1}{\alpha + 1} - \frac{z}{\alpha + 1}\right) \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^2, \qquad (2.5.3.26)$$

another convolution of pseudo-binomials. It is the special case a = 2, b = 1 of Bhattacharya's. It is obvious that

$$G(z) = \left(1 + \frac{1}{\alpha + 1} - \frac{z}{\alpha + 1}\right) \Big/ \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^2$$

$$= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1} \left[\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1} \left(1 + \frac{1}{\alpha + 1} - \frac{z}{\alpha + 1}\right)^{-1}\right]$$

$$= \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1} \left[\theta + (1 - \theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}\right]$$

$$= H_1(z)\, H_2(z) \qquad (2.5.3.27)$$

where

$$H_1(z) = \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}, \text{ a geometric distribution}$$

and $$H_2(z) = \theta + (1 - \theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}, \quad \theta = \alpha/(\alpha + 1),$$

a geometric distribution with-added-zeroes.

2.5.4 The geometric with-decreased-zeroes and related GPDs

In his discussion on infinitely divisible pgfs, Feller (1968, pp.290-291) gives the following, among others, as an example:

$$G_*(z) = \sqrt{q - p}\left(\frac{q + pz}{q - pz}\right)^{1/2} \qquad (2.5.4.1)$$

where $0 < p < 1$, $q = 1 - p > p$. Considering $G_*(z)$ in the form

$$G_{*}(z) = \sqrt{q - p}\left(\frac{q + pz}{q - pz}\right)^{1/2}$$

$$= \left\{(q + pz)\left[\frac{1 - (p/q)}{1 - (p/q)z}\right]\right\}^{1/2}, \qquad 0 < \frac{p}{q} < 1$$

$$= (q + pz)^{1/2}\left[\frac{1 - (p/q)}{1 - (p/q)z}\right]^{1/2} \qquad (2.5.4.2)$$

we have a convolution of pseudo-binomials.

Here we show that

$$G(z) = (q + pz)\left[\frac{1 - (p/q)}{1 - (p/q)z}\right]$$

the base-component of $G_{*}(z)$ is in fact the geometric distribution with-decreased-zeroes, and also that $G(z)$ is infinitely divisible, which explains the inf divisibility of $G_{*}(z)$. Then we present another inf div convolution of pseudo-binomials, the convolution of $G(z)$ with a negative binomial distribution.

On putting $p = 1/(\alpha + 2)$, $q = 1 - p = (\alpha + 1)/(\alpha + 2)$ we have

$$G(z) = (q - p)\left(\frac{q + pz}{q - pz}\right)$$

$$= (q - p)[-1 + (2q)(q - pz)^{-1}]$$

$$= \left(\frac{-\alpha}{\alpha + 2}\right) + \left[1 - \left(\frac{-\alpha}{\alpha + 2}\right)\right]\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}$$

$$= \theta + (1 - \theta)\left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1} \qquad (2.5.4.3)$$

where $\theta = -\alpha/(\alpha + 2)$, $\alpha > 0$, $0 < |\theta| < 1$

i.e. $$G(z) = \theta + (1 - \theta)h_{1}(z) \qquad (2.5.4.4)$$

with $h_{1}(z) = \left(1 + \frac{1}{\alpha} - \frac{z}{\alpha}\right)^{-1}$ and $\theta = -\alpha/(\alpha + 2)$. Hence $G(z)$ is a geometric distribution with-decreased-zeroes.

The infinite divisibility of this distribution is more easily shown when $G(z)$ is put in the Levy-Feller form

$$G(z) = \exp\{\lambda[u(z) - 1]\} \qquad (2.5.4.5)$$

where
$$u(z) = (1/\lambda)\log\left\{\frac{1+\theta z}{1-\theta z}\right\}$$
$$= (2/\lambda)\sum_{x=0}^{\infty}(\theta z)^{2x+1}/(2x+1)$$
$$= \Sigma P_x z^x \qquad (2.5.4.6)$$

with
$$\theta = 1/(\alpha+1), \quad 0<\theta<1, \qquad (2.5.4.7)$$
$$\lambda = -\log\left\{\frac{1-\theta}{1+\theta}\right\} > 0, \qquad (2.5.4.8)$$

and
$$P_x = (2/\lambda)\theta^x/x, \quad x = 1, 3, 5, \ldots \qquad (2.5.4.9)$$

It is obvious that u(z) is a pgf. It can be regarded as the logseries on positive odd numbers. This shows that G(z) is a GPD. The infinite divisibility of G(z) implies that of $G_*(z)$.

Now consider the expression

$$(1+z)(1-z)^{-(a+1)} = {}_2F_1\left\{a, \frac{a}{2}+1; \frac{a}{2}; z\right\}, \quad |z|<1,\ a>0 \qquad (2.5.4.10)$$

(see e.g. Mangulis (1965, pp.11, 14)). This implies that

$$\left\{\frac{1+\theta z}{1-\theta z}\right\}\left\{\frac{1}{1-\theta z}\right\}^a = {}_2F_1\left\{a, \frac{a}{2}+1; \frac{a}{2}; \theta z\right\}, \quad 0<\theta<1,\ |z|<1,\ a>0. \qquad (2.5.4.11)$$

Hence the distribution with the pgf
$$G(z) = \left[(q-p)\left\{\frac{q+pz}{q-pz}\right\}\right]\left\{\frac{q-p}{q-pz}\right\}^a \qquad (2.5.4.12)$$
$$\simeq \left[\left\{\frac{1+\theta z}{1+\theta}\right\}\left\{\frac{1-\theta}{1-\theta z}\right\}\right]\left\{\frac{1-\theta}{1-\theta z}\right\}^a, \quad \theta \simeq \frac{p}{q} = 1/(\alpha+1) \qquad (2.5.4.13)$$
$$\simeq {}_2F_1\left\{a, \frac{a}{2}+1; \frac{a}{2}; \theta z\right\} / {}_2F_1\left\{a, \frac{a}{2}+1; \frac{a}{2}; \theta\right\} \qquad (2.5.4.14)$$

is inf div since it is the convolution of a geometric with-decreased-zeroes (an inf div distribution) with a negative binomial (another inf div distribution).

The probability function of this generalized hypergeometric probability distribution is

$$P_x = \frac{\Gamma(x + a)}{\Gamma(a)} \frac{\Gamma\left(x + 1 + \frac{a}{2}\right)}{\Gamma\left(x + \frac{a}{2}\right)} \frac{\theta^x}{x!} \qquad (2.5.4.-5)$$

and a recurrence relation between successive probability is

$$P_{x+1} = \theta\left(\frac{x + a}{x + 1}\right)\left(\frac{2x + a + 2}{2x + a}\right)P_x . \qquad (2.5.4.-6)$$

2.6 Genesis and Applications of GPDs

2.6.1 The clustering model

The clustering model was described in section 1.4.6, and the application of GPDs in the study of accident proneness was mentioned in section 2.4.4 to illustrate the clustering model. A comprehensive review on this has been given by Kemp (1970). Another area of application of GPDs already mentioned (section 2.5.2) as an example of the clustering model is the spatial distribution of populations of insects, plants, and bacteria, to which the review papers of Patil and Stiteler (1972, 1974) and Westman (1971) are closely related, as is a brief discussion by Boswell et al. (1979, pp.103-111).

Among the many other phenomena which can be studied by the clustering model is the emission of cosmic ray showers. It is generally supposed that the number N of cosmic ray showers has a Poisson distribution with parameter λ where λ is the expected number of showers per unit time-interval. For any given ray counter the number of registrations caused by a shower is a random variable having a distribution with pgf h(z). Then the number of registrations during a unit time-interval is a GPD having the pgf

$$G(z) = \exp\{\lambda[h(z) - 1]\}.$$

2.6.2 The stuttering-Poisson model

The stuttering-Poisson model is widely used in operational research e.g. in queueing and inventory problems (e.g. Galliher et al. (1959), Cox (1967, pp.95-96), and Adelson (1977)). It is desired to fit a theoretical probability distribution to "demands". The assumption is that orders of sizes 1, 2,... arrive simultaneously as independent Poisson streams of means a_i (i = 1, 2,...) orders per unit time respectively. The distribution of total demands per unit time is therefore the convolution of Poisson singlet, doublet, triplet, etc., distributions, and its pgf is

$$G(z) = \exp\{\Sigma a_i(z^i - 1)\}, \quad a_i \geq 0 \qquad (2.6.2.1)$$

$$= \exp\{\lambda[h(z) - 1]\} \qquad (2.6.2.2)$$

where

$$\lambda = \Sigma a_i, \qquad (2.6.2.3)$$

and $$h(z) = \Sigma(a_i/\lambda)z^i. \qquad (2.6.2.4)$$

Kemp (1967b) has discussed stuttering-Poisson distributions as generalized Poisson distributions. Galliher et al. (1959) use the Poisson-geometric distribution in their inventory study, although a GPD with a finite cluster-size distribution is considered more appropriate (Adelson (1977)). Kemp and Kemp's (1965) Hermite distribution having the pgf

$$G(z) = \exp\{a_1(z - 1) + a_2(z^2 - 1)\}, \qquad (2.6.2.5)$$

Gupta and Jain's (1974) generalized Hermite distribution with the pgf

$$G(z) = \exp\{a_1(z - 1) + a_m(z^m - 1)\}, \qquad (2.6.2.6)$$

Patel's (1976b) triple and quadruple stuttering-Poisson (i.e. the three-parameter and the four-parameter Hermite) distributions having respectively the pgfs

$$G(z) = \exp\{a_1(z - 1) + a_2(z^2 - 1) + a_3(z^3 - 1)\}, \qquad (2.6.2.7)$$

$$G(z) = \exp\{a_1(z - 1) + a_2(z^2 - 1) + a_3(z^3 - 1) + a_4(z^4 - 1)\}, \qquad (2.6.2.8)$$

and the Poisson-binomial having the pgf

$$G(z) = \exp\{\lambda[((1 - \theta) + \theta z)^m - 1]\}, \quad \theta \simeq 1/(\beta + 1) \qquad (2.6.2.9)$$

$$= \exp\left\{\sum_i^m a_i(z^i - 1)\right\} \qquad (2.6.2.10$$

with $$a_i = \left(\frac{\beta}{\beta + 1}\right)^m \binom{m}{i} \beta^i$$

are particularly suitable for studying the stuttering-Poisson in-flow of demands.

2.7 Summary on "Generalized Poisson Distributions"

In this chapter, the definition and several functional forms of generalized Poisson distributions have been given. Properties and applications of this family of discrete distributions have also been discussed.

A discrete distribution is said to be a generalized Poisson distribution if its pgf can be put in the form

$$G(z) = \exp\{\lambda[h(z) - 1]\}$$

where $\lambda > 0$ and $h(z)$ is a pgf. In a case where $h(z)$ takes values on $x = 0, 1, \ldots$ the form can be recast as

$$G(z) = \exp\{\gamma[h_+(z) - 1]\}$$

by putting

$$\gamma = \lambda\{1 - h(0)\}$$

and

$$h_+(z) = \{h(z) - h(0)\}/\{1 - h(0)\},$$

thereby giving $h_+(z)$ supported on $x = 1, 2, \ldots$ This latter form is called the Levy-Feller representation and is useful, in particular, for describing the spatial pattern of an aggregated population whose members are in randomly dispersed clusters.

It is often not obvious that a given discrete distribution is or is not a GPD when the pgf is presented in forms different from the above

two. Certain conditions have been given for identifying some distributions which are clearly not GPDs. For example, terminating and truncated lattice distributions are not GPDs, neither is a discrete distribution whose variance is less than its mean or less than P_1/P_0 where P_0, P_1 denote the probabilities of the variates zero and one respectively. Also, if a lattice distribution is non-increasing and

$$x(P_{x-1} - P_x) > P_0$$

for any of $x = 1, 2, \ldots$, then it is not a GPD.

It does not follow however that all other types of lattice distributions are necessarily GPDs. To ascertain that a given distribution is actually a GPD is in general not easy. Some sufficient conditions are however available for identifying certain types of GPDs. For example, if a lattice distribution is logconvex, that is, has probabilities P_x satisfying

$$\frac{P_{x+1}}{P_x} \geq \frac{P_x}{P_{x-1}} \qquad \text{for all } x = 1, 2, \ldots$$

then it is definitely a GPD. So is any completely monotone discrete distribution, that is, any distribution arising as a mixture of the geometric distribution having the pgf

$$h(z) = (1 - \theta)/(1 - \theta z), \qquad 0 < \theta < 1.$$

Some theorems have been stated, which can be applied in determining whether a given pgf is the pgf of such a mixture.

A lattice distribution with the pgf $G(z)$ is a GPD if and only if

$$V(z) = \lim_{\gamma \to 0} \left[\left\{ \{G(z)\}^{\gamma} - \{G(0)\}^{\gamma} \right\} / \left\{ \{G(1)\}^{\gamma} - \{G(0)\}^{\gamma} \right\} \right]$$

is a pgf (Kemp (1978)). If $G(z)$ is indeed a GPD then its cluster-size distribution is given by $V(z)$.

Another necessary and sufficient condition is given by the property of infinite divisibility. A discrete distribution is a GPD if and only if it is infinitely divisible i.e. if and only if its pgf $G(z)$ is such

that

$$h(z) = \{G(z)\}^{\alpha}$$

where h(z) is a pgf and α is any real positive number. The mixture of the Poisson by an infinitely divisible distribution is also infinitely divisible and is hence a GPD (see e.g. Maceda (1948)). Huff (1978) has even shown that if the mixing distribution from which a mixture of the Poisson is derived is not infinitely divisible then the mixture of the Poisson is not infinitely divisible.

Mixtures of the Poisson (compound Poisson distributions) have natural applications in models of heterogeneity while generalized Poisson distributions are useful for studying clustering models. The former are known as models of apparent contagion and the latter are called models of true contagion. Hence an infinitely divisible mixture of the Poisson is generated by chance mechanisms exhibiting both apparent contagion and true contagion simultaneously. The negative binomial is a well known example. Some other infinitely divisible mixtures of the Poisson have been derived in this chapter, from infinitely divisible continuous distributions supported on the whole or part of the positive half real line. One of these is the Poisson Λ reciprocal gamma distribution.

The types of GPDs discussed include infinitely divisible convolutions of pseudo-binomials. A function w(z) is called a pseudo-binomial if it is in the form

$$w(z) = (q + pz)^{\gamma}, \quad \gamma \text{ non-integral}, \ 0 < p < 1, \ q = 1 - p$$

or $$w(z) = (\beta + 1 - \beta z)^{\alpha}, \qquad \alpha > 0, \ \beta > 0.$$

(Kemp (1979)). Some infinitely divisible convolutions of pseudo-binomials given in the literature are re-cast into conventional forms.

Analytical aspects which are not directly of practical importance are excluded from the discussion. Hartman and Wintner (1942), Gnedenko (1973), and Cuppens and Lukacs (1969) have given interesting expositions on such aspects.

CHAPTER 3

THE POISSON INVERSE NORMAL DISTRIBUTION

CHAPTER 3

THE POISSON INVERSE NORMAL DISTRIBUTION

3.1 Introduction

Chapter 1 is about families, genesis, and applications of discrete distributions in general, while Chapter 2 is a discussion of the particular family of interest, the generalized Poisson distributions. In this chapter, we narrow down further and consider a specific generalized Poisson distribution, the Poisson Λ inverse normal (PIN) distribution. Our interest lies mainly in parameter estimation: in Section 3.2, the major parameter estimation methods in use for generalized Poisson distributions are reviewed. The rest of the chapter then dwells on the PIN distribution alone.

The Poisson Λ inverse normal (also commonly called Poisson Λ inverse Gaussian) distribution is the compound distribution obtained by allowing the Poisson parameter to follow the inverse normal distribution.

The inverse normal distribution (e.g. Johnson and Kotz (1970, pp.137-153) and Tweedie (1957)) has the density

$$f(x) = [\alpha/(2\pi x^3)]^{\frac{1}{2}} \exp[-\alpha(x-\beta)^2/(2\beta^2 x)] \qquad (3.1.0.1)$$

with $x > 0$, $\alpha > 0$, $\beta > 0$. It is otherwise called the inverse Gaussian or the Wald distribution. The distribution is widely used in sequential analysis and in the study of the motion of particles influenced by Brownian movement. Relatively simple estimators of its parameters have been shown to have desirable asymptotic as well as small-sample properties. Furthermore, many of its features are closely related to the corresponding features of the normal distribution. In particular, there is an inverse relationship between its cumulant generating function and that of the latter. Also, its cumulative distribution function, $F(x)$, is given by

$$P(x) = \Phi\left[\left(\frac{\alpha}{x}\right)^{\frac{1}{2}}\left(\frac{x}{\beta} - 1\right)\right] - e^{2\alpha/\beta}\Phi\left[-\left(\frac{\alpha}{x}\right)^{\frac{1}{2}}\left(\frac{x}{\beta} + 1\right)\right] \qquad (3.1.0.2)$$

where $\Phi(y)$ denotes the cumulative distribution function of the standard normal variable Y. Like the normal distribution it is unimodal and also has the property of reproducibility. All these properties recommend the distribution for consideration as a mixing distribution.

The moment generating function, M(t), of the inverse normal distribution is

$$M(t) = \exp\left\{\left(\frac{\alpha}{\beta}\right)\left[1 - \left[1 - \left(\frac{2\beta^2}{\alpha}\right)t\right]^{\frac{1}{2}}\right]\right\} \qquad (3.1.0.3)$$

from which the probability generating function, G(z), of the Poisson Λ inverse normal distribution is easily obtained as

$$G(z) = \exp\left\{\left(\frac{\alpha}{\beta}\right)\left[1 - \left[1 + \left(\frac{2\beta^2}{\alpha}\right)(z - 1)\right]^{\frac{1}{2}}\right]\right\} \qquad (3.1.0.4)$$

This pgf can be recast as

$$G(z) = \exp\left\{\left(\frac{\alpha}{\beta}\right)\left[1 - \left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}}\right]\right\}$$

$$= \exp\left\{a(b - 1)\left[\left(\frac{b}{b - 1}\right)(1 - \sqrt{1 - \theta z}) - 1\right]\right\}$$

$$= \exp\{\lambda[h(z) - 1]\} \qquad (3.1.0.5)$$

where

$$\lambda = a(b - 1) > 0$$

and

$$h(z) = \left(\frac{b}{b - 1}\right)(1 - \sqrt{1 - \theta z})$$

$$= z\,{}_2F_1(\tfrac{1}{2}, 1; 2; \theta z)/{}_2F_1(\tfrac{1}{2}, 1; 2; \theta) \qquad (3.1.0.6)$$

with

$$a = \alpha/\beta \qquad (3.1.0.7)$$

$$b = \left(1 + \frac{2\beta^2}{\alpha}\right)^{\frac{1}{2}} \qquad (3.1.0.8)$$

and

$$\theta = \left(\frac{2\beta^2}{\alpha}\right)\Big/\left(1 + \frac{2\beta^2}{\alpha}\right) \qquad (3.1.0.9)$$

This form clearly shows that the PIN distribution is a generalized Poisson distribution. Its cluster-size distribution has the pgf h(z) given in (3.1.0.6). The form of (3.1.0.6) and (2.4.3.29) shows its relationship to the mixture of the Poisson obtained with Feller's stable distribution, a special case of our new distribution, the Poisson Λ reciprocal gamma distribution (Chapter 2, Section 2.4.3).

On putting

$$p = \tfrac{1}{2}(1 + \sqrt{1 - \theta}), \quad q = 1 - p, \; 0 < p < 1 \qquad (3.1.0.10)$$

it is easy to see that

$$h(z) = z\,{}_2F_1(\tfrac{1}{2}, 1; 2; 4pqz)/{}_2F_1(\tfrac{1}{2}, 1; 2; 4pq) \qquad (3.1.0.11)$$

i.e. that h(z) is a lost-games distribution (Kemp and Kemp (1968)) having unity as the value of the exponent parameter, a.

The mean and variance of the PIN distribution are given respectively by

$$E(X) = \beta \qquad (3.1.0.12)$$

and

$$\mathrm{Var}(X) = \alpha + \frac{\beta^3}{\alpha} \qquad (3.1.0.13)$$

Certain properties of the distribution have been discussed by Holla (1966), Sankaran (1968), Sichel (1971, 1973a, 1973b, 1974, 1975, 1982a, 1982b), Shaban (1981, 1982) and Atkinson and Yeh (1982). Sankaran fits it to a European-corn-borer data-set and Sichel considers its fit to word-frequency data (1971, 1975), diamond-stones-density data (1973a, 1973b), and data obtained on customers' repeat-buying of household non-durable consumer-goods (1982b).

Sichel obtains the joint maximum likelihood estimators (1971) and the corresponding asymptotic covariance matrix for the estimates (1982a) of the parameters in a particular parameterization. He however expresses dissatisfaction with the parameterization because of substantial negative correlation between the two parameter estimates. In Section 3.3 we consider Holla's (1966) parameterization of the distribution which we find to be free of the defect in Sichel's.

For the parameters in Holla's parameterization, joint maximum likelihood equations and expressions for obtaining the corresponding asymptotic covariance matrix are derived. Shaban (1981) remarks that maximum likelihood estimation for the two parameters in this form of the distribution is intractable and he therefore settles for a particular case involving only one parameter. The main approach followed here (Section 3.3.1) is that based on the probability generating function, a technique proposed by Kemp (1967b). The mathematics involved are straightforward. Section 3.3.2 derives the same expressions from the probability function, thereby verifying the correctness of the expressions and providing a comparison of the two methods of proceeding.

The maximum likelihood equations obtained (in Section 3.3.3) are used to fit the PIN distribution to some sets of data to which Martin and Katti (1965) have fitted similar generalized Poisson distributions. The pattern of the iterative procedure in the fit of the PIN distribution is exhibited for two data sets. One is McGuire et al.'s (1957) frequently used European-corn-borer (short-tailed) data, which is also the Empirical Distribution No. 15 of Martin and Katti's. The other is Bertram's (1949) filarial worms (long-tailed) data. The equations are also used (in Section 3.3.4) to compute the asymptotic correlations of the maximum-likelihood estimates of the two parameters. All the results are discussed in Section 3.3.5.

As a further contribution to the study of the Poisson inverse normal distribution, we present in Section 3.4 the even-points estimators of the parameters. Maximum-likelihood estimation for this distribution is an iterative procedure and requires initial values of the estimates, values provided by explicit estimators such as those from the even-points method. Explicit estimators are, of course, also important in their own right. Besides those of the even-points method (Section 3.4.1), explicit estimators obtained from three other methods are given (in Section 3.4.2). The

expressions required for computing the asymptotic efficiencies of these four sets of estimators are derived in Section 3.3.3.

In Holla's (1966) parameterization the probability function of the PIN distribution is given by

$$P_x = (2\alpha/\pi)^{\frac{1}{2}} \frac{1}{x!} e^{(\alpha/\beta)} \left[\alpha/\left\{2\left(1 + \frac{\alpha}{2\beta^2}\right)\right\}^{\frac{1}{2}(x-\frac{1}{2})}\right] \cdot K_{(x-\frac{1}{2})}\left\{\left[2\alpha\left(1 + \frac{\alpha}{2\beta^2}\right)\right]^{\frac{1}{2}}\right\} \quad (3.1.0.14)$$

$$= \frac{w_2}{x!} (w_1)^x K_{(x-\frac{1}{2})}(w_0) \quad (3.1.0.15)$$

where

$$w_0 = \left[2\alpha\left(1 + \frac{\alpha}{2\beta^2}\right)\right]^{\frac{1}{2}} = ab \quad (3.1.0.16)$$

$$w_1 = \alpha/(ab) = \theta w_0/2 = \beta/b \quad (3.1.0.17)$$

and $$w_2 = (2\alpha/\pi)^{\frac{1}{2}}\left[\alpha/\left\{2\left(1 + \frac{\alpha}{2\beta^2}\right)\right\}\right]^{\frac{1}{4}} e^{(\alpha/\beta)}$$

$$= (2w_0/\pi)^{\frac{1}{2}} e^a \quad (3.1.0.18)$$

with a, b, and θ as given in (3.1.0.7), (3.1.0.8), and (3.1.0.9) respectively.

In Sichel's (1971, 1975, 1982a, 1982b) parameterization, the probability function is given by

$$P_x = (2\gamma/\pi)^{\frac{1}{2}} \exp\left\{\gamma(1-\theta)^{\frac{1}{2}}\right\} \cdot (\gamma\theta/2)^x \frac{1}{x!} K_{(x-\frac{1}{2})}(\gamma) \quad (3.1.0.19)$$

that is, in relation to Holla's, Sichel's parameters θ and γ are given by

$$\theta = \left(\frac{2\beta^2}{\alpha}\right) \Big/ \left(1 + \frac{2\beta^2}{\alpha}\right) \quad (3.1.0.20)$$

and $$\gamma = w_0 = \left(\frac{\alpha}{\beta}\right)\left(1 + \frac{2\beta^2}{\alpha}\right)^{\frac{1}{2}} \quad (3.1.0.21)$$

The recurrence relation for the modified Bessel function of the second kind is given by

$$K_{c+1}(y) = \frac{2c}{y} K_c(y) + K_{c-1}(y) \quad (3.1.0.22)$$

From this, the recurrence relation for the probabilities of the PIN distribution is easily obtained as

$$(x + 1)P_{x+1} = \theta(x - \tfrac{1}{2})P_x + (w_1^2/x)P_{x-1}, \qquad x = 1, 2,... \qquad (3.1.0.23)$$

with

$$P_0 = \exp(-a(b - 1)), \quad P_1 = w_1 P_0 \qquad (3.1.0.24)$$

Although theoretically the parameters α and β can each take any non-negative value, only limited ranges of values are important from the point of view of application. Firstly, since the distribution is a GPD, it is clear that $P_0 > 0$. For practical purposes, $P_0 > 1.0 \times 10^{-25}$ is a very non-restrictive requirement. The corresponding limits on the parameter values follow accordingly. The second requirement affecting the range of parameter values is the tail-length, the effective support. One way of determining this is to compute percentage points, such as has been done to obtain Table 1.

Entries in the table are the minimum values of m such that

$$T_m \equiv 1 - \sum_{x=0}^{m} P_x < 0.0001. \qquad (3.1.0.25)$$

Entries greater than 300 in this table correspond to parameter values which are rarely encountered in practical situations. A few empirical distributions with non-trivial support of up to x = 1000 have been cited in the literature (e.g. Herdan (1964)). However the sample size required to reasonably represent distributions with such a wide effective support is rarely attainable.

3.2 Parameter Estimation Methods for GPDs

3.2.1 Maximum likelihood estimation

The maximum-likelihood method is generally regarded as the standard estimation method. This is mainly because of its desirable asymptotic properties. The method also has the advantage that the estimators it

provides for different parameterizations of a given distribution are invariant. However the (asymptotic) superiority of maximum-likelihood does not necessarily apply in small sample situations.

Suppose we have a sample of size $N = \Sigma f_x$ where f_x is the number of members of the sample taking the value x, and we wish to estimate the parameters of a distribution by the maximum-likelihood method. Let P_x be the corresponding theoretical probability based on the particular values of the parameters under consideration. Then the loglikelihood of the sample is

$$L = \Sigma f_x \log P_x . \qquad (3.2.1.1)$$

Note that maximizing the loglikelihood is equivalent to maximizing the likelihood.

The subject of Nonlinear Optimization has provided techniques for performing such maximization. Most of these techniques are presented as minimization procedures. In this case then the objective function to minimize is the negative loglikelihood. In general we are dealing with unconstrained minimization (e.g. Kennedy and Gentle (1980, pp.425-475)), except that for lattice distribution positivity constraints are often attached, since in most cases the parameters being estimated are positive-valued.

Most of the good minimization techniques usually require that first or second derivatives of the objective function be provided. Fisher's scoring method (see e.g. Rao (1973, pp.366-374)), a version of the well-known Newton-Raphson method, uses the matrix of the expected values of the second derivatives instead of the actual Hessian matrix. Like the Newton-Raphson and all its other modifications, when a stationary point is reached the Hessian matrix is available for use in determining the precision of the minimum. It is the method preferred here because of this advantage of having the Hessian matrix for computing the asymptotic covariance matrix of the maximum likelihood estimates and

hence obtaining at least an indicative measure of their precision. The method is also relatively easy to set up.

Those minimization techniques like Nelder and Mead's (1965) and Hooke and Jeeves' (1961) which do not require any derivatives, although generally regarded as less satisfactory, have been found to be adequate for most ordinary purposes (Ross (1970)). These are the so-called "direct search" procedures. A locally written direct search algorithm called MAXIM is used as a verification tool in this work (see Tables IIA, IIIA).

Except in the case when the likelihood is maximized directly or an equivalent nonlinear optimization is performed, most discussion of maximum-likelihood estimation is about the derivation of a set of likelihood equations, usually from the partial derivatives of the loglikelihood. These equations are then solved, in some cases explicitly, but most often iteratively.

Without loss of generality, suppose we have two parameters α and β to be estimated by the maximum-likelihood method. Then

$$S_\alpha \equiv \frac{\partial L}{\partial \alpha} = \Sigma f_x \frac{\partial}{\partial \alpha} \log P_x \qquad (3.2.1.2)$$

$$= \Sigma f_x \frac{1}{P_x} \frac{\partial}{\partial \alpha} P_x \qquad (3.2.1.3)$$

$$= \Sigma f_x A_x \qquad (3.2.1.4)$$

and similarly,

$$S_\beta \equiv \frac{\partial L}{\partial \beta} = \Sigma f_x \frac{1}{P_x} \frac{\partial}{\partial \beta} P_x \qquad (3.2.1.5)$$

$$= \Sigma f_x B_x \qquad (3.2.1.6)$$

where

$$A_x = \frac{1}{P_x} \frac{\partial}{\partial \alpha} P_x, \qquad (3.2.1.7)$$

and

$$B_x = \frac{1}{P_x} \frac{\partial}{\partial \beta} P_x. \qquad (3.2.1.8)$$

The joint maximum-likelihood equations then are

$$S_\alpha \equiv \Sigma f_x A_x = 0 \qquad (3.2.1.9)$$

and

$$S_\beta \equiv \Sigma f_x B_x = 0. \qquad (3.2.1.10)$$

This set of equations can often be simplified, but in this particular form it is relatively easy to solve by the scoring method. We shall therefore refer to it as the scoring form of maximum-likelihood equations. Note that the partial derivatives $\frac{\partial}{\partial\alpha} P_x$ and $\frac{\partial}{\partial\beta} P_x$ can be easily obtained from the probability generating function G(z) (e.g. Kemp (1967a, 1967b)).

The asymptotic covariance matrix of the joint maximum-likelihood estimators α and β is the inverse of the Fisher information matrix, I. Because of the complexity of the second derivatives of L with respect to the parameters, it is usual to replace them by their expected values (see e.g. Kemp (1967a)) and use can therefore be made of the well-known relationship

$$E\left\{\frac{\partial^2 L}{\partial\alpha\partial\beta}\right\} = -E\left\{\frac{\partial L}{\partial\alpha} \cdot \frac{\partial L}{\partial\beta}\right\}$$

$$\simeq -NE\left\{\frac{\partial}{\partial\alpha}\log P_x \cdot \frac{\partial}{\partial\beta}\log P_x\right\}. \qquad (3.2.1.11)$$

where N is the sample size. This means that, with

$$S_\alpha \simeq \Sigma f_x A_x \quad \text{and} \quad S_\beta = \Sigma f_x B_x,$$

the corresponding elements of the information matrix are estimated by

$$I_{\alpha\alpha} = N\Sigma P_x A_x^2 \qquad (3.2.1.12)$$

$$I_{\beta\beta} = N\Sigma P_x B_x^2, \qquad (3.2.1.13)$$

and

$$I_{\alpha\beta} = N\Sigma P_x A_x B_x.$$

The asymptotic correlation coefficient $r_{\alpha\beta}$ is computed from the elementary expression

$$r_{\alpha\beta} = \mathrm{cov}(\alpha, \beta)/[\mathrm{var}(\alpha) \cdot \mathrm{var}(\beta)]^{\frac{1}{2}}. \qquad (3.2.1.15)$$

All the discussion here is in reference to first order asymptotics. Shenton and Bowman (1963, 1967) and their school have lately given attention to higher order asymptotic properties of maximum-likelihood estimators, and other types of estimators for different discrete distributions. Much of their discussions tend to be clouded by heavy algebra. Their results mainly serve to reiterate the common warning that first order asymptotic properties of estimators are better regarded as screening tools. Evidently estimators with poor first-order asymptotic properties are unlikely to be satisfactory in small samples.

Sprott (1965) has observed that many generalized Poisson distributions have a maximum-likelihood equation of the form

$$\Sigma f_x (x + 1)(P_{x+1}/P_x) = N\bar{x} \qquad (3.2.1.16)$$

together with

$$c\alpha\beta = \bar{x} \qquad (3.2.1.17)$$

or a similar one, where α, β are the estimators for the parameters being estimated, c being a constant. Sprott shows that distributions having (3.2.1.16) as a maximum-likelihood equation have pgfs of the form

$$\begin{aligned} G(z;\ \alpha,\ \beta) &= G_1\left\{H[(1 - z)\ u(\beta)];\ \alpha\right\} \\ &= G_2\left\{(1 - z)\ u(\beta);\ \alpha\right\} \end{aligned} \qquad (3.2.1.18)$$

The ghfd's with pgfs as given in (1.4.4.17) are of this form.

The maximum-likelihood estimation of the parameters of the negative binomial

$$G(z) = (1 + \beta - \beta z)^{-\alpha}, \quad \alpha > 0,\ \beta > 0 \qquad (3.2.1.19)$$

has been discussed by Haldane (1941), Anscombe (1950), Bliss and Fisher (1953), and Simonsen (1976). Note that the maximum-likelihood estimates of the parameters of another form of the distribution are asymptotically uncorrelated. It is desirable to ascertain and use the parameterization of a given distribution for which the maximum likelihood estimates are asymptotically uncorrelated. One of the contributions of this thesis

is the observation that Holla's (1966) parameterization of the Poisson inverse normal distribution is of this type (see Section 3.3 below).

Haldane derives the likelihood equations for the negative binomial as

$$\alpha\beta = \bar{x} \qquad (3.2.1.20)$$

$$N\log(1 + \bar{x}/\alpha) = \sum_{x=0}^{\infty} r_x/(\alpha + x) \qquad (3.2.1.21)$$

where $r_i = \sum_{j=i+1} f_j$.

Bliss and Fisher obtain the "scores" and with these the asymptotic covariance matrix for the estimates produced by the equations. They also suggest the use of the scoring method for solving the equations. However if the scoring method is to be used it is much better to have the equations in the scoring form. In the case of the negative binomial, since

$$(x + 1)P_{x+1} = \left(\frac{\beta}{\beta + 1}\right)(x + \alpha)P_x, \qquad (3.2.1.22)$$

we have for the scoring expressions

$$A_x \equiv \frac{1}{P_x}\frac{\partial}{\partial\alpha}P_x$$

and

$$B_x \equiv \frac{1}{P_x}\frac{\partial}{\partial\beta}P_x,$$

the simple recursive formulae

$$A_{x+1} = \frac{1}{x + \alpha} + A_x, \qquad x = 0, 1,... \qquad (3.2.1.23)$$

with $A_0 = -\log(\beta + 1)$, and

$$B_{x+1} = \frac{\beta + 1}{\beta} + B_x, \qquad x = 0, 1,... \qquad (3.2.1.24)$$

with $B_0 = -\alpha/(\beta + 1)$.

Anscombe uses the zero frequency, instead of the sample mean, to derive a set of likelihood equations similar to Haldane's and conjectures that these equations may not have a solution under certain given conditions. Simonsen has discussed these and other conditions relating to the existence of a solution to Anscombe's equations.

Pieters et al. (1977) have conducted a simulation study of the small-sample properties of the estimators in Haldane's equations. Second- and higher-order asymptotic properties of the estimators of the negative binomial parameters have been investigated by Shenton and Wallington (1962).

The maximum-likelihood equations for the Neyman Type A distribution

$$G(z) = \exp\{\beta[e^{\alpha(z-1)} - 1]\} \tag{3.2.1.25}$$

are

$$\alpha\beta = \bar{x} \tag{3.2.1.26}$$

$$\Sigma f_x(x+1)P_{x+1}/P_x = n\alpha\beta. \tag{3.2.1.27}$$

Douglas (1955) has discussed the equations. For the scoring form of the equations we have, corresponding to α, β, the scoring expressions

$$A_x = \left(\frac{x+1}{\alpha}\right)\left(P_{x+1}/P_x\right) - \left(\frac{1}{\alpha}\right), \tag{3.2.1.28}$$

amd

$$B_x = \left(\frac{x+1}{\alpha\beta}\right)\left(P_{x+1}/P_x\right) - 1. \tag{3.2.1.29}$$

Evans (1953) has presented another set of likelihood equations. Shenton and Bowman (1967) have discussed the first and second order asymptotic properties of the maximum-likelihood estimators of the Neyman Type A parameters.

Kemp (1967a) has derived the scoring form of the maximum-likelihood equations for estimating the parameters a, b, c of the Short distribution

$$G(z) = \exp\left\{(b+c)\left\{\left[\left(\frac{c}{b+c}\right)e^{a(z-1)} + \left(\frac{b}{b+c}\right)z\right] - 1\right\}\right\}. \tag{3.2.1.30}$$

For this distribution,

$$A_x = \left(\frac{x+b}{a}\right) - \left(\frac{x+1}{a}\right)\left(P_{x+1}/P_x\right) - \left(\frac{b}{a}\right)\left(P_{x-1}/P_x\right) \tag{3.2.1.31}$$

$$B_x = \left(\frac{x+1}{ac}\right)\left(P_{x+1}/P_x\right) - \left(1 - \frac{b}{ac}\right) \tag{3.2.1.32}$$

$$C_x = \left(P_{x-1}/P_x\right) - 1. \tag{3.2.1.33}$$

The estimation of the parameters of the Poisson V binomial

$$G(z) = \exp\{\lambda[(q + pz)^n - 1]\} \qquad (3.2.1.34)$$

by the maximum-likelihood method has been considered by Sprott (1958).

The likelihood equations are

$$n\lambda p = \bar{x} \qquad (3.2.1.35)$$

$$\Sigma f_x(x + 1)(P_{x+1}/P_x) = Nn\lambda p. \qquad (3.2.1.36)$$

The special case n = 2 of the Poisson V binomial is Kemp and Kemp's (1965) Hermite distribution

$$G(z) = \exp\{a_1(z - 1) + a_2(z^2 - 1)\} \qquad (3.2.1.37)$$

with $a_1 = 2\lambda pq$ and $a_2 = \lambda p^2$,

for which the scoring form of maximum-likelihood equations is obtained with

$$A_x = (P_{x-1}/P_x) - 1, \qquad (3.2.1.38)$$

and

$$B_x = (P_{x-2}/P_x) - 1. \qquad (3.2.1.39)$$

Patel et al. (1974) have discussed the first and second-order properties of the corresponding maximum-likelihood estimators.

Katti and Gurland (1962b), in their discussion of the properties of the Poisson V negative-binomial distribution

$$G(z) = \exp\{\lambda[(Q - Pz)^{-\alpha} - 1]\}, \quad Q = P + 1 \qquad (3.2.1.40)$$

obtain the scoring expressions A_x, B_x, C_x for α, P, λ respectively as

$$A_x = -\left(\frac{1}{\alpha P}\right)\log Q\left[Q(x + 1)(P_{x+1}/P_x) - xP\right] + \left(\frac{1}{\alpha P}\right)(1/P_x)\sum_{r=1}^{x}(P/Q)^r\left(\frac{1}{r}\right)\left[(x + 1 - r)P_{x+1-r} - P(x - r)P_{x-r}\right] \qquad (3.2.1.41)$$

$$B_x = (1/P)[x - (x + 1)(P_{x+1}/P_x)] \qquad (3.2.1.42)$$

$$C_x = \left(\frac{Q}{\alpha P\lambda}\right)(x + 1)(P_{x+1}/P_x) - \left(\frac{x}{\alpha\lambda}\right) \qquad (3.2.1.43)$$

Shumway and Gurland (1960b) have derived another set of equations, in matrix form, for which they and Gebski (1975) suggest solution techniques.

The Poisson V geometric, otherwise known as the Polya-Aeppli, is a special case of the Poisson V negative binomial. Its pgf is

$$G(z) = \exp\left\{\lambda\left[\left(\frac{1-\theta}{1-\theta z}\right) - 1\right]\right\}. \qquad (3.2.1.44)$$

Evans (1953) and Kemp (1967b) have discussed the maximum-likelihood estimation of the parameters in its Polya-Aeppli form

$$G(z) = \exp\left\{\alpha\left[z\left(\frac{1-\theta}{1-\theta z}\right) - 1\right]\right\}, \quad \alpha = \lambda\theta. \qquad (3.2.1.45)$$

The scoring form of the likelihood equations can be obtained with A_x, B_x corresponding to θ, λ and given by

$$A_x = \{\theta x - [1 - \theta(1-\theta)](x+1)(P_{x+1}/P_x)\}/\theta(1-\theta) \qquad (3.2.1.46)$$

$$B_x = \left(\frac{1}{\lambda\theta}\right)(x+1)(P_{x+1}/P_x) - \left(1 + \frac{x}{\lambda}\right). \qquad (3.2.1.47)$$

3.2.2 The moments method

The moments method of estimation consists in equating the first k population moments to the first k sample moments respectively, resulting usually in an explicit estimator for each of the k parameters to be estimated. Moments estimators are often the easiest to derive and are therefore commonly the estimators presented when a distribution is being newly proposed. When (as is often the case) they are explicit estimators, they can also be employed in obtaining initial values for an iterative maximum-likelihood procedure. Because of the low precision of the higher moments, only the first and second are normally used. This means that usually when more than two parameters are to be estimated, other statistics are combined with the first two sample moments. Such other statistics are discussed in Sections 3.2.3 and 3.2.4 below.

The first-order asymptotic covariance matrix, V, for moments estimates has been derived by Fisher (1941) as follows (see also Kemp (1967a)).

Let $m_1, m_2, \ldots, m_3$ be the first k sample moments and k_1, k_2, k_3, etc. the theoretical cumulants of the distribution. Denote by M the asymptotic covariance matrix of the m_i's (e.g. $var(m_1) = k_2/N$, $cov(m_1, m_2) = k_3/N$, etc.) evaluated at N = 1, e.g. for the first three sample moments

$$M = \begin{bmatrix} k_2 & k_3 & k_4 \\ k_3 & k_4 + 2k_2^2 & k_5 + 6k_2k_3 \\ k_4 & k_5 + 6k_2k_3 & k_6 + 9k_3^2 + 9k_2k_4 + 6k_2^3 \end{bmatrix} \qquad (3.2.2.1)$$

Suppose there are r parameters $a_1, \ldots, a_r$ to be estimated using some or all of the first three moments m_1, m_2, m_3. Then their estimators denoted by $\hat{a}_1, \ldots, \hat{a}_r$ are each a function of m_1, m_2, m_3 or whichever subset of (m_1, m_2, m_3) is used. In general, suppose

$$\hat{a}_i \equiv u_i(m_1, m_2, m_3), \qquad i = 1, \ldots, r.$$

Let $H = [\partial u_i / \partial m_j]$ be the Jacobian matrix of the transformation evaluated at the expected values of m_1, m_2, m_3 and denote its transpose by H'. Then the matrix V is given by

$$V = HMH'. \qquad (3.2.2.3)$$

With the matrix V the asymptotic efficiency of the moments estimators is then computed using

$$E = |F|/|V| \qquad (3.2.2.4)$$

where $F = I^{-1}$ is the asymptotic covariance matrix for the corresponding maximum-likelihood estimates, I being the information matrix.

First-order asymptotic efficiency of moments estimators for the negative binomial parameters has been discussed by Fisher (1941), Anscombe (1950) and Katti and Gurland (1962b). Pieters et al. (1977) have performed a simulation study of the corresponding small-sample

properties. Higher-order properties of these estimators have been considered by Shenton and Myers (1965) and Shenton and Bowman (1967).

For the Neyman Type A, Katti and Gurland (1962b) compute the first-order asymptotic efficiency while Shenton and Bowman (1967) discuss the higher-order asymptotic properties of the same moments estimators. Kemp (1967a) has computed first-order efficiency for the moments estimators of the Short distribution parameters. A similar efficiency study has been done for the Poisson V binomial by Katti and Gurland (1962b) and for its special case n = 2, the Hermite distribution, by Patel (1976a).

For the Poisson V negative binomial, Katti and Gurland (1961) have evaluated, at certain values of the parameters, the expression which they have derived for the first-order asymptotic efficiency of the moments estimators. Evans (1953) obtains approximations for the asymptotic variances and covariance of the moments estimates of the Polya-Aeppli (Poisson V geometric) parameters. Shenton et al. (1967) have discussed the second- and higher-order properties of these estimators.

3.2.3 The method of first two frequencies and their ratio

When a high proportion of the observations in a random-count investigation take the variate values zero and one, the corresponding frequencies can often be used in the estimation of the parameters. The statistics derived therefrom are usually, but not always, utilized in combination with the sample moments.

The first order asymptotic covariance matrix, V_1, of the statistics $\bar{x}$ and $\log f_0$ is given by

$$V_1 = \begin{bmatrix} k_2 & -k_1 \\ -k_1 & (1 - P_0)/P_0 \end{bmatrix} \qquad (3.2.3.1)$$

(see Katti and Gurland (1962a)). The matrix V_2 for $\bar{x}$ and f_1/f_0 is given by

$$V_2 \simeq \begin{bmatrix} k_2 & \frac{P_1}{P_0} \\ \frac{P_1}{P_0} & \left(\frac{1}{P_0}\right)\left(\frac{P_1}{P_0}\right)\left(1 + \frac{P_1}{P_0}\right) \end{bmatrix} \qquad (3.2.3.2)$$

Anscombe (1950) uses the statistics $\bar{x}$ and f_0 to obtain estimators for the parameters of the negative binomial. The same author has also computed the asymptotic efficiency of the estimators. Evans (1953), considering the estimation of the Neyman Type A parameters, derives an equation involving the statistics $\bar{x}$ and f_0 and states that $\text{cov}(\bar{\alpha}, \bar{\lambda})$ for the resulting estimates of α, λ is approximately α. Kerr (1969) computes the efficiency of the estimators obtained with $\bar{x}$, s^2 and $\log f_0$ for the Short distribution. Katti and Gurland (1962b) have done a similar efficiency study for the estimators obtained with $\bar{x}$ and f_0 for the special cases n = 2, 3, 5 of the Poisson binomial. The special case n = 2 is, of course, the Hermite distribution, for which the same type of efficiency study has been separately carried out by Patel (1976a).

Katti and Gurland (1961) present tables of the efficiency attained at certain parameter values by the estimators derived from the set $\bar{x}$, s^2 and $\log f_0$ and also from the set $\bar{x}$, s^2, f_1/f_0 for the parameters of the Poisson V negative-binomial distribution. Evans (1953) also uses the statistic f_0 in the estimation of the parameter of the Polya-Aeppli distribution.

3.2.4 The even-points method

The even-points method introduced by Patel (1976a) is based on the relationship

$$G(1) + G(-1) = 2\Sigma P_{2x} \qquad (3.2.4.1)$$

and

$$G(1) - G(-1) = 2\Sigma P_{2x+1} \qquad (3.2.4.2)$$

where G(z) denotes the pgf of a lattice distribution and P_i the probability that the random variable takes the value i.

Let f_i denote the observed frequency for the variate i, and let

$$S_e = \Sigma f_{2x}, \quad S_0 = \Sigma f_{2x+1}, \text{ and } T_x = \Sigma x f_x. \tag{3.2.4.3}$$

The statistics with which the even-points estimators are derived are S_e or S_0 together with $\bar{x} \equiv T_x/N$, f_0, f_1 or the ratio of the latter two. The derivation of the asymptotic variances and covariances of the statistics S_e, S_0, and T_x follows.

Note that

$$\begin{aligned} G(z) + G(-z) &= \Sigma P_x z^x + \Sigma(-1)^x P_x z^x \\ &= P_0 z^0 + P_1 z^1 + P_2 z^2 + P_3 z^3 + \ldots \\ &\quad + \{P_0 z^0 - P_1 z^1 + P_2 z^2 - P_3 z^3 + \ldots\} \\ &= 2\Sigma P_{2x} z^{2x} \end{aligned} \tag{3.2.4.4}$$

Similarly,

$$G(z) - G(-z) = 2\Sigma P_{2x+1} z^{2x+1} \tag{3.2.4.5}$$

On putting z = 1, we obtain

$$\tfrac{1}{2}\{G(1) + G(-1)\} = \Sigma P_{2x} \equiv \theta_e \tag{3.2.4.6}$$

$$\tfrac{1}{2}\{G(1) - G(-1)\} = \Sigma P_{2x+1} \equiv \theta_0 \tag{3.2.4.7}$$

as stated above. Note however that G(1) = 1.

An unbiased estimator for $N\theta_e$ is S_e and that for $N\theta_0$ is S_0, where N is the sample size. Note that $\theta_e + \theta_0 = 1$. Hence we can consider S_e as a random variable following the binomial distribution with the parameters N, θ_e. Similarly S_0 is distributed as binomial with the parameters N, θ_0. Therefore the means, variances, and covariance of S_e, S_0 are given by

$$\mu_{s_e} = N\theta_e, \qquad \mu_{s_0} = N\theta_0,$$

$$\text{var}(S_e) = N\theta_e(1 - \theta_e), \qquad \text{var}(S_0) = N\theta_0(1 - \theta_0),$$

$$\text{and} \quad \text{cov}(S_e, S_0) = -N\theta_e\theta_0. \tag{3.2.4.8}$$

To obtain $\text{cov}(S_e, T_x)$ and $\text{cov}(S_0, T_x)$, note that if the joint moment generating function of two random variables Y, Z are given as

M(t, s), then the joint cumulant generating function K(t, s) is

$$K(t, s) = \log M(t, s) \tag{3.2.4.9}$$

and the covariance of Y, Z is given by

$$\mathrm{cov}(Y, z) = \left[\frac{\partial^2}{\partial t \partial s} K(t, s)\right]_{(t=0, s=0)} \tag{3.2.4.10}$$

So we first require an expression for the joint mgf of S_e, S_0, T_x.

Substituting $z = e^t$ in

$$\tfrac{1}{2}[G(z) + G(-z)] = \Sigma P_{2x} z^{2x} \tag{3.2.4.11}$$

$$\tfrac{1}{2}[G(z) - G(-z)] = \Sigma P_{2x+1} z^{2x+1}, \tag{3.2.4.12}$$

we have

$$\tfrac{1}{2}[G(e^t) + G(-e^t)] = \Sigma P_{2x} e^{2xt} \equiv A \tag{3.2.4.13}$$

$$\tfrac{1}{2}[G(e^t) - G(-e^t)] = \Sigma P_{2x+1} e^{(2x+1)t} \equiv B \tag{3.2.4.14}$$

Note also that

$$A + B = G(e^t). \tag{3.2.4.15}$$

Evaluating these at t = 0 gives

$$(A + B)_0 = G(1) = 1 \tag{3.2.4.16}$$

$$(A)_0 = \tfrac{1}{2}[G(1) + G(-1)] = \theta_e \tag{3.2.4.17}$$

$$(B)_0 = \tfrac{1}{2}[G(1) - G(-1)] = \theta_0 \tag{3.2.4.18}$$

Let X_k be a random observation from the distribution under consideration, i.e. the lattice distribution with the pgf G(z). For this observation, each of S_e, S_0 can take the value 1 or 0. The random variable $S = (S_e, S_0)$ takes the following values with the associated probabilities

$$(1, 0) \text{ with prob } \theta_e$$

and $$(0, 1) \text{ with prob } \theta_0.$$

Hence the joint mgf of S_e, S_0, X_k is given by

$$\begin{aligned} M_{*}(u, v, w) &= E(\exp(uS_e + vS_0 + wX_k)) \\ &= E(e^H), \text{ say} \\ &= E(e^H | S = (1, 0)) \cdot \theta_e + E(e^H | S = (0, 1)) \cdot \theta_0 \end{aligned} \tag{3.2.4.19}$$

However, when S = (1, 0),

$$\text{Prob}(X_k = i) = P_i/\theta_e, \text{ if } i \text{ is } \underline{\text{even}}$$
$$= 0, \quad \text{otherwise} \qquad (3.2.4.20)$$

i.e.

$$E(e^H|S = (1, 0)) = (1/\theta_e)[P_0 e^u + P_2 e^{u+2w} + \ldots + P_{2x} e^{u+2xw} + \ldots]$$
$$= (1/\theta_e)e^u \Sigma P_{2x} e^{2xw}. \qquad (3.2.4.21)$$

Similarly, when S = (0, 1),

$$\text{Prob}(X_k = i) = P_i/\theta_0, \text{ if } i \text{ is } \underline{\text{odd}}$$
$$= 0, \quad \text{otherwise.} \qquad (3.2.4.22)$$

i.e.

$$E(e^H|S = (0, 1)) = (1/\theta_0)[P_1 e^v + P_3 e^{v+3w} + \ldots + P_{2x+1} e^{v+(2x+1)w} + \ldots]$$
$$= (1/\theta_0)e^u \Sigma P_{2x+1} e^{(2x+1)w}. \qquad (3.2.4.23)$$

Hence, we have

$$M_*(u, v, w) = E(\exp(uS_e + vS_0 + wX_k))$$
$$= E(e^H|S = (1, 0)) \cdot \theta_e + E(e^H|S = (0, 1)) \cdot \theta_0$$
$$= e^u \Sigma P_{2x} e^{2xw} + e^v \Sigma P_{2x+1} e^{(2x+1)w}$$
$$= e^u A + e^v B \qquad (3.2.4.24)$$

with

$$A \equiv \Sigma P_{2x} e^{2xw} = \tfrac{1}{2}[G(e^w) + G(-e^w)] \qquad (3.2.4.25)$$

and

$$B \equiv \Sigma P_{2x+1} e^{(2x+1)w} = \tfrac{1}{2}[G(e^w) - G(-e^w)] \qquad (3.2.4.26)$$

Therefore for N independent observations, X_k, k = 1, 2,..., N, identically distributed as G(z), the joint mgf of S_e, S_0, and $T_X = \sum_{i=1}^{N} X_k$ is given by

$$M(u, v, w) = [M_*(u, v, w)]^N$$
$$= (e^u A + e^v B)^N. \qquad (3.2.4.27)$$

The joint cgf of S_e, S_0, T_x, denoted by $K \equiv K(u, v, w)$, is given by

$$K(u, v, w) = \log M$$
$$= N\log(e^u A + e^v B) \qquad (3.2.4.28)$$

$$\frac{\partial K}{\partial u} = (NAe^{u})/(e^{u}A + e^{v}B) \tag{3.2.4.29}$$

$$\frac{\partial^2 K}{\partial w \partial u} = \left\{\left[Ne^{u}(e^{u}A + e^{v}B)\frac{\partial A}{\partial w}\right] - \left[(NAe^{u})\left(e^{u}\frac{\partial A}{\partial w} + e^{v}\frac{\partial B}{\partial w}\right)\right]\right\}\Big/(e^{u}A + e^{v}B)^2$$

$$= (Ne^{u}/M_{*}^{2})\left[M_{*}\frac{\partial A}{\partial w} - A\left(e^{u}\frac{\partial A}{\partial w} + e^{v}\frac{\partial B}{\partial w}\right)\right]. \tag{3.2.4.30}$$

Evaluating this at $\underline{c} = \underline{0}$ where $\underline{c} = (u, w, w)$, we have

$$\mathrm{Cov}(S_e, T_x) = \left[\frac{\partial^2 K}{\partial w \partial u}\right]_{\underline{c} = \underline{0}}$$

$$= N\left[\frac{\partial A}{\partial w} - A\left(\frac{\partial A}{\partial w} + \frac{\partial B}{\partial w}\right)\right]_{\underline{c} = \underline{0}} \tag{3.2.4.31}$$

Similarly,

$$\mathrm{Cov}(S_o, T_x) = \left[\frac{\partial^2 K}{\partial w \partial u}\right]_{\underline{c} = \underline{0}}$$

$$= N\left[\frac{\partial B}{\partial w} - B\left(\frac{\partial A}{\partial w} + \frac{\partial B}{\partial w}\right)\right]_{\underline{c} = \underline{0}}. \tag{3.2.4.32}$$

Patel (1976a) obtains the even-points estimators of the parameters of the Hermite distribution

$$G(z) = \exp\{a(z - 1) + b(z^2 - 1)\}, \quad a > 0, b > 0 \tag{3.2.4.33}$$

as

$$\hat{a} = \tfrac{1}{2} - \nu \tag{3.2.4.34}$$

$$\hat{b} = \tfrac{1}{2}(\bar{x} - \bar{a}) \tag{3.2.4.35}$$

where

$$\nu = \log\left(\frac{2S_e}{N} - 1\right), \tag{3.2.4.36}$$

which are admissible provided that

$$S_e < \frac{N}{2} \quad \text{and} \quad -2x < \nu < 0. \tag{3.2.4.37}$$

For Gupta and Jain's (1974) generalized Hermite

$$G(z) = \exp\{a(z - 1) + b(z^m - 1)\}, \quad m \text{ specified}, \tag{3.2.4.38}$$

the form of the even-points estimators depends on whether m is odd or even. If m is even, the estimators are given by

$$\hat{a} = -\tfrac{1}{2}\nu \tag{3.2.4.39}$$

$$\hat{b} = (\bar{x} - \bar{a})/m \tag{3.2.4.40}$$

provided that $-2\bar{x} < \nu < 0$. If m is odd, they are given as

$$\hat{a} = -\tfrac{1}{2}(2\bar{x} + m\nu)/(m - 1) \qquad (3.2.4.41)$$

$$\hat{b} = \tfrac{1}{2}(2\bar{x} + \nu)/(m - 1) \qquad (3.2.4.42)$$

provided that $-2\bar{x} < \nu < -\frac{2\bar{x}}{m}$.

The even-points estimators of the parameters of the Poisson Λ inverse normal distribution are derived in Section 3.3.3.

3.2.5 The minimum-chi-square method

The underlying principle behind the various versions of the minimum chi-square estimation method is the minimization of the function

$$\chi^2 = N[\Sigma(f_i^2/NP_i) - 1] \qquad (3.2.5.1)$$

(where N is the sample size and f_i, P_i are respectively the frequency and the probability of the ith variate), or a modified form of it (e.g. Rao (1973, pp.352-353). This method of parameter estimation has been reviewed by Rao (1957) and Berkson (1980). The version of the method used for the estimation of the parameters of GPDs and other lattice distributions is due to Gurland and his school (see e.g. Katti and Gurland (1962a, 1962b), Gurland (1965), Hinz and Gurland (1967), Gurland and Hinz (1971), Gurland and Tripathi (1975), Tripathi and Gurland (1977)). This version is actually not a distinct estimation method but rather a refinement technique, a kind of variance-stabilization procedure. It is a form of weighted least-squares.

Given the linear relation

$$\underline{T} = \underline{C}\,\underline{\theta} \qquad (3.2.5.2)$$

where $\underline{T}$ is a $k \times 1$ vector of population statistics;

$\underline{\theta}$ is an $m \times 1$ vector of parameters to be estimated;

$\underline{C}$ is a $k \times m$ ($k \geqslant m$) matrix of known constants;

then the estimator of $\underline{\theta}$ is obtained by maximizing with respect to θ the quadratic form

$$(\underline{t} - \underline{T})' \hat{\underline{V}}^{-1} (\underline{t} - \underline{T}) \qquad (3.2.5.3)$$

where $\underline{t}$ is the sample counterpart of $\underline{T}$;

$\hat{V}$ is a consistent estimator for $\underline{V}$ the covariance matrix of $\underline{t}$.

The minimization yields explicitly the estimator

$$\hat{\theta} = (\underline{C}'\underline{V}^{-1}\underline{C})^{-1}\underline{C}'\underline{V}^{-1}\underline{t}. \qquad (3.2.5.4)$$

The efficiency of this estimator depends on the elements of $\underline{t}$ and of $\underline{V}$.

An important advantage of this method is its implicit goodness of fit test. This test has the statistic

$$X^2_* = N \min_{\underline{\theta}} (\underline{t} - \underline{T})'\hat{\underline{V}}^{-1}(\underline{t} - \underline{T}) \qquad (3.2.5.5)$$

$$= N(\underline{t} - \underline{C}\,\hat{\underline{\theta}})'\underline{V}^{-1}(\underline{t} - \underline{C}\,\hat{\underline{\theta}}) \qquad (3.2.5.6)$$

which is distributed asymptotically as an exact chi-square distribution and does not require artificial grouping of data.

3.3 The P.I.N. Distribution: Maximum Likelihood Estimation

3.3.1 Partial derivatives of $\log P_x$: Method I

The objective of this section is to obtain, from the pgf $G(z)$, expressions for

$$A_x = \frac{1}{P_x}\frac{\partial}{\partial\alpha}P_x, \qquad (3.3.1.1)$$

and

$$B_x = \frac{1}{P_x}\frac{\partial}{\partial\beta}P_x, \qquad (3.3.1.2)$$

which are required in the derivation of the scoring form of the maximum likelihood equations for estimating the parameters of the Poisson Λ inverse normal distribution. These same expressions are also obtained from the probability function, P_x, in the next section (Section 3.3.2), thereby verifying the correctness of the derivation here.

Let $$a = \frac{\alpha}{\beta}, \tag{3.3.1.3}$$

$$b = \left(1 + \frac{2\beta^2}{\alpha}\right)^{\frac{1}{2}}, \tag{3.3.1.4}$$

and $$\theta = \left(\frac{2\beta^2}{\alpha}\right) \Big/ \left(1 + \frac{2\beta^2}{\alpha}\right) \tag{3.3.1.5}$$

and note that

$$\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}} = b(1 - \theta z)^{\frac{1}{2}}, \tag{3.3.1.6}$$

and

$$\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{-\frac{1}{2}} = \left(\frac{1}{b}\right)(1 - \theta z)^{-\frac{1}{2}}. \tag{3.3.1.7}$$

Note also that the pgf

$$G(z) = \exp\left\{\left(\frac{\alpha}{\beta}\right)\left(1 - \left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}}\right)\right\}$$

$$= \exp\left\{a - ab(1 - \theta z)^{\frac{1}{2}}\right\} \tag{3.3.1.8}$$

gives

$$\left(\frac{b}{\beta}\right)G'(z) = (1 - \theta z)^{-\frac{1}{2}}G(z), \tag{3.3.1.9}$$

$$\left(\frac{b}{\beta}\right)(1 - \theta z)G'(z) = (1 - \theta z)^{\frac{1}{2}}G(z) \tag{3.3.1.10}$$

The expression A_x is obtained from the pgf as follows.

$$\log G(z) \approx \left(\frac{\alpha}{\beta}\right) - \left(\frac{\alpha}{\beta}\right)\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}} \tag{3.3.1.11}$$

$$\frac{\partial}{\partial\alpha} G(z) \approx \left(\frac{1}{\beta}\right)G(z) - \left(\frac{1}{\beta}\right)\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}}G(z)$$

$$+ \left(\frac{\beta}{\beta}\right)(1 - z)\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{-\frac{1}{2}} \tag{3.3.1.12}$$

$$= \left(\frac{1}{\beta}\right)G(z) - \left(\frac{b}{\beta}\right)(1 - \theta z)^{\frac{1}{2}}G(z)$$

$$+ \left(\frac{1}{ab}\right)(1 - z)(1 - \theta z)^{-\frac{1}{2}}G(z) \tag{3.3.1.13}$$

$$= \left(\frac{1}{w_2}\frac{\partial w_2}{\partial u}\right)\left[\frac{w_2}{x!}\,w_1^{x}K_{(x-\frac{1}{2})}(w_0)\right] + (x)\left(\frac{1}{w_1}\frac{\partial w_1}{\partial u}\right)\left[\frac{w_2}{x!}\,w_1^{2}K_{(x-\frac{1}{2})}(w_0)\right]$$

$$+ (x-\tfrac{1}{2})\left(\frac{1}{w_0}\frac{\partial w_0}{\partial u}\right)\left[\frac{w_2}{x!}\,w_1^{x}K_{(x-\frac{1}{2})}(w_0)\right]$$

$$- (x+1)\left(\frac{w_0}{w_1}\right)\left(\frac{1}{w_0}\frac{\partial w_0}{\partial u}\right)\left[\frac{w_2}{(x+1)!}\,w_1^{(x+1)}K_{(x+\frac{1}{2})}(w_0)\right] \qquad (3.3.2.5)$$

$$= \left[\frac{\partial}{\partial u}\log w_2 + x\frac{\partial}{\partial u}\log w_1 + (x-\tfrac{1}{2})\frac{\partial}{\partial u}\log w_0\right]P_x$$

$$- \left[(x+1)\left(\frac{w_0}{w_1}\right)\frac{\partial}{\partial u}\log w_0\right]P_{x+1} \qquad (3.3.2.6)$$

Note that

$$w_0 = \left(\frac{\alpha}{\beta}\right)\left(1+\frac{2\beta^2}{\alpha}\right)^{\frac{1}{2}} \qquad (3.3.2.7)$$

$$w_1 = \beta\left(1+\frac{2\beta^2}{\alpha}\right)^{-\frac{1}{2}} \qquad (3.3.2.8)$$

$$w_2 = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{2}{\beta}\right)^{\frac{1}{2}}\left(1+\frac{2\beta^2}{\alpha}\right)^{\frac{1}{4}}e^{(\alpha/\beta)} \qquad (3.3.2.9)$$

i.e.

$$\frac{w_0}{w_1} = \left(\frac{\alpha}{\beta^2}\right)\left(1+\frac{2\beta^2}{\alpha}\right) = \frac{2}{\theta} \qquad (3.3.2.10$$

From these, we have

$$\frac{\partial}{\partial\alpha}\log w_0 = \left(1-\frac{\theta}{2}\right)/\alpha,$$

$$\frac{\partial}{\partial\alpha}\log w_1 = \frac{\theta}{2\alpha},$$

and
$$\frac{\partial}{\partial\alpha}\log w_2 = +[a+\tfrac{1}{2}(1-\theta)]/\alpha \qquad (3.3.2.11$$

Similarly,

$$\frac{\partial}{\partial\beta}\log w_0 = -(1-\theta)/\beta,$$

$$\frac{\partial}{\partial\beta}\log w_1 = (1-\theta)/\beta,$$

and
$$\frac{\partial}{\partial\beta}\log w_2 = -[a+\tfrac{1}{2}(1-\theta)]/\beta \qquad (3.3.2.12$$

Hence,

$$\frac{\partial}{\partial\alpha} P_x = \left\{\left[a + \tfrac{1}{2}\left(1 - \frac{\theta}{2}\right)\right]/\alpha + x \cdot \frac{\theta}{2\alpha}\right.$$

$$\left. + (x - \tfrac{1}{2})\left(1 - \frac{\theta}{2}\right)/\alpha\right\}P_x - \left(\frac{2}{\theta}\right)\left(1 - \frac{\theta}{2}\right)/\alpha(x + 1)P_{x+1} \qquad (3.3.2.13)$$

$$= \left(\frac{1}{\alpha}\right)(x + a)P_x - \left(\frac{1}{\alpha}\right)\left(\frac{2 - \theta}{\theta}\right)(x + 1)P_{x+1} \qquad (3.3.2.14)$$

$$= \left(\frac{\beta x + \alpha}{\alpha\beta}\right)P_x - \left(\frac{\beta^2 + \alpha}{\alpha\beta^2}\right)(x + 1)P_{x+1} \qquad (3.3.2.15)$$

$$= \left(\frac{1}{\alpha\beta^2}\right)\left[\beta(\beta x + \alpha)P_x - (\beta^2 + \alpha)(x + 1)P_{x+1}\right] \qquad (3.3.2.16)$$

which gives

$$A_x \equiv \frac{1}{P_x}\frac{\partial}{\partial\alpha} P_x$$

$$= \left(\frac{1}{\alpha\beta^2}\right)\left[\beta(\beta x + \alpha) - (\beta^2 + \alpha)(x + 1)(P_{x+1}/P_x)\right] \qquad (3.3.2.17)$$

as in Method I, Section 3.3.1.

Similarly,

$$\frac{\partial}{\partial\beta} P_x = \left\{- [a + \tfrac{1}{2}(1 - \theta)]/\beta + x(1 - \theta)/\beta\right.$$

$$\left. - (x - \tfrac{1}{2})(1 - \theta)/\beta\right\}P_x + \left(\frac{2}{\beta}\right)\left(\frac{1 - \theta}{\theta}\right)(x + 1)P_{x+1} \qquad (3.3.2.18)$$

$$= - \left(\frac{a}{\beta}\right)P_x + \left(\frac{\alpha}{\beta^3}\right)(x + 1)P_{x+1} \qquad (3.3.2.19)$$

$$= \left(\frac{\alpha}{\beta^3}\right)\left[(x + 1)P_{x+1} - \beta P_x\right] \qquad (3.3.2.20)$$

which gives

$$B_x \equiv \frac{1}{P_x}\frac{\partial}{\partial\beta} P_x$$

$$= \left(\frac{\alpha}{\beta^3}\right)\left[(x + 1)(P_{x+1}/P_x) - \beta\right] \qquad (3.3.2.21)$$

as in Section 3.3.1.

3.3.3 The maximum likelihood equations

With the partial derivatives of $\log P_x$ obtained in Section 3.3.1 (see also Section 3.3.2), the maximum likelihood equations for estimating the parameters of the P.I.N. distribution are derived here. The application of the equations in fitting the distribution to some data is also discussed.

The maximum likelihood equations are

$$S_\alpha \equiv \frac{\partial L}{\partial \alpha} \equiv \Sigma f_x A_x = 0 \qquad (3.3.3.1)$$

$$S_\beta \equiv \frac{\partial L}{\partial \beta} \equiv \Sigma f_x B_x = 0 \qquad (3.3.3.2)$$

where

$$A_x \equiv \frac{1}{P_x}\frac{\partial}{\partial \alpha} P_x = \left(\frac{1}{\alpha\beta^2}\right)\left[\beta(\beta x + \alpha) - (\beta^2 + \alpha)(x+1)(P_{x+1}/P_x)\right]$$

$$B_x \equiv \frac{1}{P_x}\frac{\partial}{\partial \beta} P_x = \left(\frac{\alpha}{\beta^3}\right)\left[(x+1)(P_{x+1}/P_x) - \beta\right]$$

The equations are usually left in this form if they are to be solved by the scoring method. These equations simplify to

$$\bar{x} = \beta \qquad (3.3.3.3)$$

$$\Sigma f_x (x+1)(P_{x+1}/P_x) = N\bar{x}, \qquad (3.3.3.4)$$

a form which shows that the sample mean is a maximum-likelihood estimator for the parameter β. This feature is one of the advantages of Holla's parameterization.

The scoring form of the likelihood equations has been used in fitting the P.I.N. distribution to two empirical data sets as shown in Tables IIA, IIB and IIIA, IIIB. The data in Table IIA is McGuire et al.'s (1957) European-corn-borer data to which several generalized Poisson distributions have been fitted in the literature. Sankaran (1968) has fitted the P.I.N. distribution to it by the moments-estimation-method. It is fairly short-

tailed. The other data (Table IIIA) is Bertram's (1949) filarial worms data to which Sichel (1971) has fitted the PIN distribution using the method of zero-frequency. It is long-tailed. Tables IIB and IIIB show the iteration values obtained in the process of the scoring method of the respective maximum-likelihood equations. The purpose is to provide a feel of the convergence. In each case the initial values (i.e. values at Iteration Zero) are obtained using the method of frequency at zero and the ratio of f_0 and f_1. As a check on the accuracy of the scoring method, the estimation has also been carried out by minimizing the negative loglikelihood using a locally-written direct-search no-derivatives algorithm called MAXIM. The latter method, of course, is much slower; however, on convergence it gives to many decimal places the same estimates as given by the scoring method.

Apart from data-fitting, another application of the scoring expressions in this work is the computation of asymptotic correlations of maximum-likelihood estimates (Section 3.3.4). The latter application is in fact considered the more important. Since both computations are based on the same scoring expressions, a satisfactory performance of the scoring method employed enhances confidence in the values of the asymptotic correlations discussed in Section 3.3.4.

Martin and Katti (1965) have fitted six lattice distributions to 35 sets of empirical data by maximum-likelihood. They present the value of maximum $\log_{10}$(likelihood) in each case. The distributions which they fit are the Poisson, the Poisson-with-added-zeroes, the negative binomial, the Neyman Type A, and the cases n = 2, 3 of the Poisson V binomial. The latter four are generalized Poisson distributions. However, our interest here lies not in the distribution fitted but in the respective final values of the likelihood.

The second column of Table IV contains the maximum of the six values of max $\log_{10}$(likelihood) given by Martin and Katti, and the third column

the corresponding upper-tail probability for the χ^2 test. The remaining columns are the corresponding values obtained here in the scoring-method fitting of the PIN distributions. In computing the Pearson goodness-of-fit statistic, some cells have been pooled so that, as in Martin and Katti's, no expected frequency is less than 5. In the table, an asterisk on a figure in the fourth column indicates that convergence had not been reached at the 25th iteration when the process was terminated, and dashes in the last three columns indicate that the computer program encountered inadmissible values, negative values of the parameter estimates, and hence terminated the process. The closeness of our values of the max $\log_{10}$(likelihood) to those given by Martin and Katti provides further re-assurance about the accuracy of our computer program used to implement the scoring method and hence about the correctness of the scoring expressions, A_x, B_x. These same expressions are used in the computation of the asymptotic correlations of maximum likelihood estimates of the Poisson inverse normal parameters discussed below (Section 3.3.4).

3.3.4 The asymptotic correlation of the estimates

This section is a study of the asymptotic correlation of the maximum-likelihood estimates of the parameters α and β in Holla's form of the Poisson inverse normal distribution. As detailed in Section 3.3.1, the correlation, $r_{\alpha\beta} \equiv \text{Corr}(\hat{\alpha}, \hat{\beta})$, is obtained as

$$r_{\alpha\beta} = V_{\alpha\beta}/(V_{\alpha\alpha} \cdot V_{\beta\beta})^{\frac{1}{2}} \qquad (3.3.4.1)$$

where $V_{\alpha\alpha}$, $V_{\beta\beta}$ and $V_{\alpha\beta} = V_{\beta\alpha}$ are elements of the asymptotic covariance matrix V, which is the inverse of the Fisher information matrix I.

For each set of parameter values, I is computed by the formulae given in Section 3.2.1 except that we have used N = 1. Irrespective of the value used for N, the correlation remains unchanged. Note that the elements of I are infinite series with terms corresponding to the values

$x = 0, 1, \ldots$ of the random variable, but in the computation one has to decide on a maximum value for x. We have used the same truncation rule employed by Katti and Gurland (1961) (see also Kemp (1967a)). Let $a_{ij}(x)$ denote the $(x + 1)$th term in the series comprising the (i, j)th element of I, and let $A_{ij}(x) = \sum_{r=0}^{x} a_{ij}(r)$. We compute sequentially for $x = 0, 1, \ldots$ until we reach a value of x for which

$$\sum_{i,j} a_{ij}^2(x)/A_{ij}^2(x) < 10^{-8}. \qquad (3.3.4.2)$$

The resulting values of $A_{ij}(x)$ are then taken to be the appropriate element of I. An upper limit of $x = 500$ was placed on x but was seldom needed. Katti and Gurland's work is an efficiency study. They show that the truncation rule is expected to result in computed efficiencies correct to three decimal places. For efficiency, I^{-1} and another matrix are involved but for correlation only I^{-1} is needed. Hence we expect the computed correlations to be correct to at least three decimal places with the same truncation rule.

Table V gives the correlations computed for parameter values corresponding to entries not greater than 300 in Table I. These ranges of parameter values are the important ranges from the point of view of application. It is interesting that the correlation values corresponding to them are virtually zero, indicating that the maximum likelihood estimates of the parameters α, β in Holla's form of the PIN distribution are (virtually) asymptotically uncorrelated. This result is discussed further in Section 3.3.5 below.

Sichel (1982a) gives the following explicit expression for obtaining the covariance matrix, U, of the parameters θ and γ in his own form of the PIN distribution. With U_{11}, $U_{12} = U_{21}$, and U_{22} denoting $U_{\theta\theta}$, $U_{\theta\gamma}$ and $U_{\gamma\gamma}$ respectively, the matrix U is given as

$$U = c \begin{bmatrix} \gamma(2 - \theta) & -2\theta(1 - \theta) \\ -2\theta(1 - \theta) & 4\theta(1 - \theta)^{\frac{3}{2}} R \end{bmatrix}, \qquad (3.3.4.3)$$

c being a constant, and

$$R = \left[\gamma^{-1} + (1 - \theta)^{\frac{1}{2}}\right] \left\{\left[\sum_{n=0}^{\infty} \left(P_{x+1}^2 / P_x\right) - 1\right]\right\}.$$

Let V be the covariance matrix for the maximum-likelihood estimates of the parameters α, β in Holla's form. Since the maximum likelihood estimators for the respective parameters in the various parameterizations of a given distribution are invariant, it follows that, subject to rounding errors, the matrix U obtained directly from Sichel's formula should be equal to the matrix U_* obtained by a transformation such that

$$U_* = J_1 V J_1' \qquad (3.3.4.4)$$

where J_1 is the Jacobian matrix of the transformation $T_1 : (\alpha, \beta) \to (\theta, \gamma)$. Similarly, the matrix V which we have computed directly with the scores-related formulae should, except for rounding errors, be equal to the matrix V_* obtained by a transformation such that

$$V_* = J_2 U J_2' \qquad (3.3.4.5)$$

where J_2 is the Jacobian matrix of the transformation $T_2 : (\theta, \gamma) \to (\alpha, \beta)$. We have used the former, the relation $U_* = U$ to verify the correctness of our values of the correlation $r_{\alpha\beta}$. The elements of U_* are identical to those of U to five decimal places.

3.3.5 Discussion of the results on maximum likelihood estimation

The main contribution in Section 3.3 is the result obtained computationally that maximum likelihood estimates of the parameters in Holla's form of the Poisson Λ inverse normal distribution are asymptotically uncorrelated. It is obvious that a parameterization having this property is very important from the point of view of application. Anscombe (1950) has established theoretically that the parameterization

$$G(z) = \left(1 + \frac{\beta}{\alpha} - \frac{\beta}{\alpha} z\right)^{-\alpha}, \qquad \alpha > 0, \beta > 0 \qquad (3.3.5.1)$$

of the negative binomial also has this property. A unified theory which explains this feature has been given by Huzurbazar (1950, 1956) but has received little attention in the literature. The term "orthogonal parameters" is used to describe the parameters in any parameterization having this feature.

Huzurbazar (1956) shows that when a probability distribution depending on two parameters admits joint sufficient statistics, which essentially implies that its probability (density) function is of the form

$$f_0(x;\ \alpha_1,\ \alpha_2) = \exp\{u_1(\alpha_1,\ \alpha_2)\ v_1(x) + u_2(\alpha_1,\ \alpha_2)\ v_2(x) + A(x) + B(\alpha_1,\ \alpha_2)\}, \qquad (3.3.5.2)$$

then the problem of finding orthogonal parameters is solvable. In fact, the transformation

$$\beta_1 = u_1, \quad \beta_2 = \frac{\partial B}{\partial u_2} \qquad (3.3.5.3)$$

gives β_1 and β_2 as orthogonal parameters. In the like manner, it follows that the transformation

$$\beta_1 = \frac{\partial B}{\partial u_1}, \quad \beta_2 = u_2 \qquad (3.3.5.4)$$

also yields orthogonal parameters.

Since the parameters of a mixing distribution are unaltered in a mixture, it is not surprising that, if these parameters are orthogonal in the mixing distribution they remain so in the mixture. Using Huzurbazar's result it is easy to show that the parameters α and β in the inverse normal distribution (3.1.0.1) are orthogonal. These parameters remain orthogonal in the associated Poisson mixture i.e. Holla's form of the P.I.N. distribution. However this orthogonality was first discovered computationally (Table V) without any prior knowledge about Huzurbazar's work.

The correctness of the computations was verified in four different ways. The expressions A_x, B_x on which the formulae used for computing the correlations are based were derived by two methods. A good fit of the P.I.N. distribution to two data sets was obtained using the scoring method, a method resting essentially on A_x, B_x. Also the same method produced values practically identical with Martin and Katti's values of max $\log_{10}$(likelihood) in each of 31 cases. And fourthly, the expected equivalence of the covariance matrices U_* and U was obtained.

The derivation of the maximum likelihood equations for Holla's parameterization is important in its own right and much more so in view of the above mentioned property of orthogonality which has shown the high rating of the parameterization.

3.4 The P.I.N. Distribution: Even-Points Estimation

3.4.1 The even-points estimation-method

The even-points estimation-method (Patel, 1976a) has been described in Section 3.1.4. Since the sample mean $\bar{x}$ is a maximum likelihood estimator of the parameter β, the even-points estimation for the P.I.N. distribution consists of obtaining an estimator for α involving S_e (or S_o). The estimator is derived as follows.

Note that $(S_e/N) \equiv (1/N)\sum_x f_{2x}$ is an unbiased estimator of $\theta_e \equiv \sum_x P_{2x}$. Note also that

$$\tfrac{1}{2}[G(1) + G(-1)] = \theta_e$$

implies

$$G(-1) = 2\theta_e - 1 \equiv A, \text{ say.} \qquad (3.4.1.1)$$

From the pgf which is

$$G(z) = \exp\left\{\left(\frac{\alpha}{\beta}\right)\left\{1 - \left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}}\right\}\right\} \qquad (3.4.1.2)$$

we obtain

$$G(-1) = \exp\left\{\left(\frac{\alpha}{\beta}\right)\left[1 - \left(1 + \frac{4\beta^2}{\alpha}\right)^{\frac{1}{2}}\right]\right\} \qquad (3.4.1.3)$$

i.e.

$$1 - \left(1 + \frac{4\beta^2}{\alpha}\right)^{\frac{1}{2}} \simeq \left(\frac{\beta}{\alpha}\right)\log A \qquad (3.4.1.4)$$

$$1 + \frac{4\beta^2}{\alpha} = \left[1 - \frac{\beta}{\alpha}\log A\right]^2 \qquad (3.4.1.5)$$

i.e. $$\alpha \simeq \frac{\beta}{2}(\log A)^2/(2\beta + \log A) \qquad (3.4.1.6)$$

$\therefore$ Using $\hat{\beta} = \bar{x}$, we obtain

$$\hat{\alpha}_1 = (\bar{x}/2)(\log\hat{A})^2/(\log\hat{A} + 2\bar{x}) \qquad (3.4.1.7)$$

where $$\hat{A} = 2\hat{\theta}_e - 1 \simeq 2\frac{S_e}{N} - 1$$

3.4.2 Three other types of explicit estimators

Three other sets of explicit estimators of the parameters of the P.I.N. distribution are listed here. These three are already well-known. They are the moments estimators and the sets of estimators obtained by combining the sample mean with $\log f_0$ and f_1/f_0 respectively. In each case the sample mean is used as the estimator of the parameter β. Note that it is in fact also the maximum likelihood estimator of β. The following discussion then is the derivation of estimators for the parameter α using the respective statistics.

Since the mean and variance of the distribution are given by

$$\mu_1 = \beta \qquad (3.4.2.1)$$

$$\mu_2 = \beta + \beta^3/\alpha, \qquad (3.4.2.2)$$

the moments estimators of β and α are

$$\hat{\beta} = \bar{x}, \tag{3.4.2.3}$$

$$\hat{\alpha}_2 = \bar{x}^3/(s^2 - \bar{x}) \tag{3.4.2.4}$$

where as usual $\bar{x}$ and s^2 denote the sample mean and sample variance respectively.

From the pgf, we have

$$\left[1 + \left(2\frac{\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}} = 1 - \left(\frac{\beta}{\alpha}\right)\log G(z). \tag{3.4.2.5}$$

Since

$$P_0 = G(0), \tag{3.4.2.6}$$

$$\left[1 + \frac{2\beta^2}{\alpha}\right]^{\frac{1}{2}} = 1 - \frac{\beta}{\alpha}\log P_0 \tag{3.4.2.7}$$

$$\alpha(\alpha + 2\beta^2) = (\alpha - \beta\log P_0)^2 \tag{3.4.2.8}$$

i.e.

$$\alpha = \left(\frac{\beta}{2}\right)(\log P_0)^2/(\beta + \log P_0) \tag{3.4.2.9}$$

$$\therefore \hat{\alpha}_3 = (\bar{x}/2)\left(\log\frac{f_0}{N}\right)^2\Big/\left(\bar{x} + \log\frac{f_0}{N}\right) \tag{3.4.2.10}$$

Note that this takes the same form as the even-points estimator, with $\hat{P}_0$ corresponding to $\hat{A} \equiv 2\hat{\theta}_e - 1$.

From the pgf we also have

$$\log G(z) = \left(\frac{\alpha}{\beta}\right) - \left(\frac{\alpha}{\beta}\right)\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{\frac{1}{2}} \tag{3.4.2.11}$$

$$G'(z)/G(z) = \beta\left[1 + \left(\frac{2\beta^2}{\alpha}\right)(1 - z)\right]^{-\frac{1}{2}} \tag{3.4.2.12}$$

i.e.

$$\left(\frac{1}{\beta}\right)\left(\frac{P_1}{P_0}\right) = \left[1 + \left(\frac{2\beta^2}{\alpha}\right)\right]^{-\frac{1}{2}} \tag{3.4.2.13}$$

$$\beta^2\left(\frac{P_0}{P_1}\right) = 1 + \left(\frac{2\beta^2}{\alpha}\right) \tag{3.4.2.14}$$

$$\alpha = 2\beta^2\Big/\left[\beta^2\left(\frac{P_0}{P_1}\right) - 1\right] \tag{3.4.2.15}$$

$$\therefore \quad \bar{\alpha}_4 = 2\bar{x}^2 \Big/ \left[\bar{x}^2\left(\frac{f_0}{f_1}\right) - 1\right]. \qquad (3.4.2.16)$$

3.4.3 Asymptotic efficiencies of the four sets of explicit estimators

Sichel (1982a) has computed the asymptotic efficiencies of the estimators obtained for his parameters θ and γ by the moments method, the methods of zero frequency and of the ratio of f_0 and f_1. Since the estimators which these methods produce are not invariant, the efficiencies of the analogous estimators discussed here cannot be derived from Sichel's results. Expressions for computing these efficiencies, and that for the newly proposed even-points estimators are derived here.

The joint asymptotic efficiency, E, of the estimators $a_1, \ldots, a_m$ of a given set of parameters is defined as

$$E = |I^{-1}|/|W| \qquad (3.4.3.1)$$

where W is the covariance matrix of $a_1, \ldots, a_m$, I is the information matrix of the corresponding maximum likelihood estimators, and the notation $|A|$ denotes the determinant of the matrix A. Suppose $t_1, \ldots, t_k$ are the statistics used in obtaining each of the a_i's, i.e. $a_i \equiv a_i(t_1, \ldots, t_k)$, $i = 1, \ldots, m$. The matrix W is given by

$$W = JHJ' \qquad (3.4.3.2)$$

(see e.g. Kemp (1967a)) where H is the covariance matrix of $t_1, \ldots, t_k$ and J is the Jacobian matrix of the transformation $a_i \equiv a_i(t_1, \ldots, t_k)$ evaluated at the point $(E(t_1), \ldots, E(t_k))$ i.e. at the expected values of the statistics.

Let ${}_iW, {}_iH, {}_iJ$, $i = 1, 2, 3, 4$ be such that the subscripts $i = 1, 2, 3, 4$ correspond to the methods of even-points, moments, zero frequency and of ratio f_1/f_0 respectively. It is convenient to refer

to these methods simply as method $\hat{\alpha}(S_e, N\bar{x})$, method $\hat{\alpha}(\bar{x}, s^2)$, method $\hat{\alpha}(\bar{x}, \log f_0)$, and method $\hat{\alpha}(\bar{x}, f_1/f_0)$ respectively to show the arrangement of the elements of the matrices and to show what have been used as the statistics t_1, t_2.

The matrices ${}_iH$, $i = 2, 3, 4$ have already been given in Section 3.1.1 as the matrices V_1, V_2, V_3 respectively. The elements of the matrix ${}_1H \equiv ({}_1H_{ij})$, the covariance matrix of the even-points statistics S_e and $T \equiv N\bar{x}$, are given by

$$
\begin{aligned}
{}_1H_{11} &= \frac{N}{4} - \frac{N}{4}\exp\left\{2\left(\frac{\alpha}{\beta}\right)\left[1 - \left(1 + \frac{4\beta^2}{\alpha}\right)^{1/2}\right]\right\} \\
{}_1H_{12} &= {}_1H_{21} = (N\beta/2)\left[\left(1 + \frac{\beta^2}{\alpha}\right)^{-1/2} - 1\right] \times \exp\left\{\left(\frac{\alpha}{\beta}\right)\left[1 - \left(1 + \frac{4\beta^2}{\alpha}\right)^{1/2}\right]\right\} \\
{}_1H_{22} &= (N\beta)\left(1 + \frac{\beta^2}{\alpha}\right)
\end{aligned}
\qquad (3.4.3.3)
$$

These expressions have been obtained using the formula derived in Section 3.1.4.

The Jacobian matrix ${}_1J$ corresponding to the even-points method $\hat{\alpha}(S_e, N\bar{x})$ is

$$
{}_1J = \begin{bmatrix} ad(2 - d) & (c - 2a)(c/b) \\ 0 & 1 \end{bmatrix} \qquad (3.4.3.4)
$$

where $\nu = (\alpha/\beta)\left[1 - \left(1 + \frac{4\beta^2}{\alpha}\right)^{1/2}\right]$,

$a = 2\beta$, $b = 2(a + \nu)$,

$c = \nu^2$, $d = (\nu/b)\exp(-\nu)$

The Jacobian matrix, ${}_2J \equiv ({}_2J_{ij})$ for the moments method $\hat{\alpha}(\bar{x}, s^2)$ is such that

$$
\begin{aligned}
{}_2J_{11} &= \bar{x}^2(3s^2 - 2\bar{x})/(s^2 - \bar{x})^2, \\
{}_2J_{12} &= -\bar{x}^3/(s^2 - \bar{x})^2, \\
{}_2J_{21} &= 1, \quad {}_2J_{22} = 0.
\end{aligned}
\qquad (3.4.3.5)
$$

Let $u = \log(f_0/N)$. Then corresponding to the method $\hat{\alpha}(\bar{x}, \log f_0)$ we have the Jacobian matrix ${}_3J \equiv ({}_3J_{ij})$ given by

$$
\begin{aligned}
{}_3J_{11} &= (u^3/2)/(\bar{x} + u)^3, \\
{}_3J_{12} &= (u\bar{x}/2)(2\bar{x} + u)/(\bar{x} + u)^2, \\
{}_3J_{21} &= 1, \quad {}_3J_{22} = 0.
\end{aligned} \tag{3.4.3.6}
$$

Similarly, with $r_{10} = f_1/f_0$, we have, for the method $\hat{\alpha}(\bar{x}, r_{10})$, the Jacobian matrix ${}_4J \equiv ({}_4J_{ij})$ where

$$
\begin{aligned}
{}_4J_{11} &= -4\bar{x}/[(\bar{x}^2/r_{10}^2) - 1]^2, \\
{}_4J_{12} &= 4\bar{x}^4/(\bar{x}^2 - r_{10}^2)^2, \\
{}_4J_{21} &= 1, \quad {}_4J_{22} = 0.
\end{aligned} \tag{3.4.3.7}
$$

3.5 Summary on "The PIN Distribution"

While Chapter 1 discusses discrete distribution in general, and Chapter 2 considers generalized Poisson distribution - the family of discrete distributions of interest in this thesis, this chapter dwells mainly on one particular GPD, the Poisson Λ inverse normal distribution. Attention is devoted mainly to the methods for estimating its parameters.

Properties of the distribution are briefly considered. In particular it is shown that it is not only a compound Poisson distribution but also a generalized Poisson distribution. The pgf of its cluster-size distribution is given.

After a brief outline of the various parameter-estimation-methods used for generalized Poisson distributions, maximum-likelihood estimation of the PIN distribution is considered and some explicit estimators of the parameters are also discussed. One of the two commonly used forms of the distribution is shown to be preferable because it consists of orthogonal parameters. This means that maximum likelihood estimates of the parameters in this form are asymptotically uncorrelated. For this form also, the sample mean is a maximum likelihood estimator of the one of the parameters.

A new estimator, the even-points estimator, is presented for the other parameter. Expressions are given for computing the asymptotic efficiencies of these estimators and of the other explicit estimators.

CHAPTER 4

The Problem of Distribution Selection

CHAPTER 4

THE PROBLEM OF DISTRIBUTION SELECTION

4.1 Introduction

Suppose a phenomenon under investigation has been represented in a model which requires a generalized Poisson distribution to study it. Assume that some data are already available from a pilot study or even the main study itself. A number of GPDs have been fitted to the data using an efficient estimation method such as the maximum likelihood method. A goodness-of-fit test has indicated some of these distributions as giving a good fit. The problem is how to select from a given set of plausible GPDs the one most appropriate for the situation.

In some studies, the model would have been derived systematically in such a way that it points to one particular distribution. An example is the manner in which Neyman (1939) derives the Neyman Type A distribution for the study of the spatial distribution of the eggs and larvae of insects. More often than not, all that is known at the beginning of such a study is that the members of the population are in randomly distributed clusters, in which case several generalized Poisson distributions may be plausible.

The problem of distribution selection has received relatively little attention in the literature although it is of considerable theoretical and practical importance. This problem is briefly discussed here in the context of future work.

4.2 Distribution Selection, Model Discrimination, and the Global Superiority of a Distribution

It is important to distinguish between three separate problems: that of selecting the distribution best fitting a data set, that of model discrimination, and that of ranking of distributions by their global importance.

Whether a given phenomenon is a result of true contagion or otherwise pertains to model discrimination. The fact that the selection of a theoretical distribution providing the best fit to a given data set is often involved in such a situation is only incidental. Froggatt (1970) has given a discussion of model discrimination in the study of accident proneness. Patil and Stiteler (1974) have similarly considered the problem as it relates to the study of the spatial distributions of plants and insects. On the other hand, consider a phenomenon for which a model has been established. For example, it is generally assumed that the spatial pattern of the distribution of the European corn-borer larvae exhibits true contagion. The problem of deciding which among a given set of distributions such as the negative binomial, the Neyman Type A, the Hermite, the Poisson A inverse normal etc. fits e.g. McGuire et al.'s (1957) data best is an example of distribution selection.

Unlike the other two, the third type of problem seems to relate directly to theoretical rather than practical considerations. Members of a given set of theoretical distributions are ranked according to a given criterion, such as skewness or "tail-length", which is purported to be an index of global practical importance. Anscombe (1950, p.363) proposes the ranking of distributions by the quantity

$$R_1 = \mu k_{(3)} / (\sigma^2 - \mu)^2 \qquad (4.2.1)$$

or, alternatively,

$$R_2 = \mu^2 k_{(4)} / (\sigma^2 - \mu)^3 \qquad (4.2.2)$$

where $k_{(3)}$ and $k_{(4)}$ denote the third and fourth factorial cumulants, a method which has been used by Katti and Gurland (1961) and Katti and Rao (1970) to support the claims that their respective distributions, the Poisson V negative binomial and the log-zero-Poisson, are superior to some other distributions. The quantities R_1 and R_2 are known as indices of flexibility. A third index, R_3, proposed by Ord (1967a) is given by

$$R_3 = \mu\mu_3/\mu_2^3 \qquad (4.2.3)$$

where μ_r denotes the rth central moment.

An immediate objection to the use of these indices as criteria for determining global superiority is that, except for a few special cases, they are not independent of the parameters of the respective distributions.

4.3 Gurland's Graphical Method

Gurland (1965) and Hinz and Gurland (1967) have proposed a graphical method for discriminating between four particular generalized Poisson distributions viz. the negative binomial (Poisson(λ) V logarithmic(θ)) the Neyman Type A (Poisson(λ) V Poisson(α)), the Poisson(λ) V binomial(n, θ), and the Poisson(λ) V negative binomial(α, θ). The method consists simply in plotting the ratio

$$T_x = k_{(x+1)}/k_{(x)} \qquad (4.3.1)$$

against x, where $k_{(r)}$ is the rth factorial cumulant.

Corresponding to each of the four distributions a straight line

$$y = bx + c \qquad (4.3.2)$$

is expected. It is suggested that the distribution appropriate for the data is

D_1 : negative binomial, if $b > 0$ and $c = 0$;

D_2 : Poisson V negative binomial, if $b > 0$ and $c > 0$;

D_3 : Neyman Type A, if $b = 0$ and $c > 0$;

D_4 : Poisson V binomial, if $b < 0$ and $c < 0$,

since, respectively,

$$_1T_x \equiv y = \left(\frac{\theta}{1-\theta}\right)x \qquad (4.3.3)$$

$$_2T_x \equiv y = \left(\frac{\theta}{1-\theta}\right)x + \left(\frac{\theta\alpha}{1-\theta}\right) \qquad (4.3.4)$$

$$_3T_x \equiv y = \alpha \qquad (4.3.5)$$

and

$$_4T_x \equiv y = -\theta x - n\theta \qquad (4.3.6)$$

Two main difficulties arise in the use of this criterion. The third and higher factorial cumulants (like other higher moments) are subject to considerable variability and hence, even in a case when one of these distributions is indeed appropriate for the given data the expected straight line would be very difficult to obtain. Secondly, the criterion, even at its best, is appropriate only for making a choice among these four particular distributions and not among all GPDs. It is difficult to imagine a practical situation in which these and only these four are to be entertained to the exclusion of all other GPDs. Furthermore, except for the case y = a corresponding to the Neyman Type A, it is doubtful that the respective forms of y = bx + c uniquely determine these particular distributions among all generalized Poisson distributions.

4.4 Ranking by Likelihood

A criterion for determining the best of well-fitting distributions is to rank by the likelihood or equivalently by the loglikelihood, L, given by

$$L = \text{En } \log p_i \qquad (4.4.1)$$

This is the criterion favoured by Katti and Rao (1970), for example.

Reviewing the cases where the values of L or its equivalents are published in the literature of discrete distributions, it seems that such a quantity pertains to the quality of the parameter estimation process rather than to the theoretical distribution being fitted. For example, the tendency is for several GPDs fitted to a set of overdispersed data by the maximum likelihood method to give essentially the same value of L (e.g. Martin and Katti (1965)). Katti and Sly (1965), in a work primarily on model discrimination but which also involves distribution selection, give the fits of three different GPDs to three sets of data. For each data set, while the probabilities of the computed Pearson

chi-squared statistic differ widely, the values of L are essentially the same. It is interesting that in one case the probabilities are respectively 0.04, 0.23, 0.65 while the corresponding values of the loglikelihood L are -1987.94, -1991.89, and -1989.11, showing a distribution which does not give a good fit to the data having virtually the same value of L as distributions which fit the data well. It appears that the worth of the quantity L or any of its equivalents as a criterion for distribution selection requires further attention.

4.5 Goodness-of-Fit Statistics and Criteria for Distribution Selection

The goodness-of-fit test statistics are designed primarily for determining whether or not a particular distribution gives a good fit of the data. Each of the distributions in the set being considered in this chapter is assumed to have been adjudged as satisfactory on the strength of a goodness-of-fit test. A method for determining the best distribution in this set consists of ranking the distribution according to the computed values of the statistic or of the probabilities of these computed values (see e.g. Sichel (1971) and Shaban (1981)).

The well known likelihood ratio goodness-of-fit statistic is given by

$$Y = 2 \log R \tag{4.5.1}$$

with

$$R = \prod_i \left(\frac{n_i}{N p_i} \right)^{n_i} \tag{4.5.2}$$

where n_i is the frequency of the ith variate, p_i is the probability of this variate, and N the sample size. This statistic asymptotically follows the χ^2 distribution. Lindsey (1974) has suggested the quantity

$$Y_* = \log R \tag{4.5.3}$$

as a criterion for distribution selection. The distributions under consideration are ranked according to their values of Y_*, which is equivalent to using the goodness-of-fit test statistic, Y.

A criticism of the use of the values of a goodness-of-fit statistic or their corresponding probabilities for determining the best in a given set of distributions pertains to the issue of comparability. First consider the use of the values of the statistics. The pooling of frequencies is often required, especially at the tails. Generally, the amount of the required pooling differs from distribution to distribution and hence the numbers of degrees of freedom would differ. The computed values of the statistic would therefore not be comparable. On the other hand, if their corresponding probabilities are obtained, the difficulty with the degrees of freedom is taken care of but that however does not mean that ranking by probabilities is free from difficulties. The value of the statistic, and consequently of the corresponding probability, depends on the minimum expected frequency to which the pooling is done. It is possible for the ordering of the distributions to change if another minimum expected frequency is adopted. Although 5 is the commonest, 3 and 1 are sometimes also used. In a case when these produce different orderings of the distribution, which of the orderings is to be accepted?

CHAPTER 5

POSTVIEW AND PROSPECTS

A major objective of this thesis is to highlight some particular aspects of the class of generalized Poisson distributions which have special bearing on applications. A large body of literature has been reviewed and some new results are also given.

As a general background to the discussion of this class, the major aspects of the subject of discrete distributions are reviewed in Chapter 1. Although certain topics such as characterization are excluded, the chapter provides a broad introduction to the subject.

In Chapter 2, the structure, properties, and applications of generalized Poisson distributions are considered extensively. The discussion is pivoted on the property of infinite divisibility. Much attention has been devoted here to distributions which are simultaneously generalized Poisson distributions and compound Poisson distributions. Compound Poisson distributions have their genesis in the models of heterogeneity, the so-called models of apparent contagion, while generalized Poisson distributions are usually associated with the clustering models, the models of true contagion. Although the dual faces of the negative binomial in relation to both true and apparent contagion have been noticed long ago and have been mentioned occasionally in the literature, a detailed study of the type of distributions simultaneously exhibiting these features is lacking.

The properties of logconvexity and of being completely monotone (i.e. being a mixture of the geometric $h(z) = (1 - \theta)/(1 - \theta z)$) have been used to identify some discrete distributions as being infinitely divisible and hence of being generalized Poisson distributions. However the determination of the cluster-size distributions of most of these

distributions is yet to be carried out, e.g. that of the Waring distribution. Also, it appears that the relationship of the Waring to the generalized Waring is analogous to that of the geometric to the negative binomial. While the geometric is logconvex and hence infinitely divisible, the negative binomial is not logconvex although it is infinitely divisible. The infinite divisibility of the generalized Waring is being investigated. It is known that the distribution is a mixture of the negative binomial on the non-exponent parameter; but that is not a sufficient condition for infinite divisibility.

The properties and applications of the newly derived reciprocal-gamma mixture of the Poisson distribution are being investigated further. Maximum likelihood estimation of its parameters is also being studied. An interesting feature of the distribution is that it arises from a continuous distribution which possesses no moments. Of mixtures of the Poisson by continuous distributions the ones discussed extensively in the literature tend to be those for which the continuous mixing distributions have moment generating functions.

In most of the distributions considered, a means of computing the probabilities has been mentioned. A distribution known to have desirable properties but whose probabilities are difficult to compute is unlikely to be readily adopted by practitioners. In particular, simple formulae for the computation of the probabilities of the Poisson Λ lognormal are being sought.

It is widely believed that, due to rounding errors, probabilities computed by recursive formulae are likely to be inaccurate and should be viewed with great reservation. There is little evidence in the literature to substantiate this belief - it may well be a carry-over from the pre-computer era. A systematization of the process for computing the probabilities of, for example, the Poisson by the so-called direct formula either by computer or otherwise would indicate that some

CHAPTER 5

Postview and Prospects

recursiveness is implicitly involved. Besides, the probability functions of most GPDs do not exist in closed forms and often the best means of computing the probabilities is by recursive formulae. The performance of recursive formulae for the probabilities of the negative binomial and some other common distributions is being compared with those of the direct means. The preliminary results already obtained are encouraging.

Concerning the aspect of parameter estimation the thesis focuses on one particular distribution, the Poisson Λ inverse normal distribution. One of the two commonly used forms of the distribution is shown to be preferable because it is such that the maximum likelihood estimates of the parameters in the form are asymptotically uncorrelated. This result has prompted the desire to know which of the prameterizations of other common distributions have this property.

Certain explicit estimators for the parameters of the Poisson Λ inverse normal have been derived. Expressions for obtaining their asymptotic efficiencies have also been presented. These expressions are being programmed for the actual computation of the efficiencies.

In the application of generalized Poisson distributions the phenomenon under study is often such that it is not always possible to derive a single specific GPD from the underlying chance mechanisms besides knowing that the required distribution is a generalized Poisson distribution. Usually many GPDs are fitted to the data and whether or not the fit of each of these distributions is satisfactory is decided by a goodness-of-fit test. How to choose the best out of the set giving satisfactory fit is yet unresolved. This problem has been briefly outlined in Chapter 4. Work in this area is also in progress.

TABLES

TABLE I

DETERMINGING USEFUL PARAMETER RANGES FOR THE P.I.N. DISTRIBUTION**

		1	2	3	4	5	6	7	8	9	10	11	12	13
	α \ β	0.001	0.005	0.01	0.05	0.1	0.5	1.0	5.0	10	50	100	500	1000
1	0.001	3	3	3	7	16	103	192	451	530	613	625	635	637
2	0.005	3	3	3	4	8	71	178	954	1526	2646	2893	3122	3153
3	0.1	3	3	3	3	6	52	141	1015	1907	4504	5291	6122	6243
4	0.05	3	3	3	3	4	22	63	698	1765	9532	15251	26452	28917
5	0.1	3	3	3	3	3	15	41	508	1394	10134	19062	45025	*
6	0.5	3	3	3	3	3	7	16	198	608	6966	17639	*	*
7	1.0	3	3	3	3	3	6	11	126	394	5062	13930	*	*
8	5.0	3	3	3	3	3	5	7	45	135	1962	6060	*	*
9	10	3	3	3	3	3	5	7	31	87	1239	3921	*	*
10	50	3	3	3	3	3	5	6	18	38	421	1329	19597	*
11	100	3	3	3	3	3	5	6	16	30	275	839	12365	39192
12	500	3	3	3	3	3	5	6	14	24	128	335	4181	13261
13	1000	3	3	3	3	3	5	6	14	23	104	250	2714	8358

** Notes: 1. Each entry is minimum x such that $\Sigma P_x \geq 0.9999$.

2. Entries greater than 300 indicate parameter values rarely encountered in practical situations.

3. Asterisks denote entries greater than 50,000.

TABLE IIA

The P.I.N. Distribution fitted to McGuire's data

Observed and Expected Frequencies

x	f_x	P.I.N. (M-L)
0	423	425.922
1	414	412.617
2	253	246.354
3	117	120.055
4	53	53.084
5	22	22.381
6	4	9.226
7	5	3.767
8	3	1.534
9	2	0.625
> 9	0	0.434
Total	1296	1296.000

$\hat{\alpha}$ = 4.16471 ± 0.71033

$\hat{\beta}$ = 1.30710 ± 0.03805

Prob ($\chi^2 \leqslant$ Pearson's chi-square) = 0.825 with min $NP_x \geqslant 1$

= 0.722 with min $NP_x \geqslant 5$

Asymptotic correlation, Corr($\hat{\alpha}$, $\hat{\beta}$) = 0.08221

TABLE IIB

The P.I.N. Distribution fitted to Bertram's data

Observed and Expected Frequencies

x	f_x	P.I.N. (M-L)
0	1155	1153.258
1	553	555.294
2	265	267.779
3	150	150.783
4	98	96.201
5	70	66.793
6	48	49.131
7	36	37.656
8	28	29.756
9	27	24.073
10	15	19.841
11	13	16.602
12	21	14.067
13	8	12.044
14	11	10.405
15	5	9.059
16	9	7.940
17	9	7.000
18	9	6.204
19	8	5.524

x	f_x	P.I.N. (M-L)
20	5	4.940
21	2	4.433
22	4	3.993
23	5	3.607
24	3	3.268
25	5	2.969
26	1	2.703
27	0	2.467
28	2	2.256
29	0	2.067
30	5	1.898
31	6	1.745
32	4	1.607
33	2	1.482
34	1	1.368
35	2	1.265
36	4	1.172
37	1	1.086
38	0	1.008
39	0	0.936

TABLE IIB - CONTINUED

x	f_x	P.I.N. (M-L)
40	0	0.871
41	2	0.710
42	0	0.755
43	0	0.704
44	0	0.657
45	1	0.613
46	0	0.573
47	0	0.536
48	0	0.502
49	0	0.470
50	0	0.440
51	1	0.413
52	0	0.387
53	0	0.363
54	0	0.341
55	0	0.321
56	0	0.301
57	0	0.284
58	0	0.267
59	1	0.251

x	f_x	P.I.N. (M-L)
60	0	0.237
61	0	0.223
62	1	0.210
63	1	0.198
64	1	0.187
65	0	0.176
66	0	0.167
67	1	0.157
68	0	0.149
69	0	0.140
70	0	0.133
71	0	0.125
72	0	0.119
73	0	0.112
74	0	0.106
75	0	0.101
76	0	0.095
77	0	0.090
78	1	0.085
> 78	0	1.653
TOTAL	2600	2600.000

$\hat{\alpha}\hat{}$ = 0.48005 ± 0.02807

$\hat{\beta}$ = 2.60808 ± 0.13503

Prob ($\chi^2 \leqslant$ Pearson's chi-square) = 0.080 with min $NP_x \geqslant 1$

= 0.031 with min $NP_x \geqslant 5$

Asymptotic Corr($\hat{\alpha}$, $\hat{\beta}$) = -0.03294

TABLE IIIA

The P.I.N. Distribution fitted to McGuire's data

The Scoring Method Iteration Points

Iteration No.	$\hat{\alpha}$	$\hat{\beta}$
0	4. 360 675 282	1. 307 098 765
1	4. 134 694 298	1. 307 498 992
2	4. 168 186 092	1. 307 017 607
3	4. 164 246 790	1. 307 108 739
4	4. 164 771 217	1. 307 097 450
5	4. 164 702 673	1. 307 098 937
6	4. 164 711 651	1. 307 098 743
7	4. 164 710 475	1. 307 098 768
8	4. 164 710 629	1. 307 098 765

Loglikelihood (at last iteration point) = -1986.04964

By Direct Search algorithm:

$\hat{\alpha}$ = 4. 164 707 736

$\hat{\beta}$ = 1. 307 098 765

Max loglikelihood = -1986.0196

TABLE IIIB

The P.I.N. Distribution fitted to Bertram's data

The Scoring Method Iteration Points

Iteration No.	$\hat{\alpha}$	$\hat{\beta}$
0	0. 474 465 735	2. 608 076 923
1	0. 480 271 246	2. 605 428 113
2	0. 479 982 380	2. 608 714 820
3	0. 480 064 308	2. 607 916 573
4	0. 480 043 951	2. 608 117 041
5	0. 480 049 059	2. 608 066 874
6	0. 480 047 780	2. 608 079 439
7	0. 480 048 100	2. 608 076 293
8	0. 480 048 020	2. 608 077 081
9	0. 480 048 040	2. 608 076 884
10	0. 480 048 035	2. 608 076 933

Loglikelihood (at last iteration point) = -5040.0850

By Direct Search Algorithm:

$\hat{\alpha}$ = 0. 480 047 809

$\hat{\beta}$ = 2. 608 075 602

Max loglikelihood = -5040.0850

TABLE IV

COMPARING THE P.I.N. FITTING

WITH MARTIN AND KATTI'S FITTING†

Data No.	Martin & Katti's		P.I.N. Distribution		
	$-\text{Log}_{10} L$	Upper Prob (Min Freq = 5)	$-\text{Log}_{10} L$	Upper-tail Prob (Min Freq = 5)	Upper-tail Prob (Min Freq = 1)
1	989.03	0.415	986.24	0.123	0.127
2	227.24	0.548	228.22	0.325	0.511
3	133.94	0.533	134.14	0.802	0.848
4	177.97	0.546	179.01	0.391	0.536
5	120.09	0.567	120.34	0.974	0.955
6	129.75	0.543	129.57	0.572	0.793
7	117.13	0.973	121.52	0.571	0.884
8	86.99	0.880	87.08*	0.998	0.870
9	87.84	0.086	86.50	0.285	0.478
10	30.24	0.799	29.54*	0.922	0.982
11	57.01	0.517	59.24*	0.000	0.000
12	2699.76	0.001	2701.21	0.002	0.000
13	2517.52	0.094	2520.46	0.009	0.013
14	507.59	0.939	697.06*	0.000	0.000
15	862.99	0.919	862.53	0.722	0.825
16	352.29	0.237	352.02	0.426	0.593
17	477.04	0.705	478.68	0.237	0.295
18	240.38	0.511	240.57	0.854	0.861
19	265.51	0.791	266.07	0.841	0.739
20	154.19	0.760	154.83	0.808	0.759

/continued .....

TABLE IV - CONTINUED

Data No.	Martin & Katti's		P.I.N. Distribution		
	$-\text{Log}_{10} L$	Upper Prob (Min Freq = 5)	$-\text{Log}_{10} L$	Upper-tail Prob (Min Freq = 5)	Upper-tail Prob (Min Freq = 1)
21	96.60	0.477	98.92*	0.002	0.007
22	344.50	0.546	342.00	0.881	0.956
23	193.93	0.189	193.93	0.673	0.767
24	165.91	0.570	165.87	0.988	0.967
25	996.80	0.730	1000.29	0.023	0.036
26	537.63	0.456	537.86	0.854	0.941
27	91.85	0.818	92.64*	0.813	0.645
28	184.28	0.983	188.04*	0.952	0.819
29	193.66	0.097	194.60	0.099	0.003
30	165.33	0.665	167.90*	0.064	0.026
31	123.44	0.268	124.33*	0.237	0.187
32	134.86	0.995	-	-	-
33	113.71	0.660	-	-	-
34	68.53	0.084	-	-	-
35	94.44	0.017	-	-	-

†Notes:

1. The 2nd column gives the maximum of M & K's 6 values of max $\log_{10}$(likelihood).
2. In the 4th column, an asterisk indicates that convergence has not been attained at the 25th iteration when the process was terminated.
3. Dashes in the last three columns indicate that the iterative process was terminated because negative values of parameter estimates were encountered.

TABLE V

CORRELATION Corr($\hat{\alpha}$, $\hat{\beta}$) OF MAXIMUM LIKELIHOOD ESTIMATES FOR THE P.I.N. DISTRIBUTION

	α \ β	1 0.001	2 0.005	3 0.01	4 0.05	5 0.1	6 0.5	7 1.0	8 5.0	9 10	10 50	11 100
1	0.001	.01578	.09676	.17476	.16448	.08045	.00893	.00323	*	*	*	*
2	0.005	.00408	.03500	.07833	.22822	.17625	.02771	.01030	*	*	*	*
3	0.01	.00213	.02032	.04901	.20936	.21051	.04393	.01688	*	*	*	*
4	0.05	.00044	.00476	.01287	.10155	.18150	.11141	.04899	*	*	*	*
5	0.1	.00022	.00244	.00673	.06165	.13153	.14874	.07364	*	*	*	*
6	0.5	.00004	.00050	.00140	.01495	.03948	.17388	.14528	.20353	*	*	*
7	1.0	.00002	.00025	.00070	.00769	.02098	.13782	.15909	.03557	*	*	*
8	5.0	.00002	.00005	.00014	.00157	.00441	.04408	.09543	.07735	.03434	*	*
9	10	.00005	.00003	.00007	.00079	.00222	.02348	.05788	.09407	.04786	*	*
10	50	-.00025	.00003	.00002	.00016	.00045	.00494	.01359	.07986	.07573	*	*
11	100	.00038	.00005	.00002	.00008	.00022	.00248	.00693	.05463	.07316	.01747	*
12	500	.12330	.00185	.00016	.00002	.00005	.00050	.00141	.01461	.03356	.03244	*
13	1000	.07234	.00604	.00001	.00001	.00002	.00025	.00071	.00760	.01926	.03690	.01880

REFERENCES

REFERENCES

1. Adelson, R.M. (1977). Compound Poisson distributions. Operational Research Quarterly, 17, 73-75.
2. Ahmed, M.S. (1971). Discrete bivariate distributions with given marginals. The Libyan Journal of Science, 1, 104-113.
3. Anscombe, F.J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. Biometrika, 37, 358-382.
4. Arbous, A.C. and Kerrich, J.E. (1951). Accident statistics and the concept of accident proneness. Biometrics, 7, 340-432.
5. Aitken, A.C. and Gonin, H.T. (1935). On fourfold sampling with and without replacement. Proceedings of the Royal Society, Edinburgh, 55, 114-125.
6. Atkinson, A.C. (1979). The computer generation of Poisson random variables. Applied Statistics, 28, 29-35, 260-263.
7. Atkinson, A.C. and Yeh, L. (1982). Inference with the generalized inverse Gaussian-Poisson distribution. Journal of American Statistical Association, 77, 153-158.
8. Bardwell, G.E. and Crow, E.L. (1964). A two-parameter family of hyper-Poisson distributions. Journal of American Statistical Association, 59, 133-141.
9. Barndorff-Nielson, O. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, 38, 309-311.
10. Barnett, A.J. (1979). The tail length of discrete distributions. Ph.D. Thesis, Statistics, University of Bradford, England.
11. Bartko, J.J. (1961). The negative binomial distribution: a review of properties and applications. Virginia Journal of Science, 12, 115-121.

12. Baxter, G. and Shapiro, J.M. (1960). On bounded infinitely divisible random variables. Sankhyā, 22, 253-260.

13. Beall, G. and Rescia, R.R. (1953). A generalization of Neyman's contagious distributions. Biometrics, 9, 354-386.

14. Berkson, J. (1980). Minimum chi-square, not maximum likelihood! (with discussion). Annals of Statistics, 8, 457-487.

15. Bertram, D.S. (1949). Studies on transmission of rat filiariasis. Annals of Tropical Medicine and Parasitology, 43, 313-332.

16. Bhattacharya, S.K. (1966). Confluent hypergeometric distributions of discrete and continuous type with applications to accident proneness. Calcutta Statistical Association Bulletin, 15, 20-51.

17. Bhattacharya, S.K. and Holla, M.S. (1965). On a discrete distribution with special reference to the theory of accident proneness. Journal of American Statistical Association, 60, 1060-1066.

18. Bissinger, B.H. (1965). A type-resisting distribution generated from considerations of an inventory decision model. In Patil, G.P. (ed., 1965, pp.15-17).

19. Blischke, W.R. (1965). Mixtures of discrete distributions. In Patil, G.P. (ed., 1965, pp.351-372).

20. Blischke, W.R. (1964). Estimating the parameters of mixtures of binomial distributions. Journal of American Statistical Association, 59, 510-528.

21. Bliss, C.I. (1965). An analysis of some insect trap records. In Patil, G.P. (ed., 1965, pp.385-397).

22. Bliss, C.I. and Fisher, R.A. (1953). Fitting the negative binomial distribution to biological data; the efficient fitting of the negative binomial. Biometrics, 9, 176-200.

23. Boes, D.C. (1966). On the estimation of mixing distributions. Annals of Mathematical Statistics, 37, 177-188.

24. Bondesson, L. (1978). On the infinite divisibility of powers of a gamma variable. Scandinavian Actuarial Journal, 48-61.

25. Boswell, M.T. and Patil, G.P. (1970). Chance mechanisms generating the negative binomial distribution. In Patil, G.P. (ed., 1970, Vol. 1, pp.3-22).

26. Boswell, M.T. and Patil, G.P. (1972). Certain mechanisms related to birth and death processes which have classical discrete distributions as special cases. In Stochastic Point Processes, ed. P.A.W. Lewis; Wiley, New York; pp.285-298.

27. Boswell, M.T., Ord, J.K. and Patil, G.P. (1979). Chance mechanisms underlying univariate distributions. In Ord, J.K. et al. (eds., 1979, pp.3-156).

28. Brown, S. and Holgate, P. (1971). Table of the Poisson lognormal distribution. Sankhyā, Series B, 33, 235-248.

29. Bulmer, M.G. (1974). On fitting the Poisson lognormal distribution to species aboundance data. Biometrika, 30, 101-110.

30. Cacoullos, T. and Papageorgiou, H. (1980). On some bivariate probability models applicable to traffic accidents and fatalities. International Statistical Review, 48, 345-356.

31. Campbell, J.T. (1934). The Poisson correlation function. Proceedings of Edinburgh Mathematical Society, Series 2, 4, 18-26.

32. Cassie, R.M. (1962). Frequency distribution models in the ecology of Plankton and other organisms. Journal of Animal Ecology, 31, 65-92.

33. Chatfield, C. (1970). Discrete distributions in market research. In Patil, G.P. (ed., 1970, Vol. 3, pp.163-180).

34. Chatfield, C. and Goodhart, G.J. (1970). The beta-binomial model for consumer purchasing behaviour. Applied Statistics, 19, 240-250.

35. Chatfield, C. and Theobald, C.M. (1973). Mixtures and random sums. The Statistician, 22, 281-285.

36. Chatfield, C., Ehrenberg, A.S.C. and Goodhart, G.J. (1966). Progress on a simplified model of stationary purchasing behaviour. Journal of Royal Statistical Society, Series A, 129, 317-367.

37. Chen, H. and Asau, Y. (1974). Generating random variables from an empirical distribution. American Institute of Industrial Engineers Transactions, 6, 163-166.

38. Consul, P.C. (1974). A simple urn model dependent on predetermined strategy. Sankhyā, Series B, 36, 391-399.

39. Consul, P.C. (1975). Some new characterizations of discrete Lagrangian distributions. In Patil, G.P. et al. (eds., 1975, Vol. 3, pp.279-290).

40. Consul, P.C. and Jain, G.C. (1970). On the generalization of binomial distribution (Abstract). Annals of Mathematical Statistics, 41, 1386-1387.

41. Consul, P.C. and Shenton, L.R. (1972). Use of Lagrange expansion for generating discrete generalized probability distributions. S.I.A.M. Journal of Applied Mathematics, 23, 239-248.

42. Consul, P.C. and Shenton, L.R. (1975). On the probabilistic structure and properties of discrete Lagrangian distributions. In Patil, G.P. et al. (eds., 1975, Vol. 1, pp.41-57).

43. Cresswell, W.L. and Froggatt, P. (1963). The Causation of Bus Driver Accidents: An Epidemiological Study. Oxford University Press, London.

44. Cox, D.R. (1967). Renewal Theory, Methuen, London.

45. Cuppens, R. and Lukacs, E. (1969). Factorization of generalized Poisson distributions. Studia Scientiarum Mathematicarum Hungarica, 4, 367-370.

46. Dacey, M.F. (1969). A hypergeometric family of discrete probability distributions. Geographical Analysis, 1, 283-294.

47. Dacey, M.F. (1970). Regularity in spatial distributions: a problem of geographical analysis. In Patil, G.P. (ed., 1970, Vol. 3, 57-71).

48. Dacey, M.F. (1972). A family of discrete probability distributions defined by the generalized hypergeometric series. Sankhyā, Series B, 34, 243-250.

49. Douglas, J.B. (1955). Fitting the Neyman Type A (two parameter) contagious distributions. Biometrics, 11, 149-173.

50. Douglas, J.B. (1970). Statistical models in discrete distributions. In Patil, G.P. (ed., 1970, Vol. 3, pp.203-232).

51. Douglas, J.B. (1971). Stirling numbers in discrete distributions. In Patil, G.P. et al. (eds., 1971, pp.69-98).

52. Douglas, J.B. (1980). Analysis with Standard Contagious Distributions. International Cooperative Publishing House, Burtonsville, Maryland, U.S.A.

53. Dwass, M. (1967). Poisson recurrent times. Journal of Applied Probability, 4, 605-608.

54. Dwass, M. (1979). A generalized binomial distribution. American Statistican, 33, No. 2 (May 1979), 86-87.

55. Eagleson, G.K. (1964). The polynomial expansions of bivariate distributions. Annals of Mathematical Statistics, 35, 1208-1214.

56. Eagleson, G.K. and Lancaster, H.O. (1967). The regression systems of sums with random elements in common. Australian Journal of Statistics, 9, 119-125.

57. Edwards, C.B. and Gurland, J. (1961). A class of distributions applicable to accidents. Journal of American Statistical Association, 56, 503-517.

58. Erdelyi, A. (ed., 1953). Higher Transcendental Functions Volumes 1, 2 and 3. McGraw-Hill, New York.

59. Erdelyi, A. (ed., 1954). Tables of Integral Transforms Volumes 1 and 2. McGraw-Hill, New York.

60. Evans, D.A. (1953). Experimental evidence concerning contagious distributions in ecology. Biometrika, 40, 186-211.

61. Feller, W. (1943). On a general class of contagious distributions. Annals of Mathematical Statistics, 14, 389-400.

62. Feller, W. (1968). An Introduction to Probability Theory and its applications: Volume 1. John Wiley, New York.

63. Feller, W. (1971). An Introduction to Probability Theory and its applications: Volume 2. John Wiley, New York.

64. Fiacco, A.V. and McCormick, G.P. (1968). Nonlinear Programming: Sequential Unconstrained Minimization Techniques. John Wiley, New York.

65. Fisher, R.A. (1931). Properties of the Functions. (Part of Introduction to) British Association Mathematical Tables, Volume 1. British Association for the Advancement of Science, London.

66. Fisher, R.A. (1941). The negative binomial distribution. Annals of Eugenics, 11, 182-200.

67. Fishman, G.S. (1976). Sampling from the Poisson distribution on a computer. Computing (Vienna), 17, 147-156.

68. Fishman, G.S. (1979). Sampling from a binomial distribution on a computer. Journal of American Statistical Association, 74, 418-423.

69. Froggatt, P. (1965). One-day absence in industry. Journal of Statistical and Social Inquiry Society of Ireland, 21, 166-178.

70. Froggatt, P. (1968). *Research in industrial mobidity: principles* and opportunities. Transactions of Society of Occupational Medicine, 18, 89-95.

71. Froggatt, P. (1970). Application of discrete distribution theory to the study of noncommunicable events in medical epidemiology. In Patil, G.P. (ed., 1970, Vol. 2, 15-40).

81. Guldberg, A. (1931). On discontinuous frequency functions and statistical series. Skandinavisk Aktuarietidskrift, 14, 167-187.

82. Gurland, J. (1957). Some interrelations among compound and generalised Poisson distributions. Biometrika, 44, 265-268.

83. Gurland, J. (1958). A generalized class of contagious distributions. Biometrics, 14, 229-250.

84. Gurland, J. (1959). Some applications of the negative binomial and other contagious distributions. American Journal of Public Health, 49, 1388-1399.

85. Gurland, J. (1965). A method of estimation for some generalised Poisson distributions. In Patil, G.P. (ed., 1965, pp.141-158).

86. Gurland, J. and Hinz, P. (1971). Estimating parameters, testing fit, and analysing untransformed data pertaining to the negative binomial and other distributions. In Patil, G.P. et al. (eds., 1971, pp.143-178).

87. Gurland, J. and Tripathi, R. (1975). Extensions of the Katz family of discrete distributions involving hypergeometric functions. In Patil, G.P. et al. (eds., 1975, Vol. 1, pp.59-82).

88. Gupta, R.C. (1974). Mixtures of some discrete distributions. South African Statistical Journal, 8, 83-92.

89. Gupta, R.P. and Jain, G.C. (1974). A generalized Hermite distribution and its properties. S.I.A.M. Journal of Applied Mathematics, 27, 359-363.

90. Haight, F.A. (1961). Index to the distributions of mathematical statistics. Journal of Research, National Bureau of Standards, Series B: Mathematics and Mathematical Physics, 65B, 23-60.

91. 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.

92. Haldane, J.B.S. (1941). The fitting of the binomial distribution. Annals of Eugenics, 11, 179-181.

93. Hall, W.J. (1956). Some hypergeometric series distributions occurring in birth and death processes at equilibrium (preliminary report). Annals of Mathematical Statistics, 27, 221.

94. Hamdan, M.A. and Al-Bayyati, H.A. (1971). Canonical expansion of the compound correlated bivariate Poisson distribution. Journal of American Statistical Association, 66, 390-393.

95. Hamdan, M.A. and Tsokos, C.P. (1971). A model for physical and biological problems: the bivariate-compounded Poisson distribution. Review of International Statistical Institute, 39, 60-62.

96. Harn, K. van (1978). Classifying Infinitely Divisible Distributions by Functional Equations. Mathematical Centre Tract 103; Mathematisch Centrum, Amsterdam, Netherlands.

97. Hartman, P. and Wintner, A. (1942). On the infinitesimal generators of integral convolutions. American Journal of Mathematics, 64, 272-279.

98. Herdan, G. (1964). Quantitative Linguistics, Butterworths, London.

99. Hinz, P. and Gurland, J. (1967). Simplified techniques for estimating parameters of some generalised Poisson distributions. Biometrika, 54, 555-566.

100. Holgate, P. (1966). Bivariate generalizations of Neyman's Type A distributions. Biometrika, 55, 241-245.

101. Holla, M.S. (1966). On a Poisson inverse Gaussian distribution. Metrika, 11, 115-121.

102. Holla, M.S. and Bhattacharya, S.K. (1965). On a discrete compound distribution. Annals of Institute of Statistical Mathematics (Tokyo), 15, 377-384.

103. Hooke, R. and Jeeves, T.A. (1961). Direct search solution of numerical and statistical problems. Journal of Association of Computing Machinery, 8, 212-229.

104. Huff, B.W. (1978). Some comments on mixtures and infinite divisibility. S.I.A.M. Journal of Applied Mathematics, 35, 17-20.

105. Huzurbazar, V.S. (1950). Probability distributions and orthogonal parameters. Proceedings of Cambridge Philosophical Society, 46, 281-284.

106. Huzurbazar, V.S. (1956). Sufficient statistics and orthogonal parameters. Sankhyā, 17, 217-220.

107. Irwin, J.O. (1965). Inverse factorial series as frequency distributions. In Patil, G.P. (ed., 1965, pp.159-174).

108. Irwin, J.O. (1968). The generalized Waring distribution applied to accident theory. Journal of Royal Statistical Society, Series A, 131, 205-225.

109. Irwin, J.O. (1975). The generalized Waring distribution - Parts I, II, and III. Journal of Royal Statistical Society, Series A, 138, 18-32, 208-228, 374-384.

110. Ismail, M.E.H. and Muldoon, M.E. (1978). Monotonicity of the zeroes of a cross-product of Bessel functions. S.I.A.M. Journal of Mathematical Analysis, 9, 759-767.

111. Jain, G.C. (1974). On power series distributions associated with Lagrange expansion. Biometrische Zeitschrift, 16, 85-97.

112. Jain, G.C. and Consul, P.C. (1971). A generalized negative binomial distribution. S.I.A.M. Journal of Applied Mathematics, 21, 501-513.

113. Johnson, N.L. and Kotz, S. (1969). Discrete Distributions. Houghton Mifflin, New York.

114. Johnson, N.L. and Kotz, S. (1970). Continuous Distributions, Vol. I. Houghton Mifflin, New York.

115. Johnson, N.L. and Kotz, S. (1982). Developments in Discrete Distributions, 1969-1980. International Statistical Review, 50, 71-101.

116. Jorgensen, B. (1982). Statistical Properties of the Generalized Inverse Gaussian Distribution. Springer-Verlag, New York.

117. Katti, S.K. (1966). Interrelations among generalized distributions and their components. Biometrics, 22, 44-52.

118. Katti, S.K. (1967). Infinite divisibility of integer-valued random variables. Annals of Mathematical Statistics, 38, 1306-1308.

119. Katti, S.K. (1977). Infinite divisibility of discrete distributions - III. Colloquia Mathematica Societatis Janos Bolyai 21: Analytic Function Methods in Probability Theory, Debrecen, Hungary; pp.165-171.

120. Katti, S.K. and Gurland, J. (1961). The Poisson Pascal distribution. Biometrics, 17, 527-538.

121. Katti, S.K. and Gurland, J. (1962a). Some methods of estimation for the Poisson-binomial distribution. Biometrics, 18, 42-51.

122. Katti, S.K. and Gurland, J. (1962b). Efficiency of certain methods of estimation for the negative binomial and the Neyman Type A distributions. Biometrika, 49, 215-226.

123. Katti, S.K. and Rao, A.V. (1970). The log-zero-Poisson distribution. Biometrics, 26, 801-813.

124. Katz, L. (1946). On the class of distributions defined by the difference equations: $(x + 1)\, f(x + 1) = (a + bx)\, f(x)$ (abstract). Annals of Mathematical Statistics, 17, 501.

125. Katz, L. (1948). Frequency functions defined by the Pearson difference equation (abstract). Annals of Mathematical Statistics, 19, 120.

126. Katz, L. (1965). Unified treatment of a broad class of discrete probability distributions. In Patil, G.P. (ed., 1965, pp.175-182).

127. Kawamura, K. (1973). The structure of bivariate Poisson distributions. Kodai Mathematical Seminar Reports, 25, 246-256.

128. Kemp, A.W. (1968a). A wide class of discrete distributions and the associated differential equations. Sankhyā, Series A, 30, 401-410.

129. Kemp, A.W. (1968b). A limited ${}_cP_p$. Skandinavisk Aktuariedskrift, 51, 198-203.

130. Kemp, A.W. (1974). Towards a unification of the theory of counts. Statistics Reports and Preprints No. 15, School of Mathematics, University of Bradford, England.

131. Kemp, A.W. (1978). Cluster size probabilities for generalized Poisson distributions. Communications in Statistics, Theory and Methods, A7, 1433-1438.

132. Kemp, A.W. (1979). Convolutions involving binomial pseudo-variables. Sankhyā, Series A, 41, 232-243.

133. Kemp, A.W. (1980). Invertible cumulative probability functions for discrete distributions. Statistics Reports and Preprints No. 47, School of Mathematics, University of Bradford, England.

134. Kemp, A.W. (1981a). Conditionality properties for the bivariate logarithmic distribution with an application to goodness of fit. In Taillie et al. (eds., 1981, Vol. 4, pp.57-73).

135. Kemp, A.W. (1981b). Efficient generation of logarithmically distributed variables. Applied Statistics, 30, 249-253.

136. Kemp, A.W. and Kemp, C.D. (1966). An alternative derivation of the Hermite distribution. Biometrika, 53, 627-628.

137. Kemp, A.W. and Kemp, C.D. (1968). On a distribution associated with certain stochastic processes. Journal of Royal Statistical Society, Series B, 30, 160-163.

138. Kemp, A.W. and Kemp, C.D. (1969). Branching and clustering models associated with the 'lost-games' distribution. Journal of Applied Probability, 6, 700-703.

139. Kemp, A.W. and Kemp, C.D. (1971). On mixing processes and the lost-games distribution. Zastosowania Matematyki Applicationes Mathematicae, 12, No. 2, 167-173.

140. Kemp, A.W. and Kemp, C.D. (1974). A family of discrete distributions defined via their factorial moments. Communications in Statistics, 3, 1187-1196.

141. Kemp, C.D. (1967a). On a contagious distribution suggested for accident data. Biometrics, 23, 241-255.

142. Kemp, C.D. (1967b). Stuttering-Poisson distributions. Journal of Statistical and Social Inquiry Society of Ireland, 21, 151-157.

143. Kemp, C.D. (1970). Accident proneness and discrete distribution theory. In Patil, G.P. (ed., 1970, Vol. 2, pp.41-66).

144. Kemp, C.D. (1971). Properties of some discrete ecological distributions. In Patil, G.P. et al. (eds., 1971, pp.1-22).

145. Kemp, C.D. (1973). An elementary ambiguity in accident theory. Accidents Analysis and Prevention, 5, 371-373.

146. Kemp, C.D. (1982). Low-storage Poisson generators for microcomputers. In COMPSTAT 1982; eds. Caussinus, H., Ettinger, P. and Tomassone, R.; Physica-Verlag, Wuerzburg, Germany; pp.145-155.

147. Kemp, C.D. and Kemp, A.W. (1956a). Generalized hypergeometric distributions. Journal of Royal Statistical Society, Series B, 18, 202-211.

148. Kemp, C.D. and Kemp, A.W. (1956b). The analysis of point quadrat data. Australian Journal of Botany, 4, 167-174.

149. Kemp, C.D. and Kemp, A.W. (1965). Some properties of the Hermite distribution. Biometrika, 52, 381-393.

150. Kemp, C.D. and Kemp, A.W. (1967). A special case of Fisher's modified Poisson series. Sankhyā, Series A, 29, 103-104.

151. Kemp, C.D. and Kemp, A.W. (1969). Some distributions arising from an inventory decision problem. Bulletin of International Statistical Institute, 43, Book 2, 236-338.

152. Kemp, C.D., Kemp, A.W. and Loukas, S. (1979). Computer sampling for discrete distributions. Statistics Reports and Preprints No. 42, School of Mathematics, University of Bradford, England.

153. Kemp, C.D. and Loukas, S. (1978). The computer generation of bivariate discrete random variables. Journal of Royal Statistical Society, Series A, 141, 513-519.

154. Kemp, C.D. and Loukas, S. (1978). Computer generation of bivariate discrete random variables using ordered probabilities. Proceedings of the American Statistical Association, Statistical Computing Section, pp.115-116.

155. Kemp, C.D. and Loukas, S. (1981). Fast methods for generating bivariate discrete random variables. In Taillie et al. (eds., 1981, Vol. 4, pp.313-319).

156. Kemp, C.D. and Papageorgiou, H. (1982). Bivariate Hermite distributions. Sankhyā, Series B, 44, 269-280.

157. Kemp, C.D. and Robson, E.I. (1981). Fast methods for generating discrete random variables on a microcomputer. Proceedings of the American Statistical Association, Statistical Computing Section, pp.138-139.

158. Kendall, M.G. and Stuart, A. (1969). The Advanced Theory of Statistics, Vol. 1, 3rd Edition, Griffin, London.

159. Kennedy, W.J. and Gentle, J.E. (1980). Statistical Computing, Marcel Dekker Inc., New York.

160. Kenney, J.F. (1939). The regression systems of two sums having random elements in common. Annals of Mathematical Statistics, 10, 70-73.

161. Kerr, J.D. (1969). The estimation of the 'Short' distribution. Biometrics, 25, 416-420.

162. Khatri, C.G. (1959). On certain properties of power series distributions. Biometrika, 46, 486-490.

163. Khatri, C.G. (1971). On multivariate contagious distributions. Sankhyā, Series B, 33, 197-216.

164. Khatri, C.G. and Patel, I.R. (1961). Three classes of univariate discrete distributions. Biometrics, 17, 567-575.

165. Kotz, S. and Johnson, N.L. (1973). Statistical distributions: a survey of the literature, trends and prospects. American Statistician, 27, No. 1 (February 1973), 15-17.

166. Kronmal, R.A. and Peterson, A.V. (1979). On the alias method for generating random variables from a discrete distribution. American Statistician, 33, No. 4 (November 1979), 214-218.

167. Lancaster, H.O. (1958). The structure of bivariate distributions. Annals of Mathematical Statistics, 29, 719-736.

168. Leiter, R.E. and Hamdan, M.A. (1973). Some bivariate probability models applicable to traffic accidents and fatalities. International Statistical Review, 41, 87-100.

169. Levy, P. (1937). Sur les exponentielles de polynomes. Ann. Ecole Normale Superieure, 73, 231-292.

170. Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem. Journal of Royal Statistical Society, Series B, 20, 102-107.

171. Lindsey, J.K. (1974). Comparison of probability distributions. Journal of Royal Statistical Society, Series B, 36, 38-47.

172. Lüders, R. (1934). Die Statistik der seltenen Ereignisse. Biometrika, 26, 108-128.

173. Lukacs, E. (1970). Characteristic Functions. 2nd Edition. Charles Griffin, London.

174. Lundberg, F. (1909). Uber die theorie der Rickversicherung. Transactions of the VIth International Congress on Actuaries, 1, 877-955.

175. Maceda, E.C. (1948). On the compound and generalised Poisson distributions. Annals of Mathematical Statistics, 19, 414-416.

176. Mangulis, V. (1965). Handbook of Series for Scientists and Engineers. Academic Press, London.

177. Mardia, K.V. (1970). Familes of Bivariate Distributions. Charles Griffin, London.

178. Martin, D.C. and Katti, S.K. (1962). Approximations to the Neyman Type A distribution for practical problems. Biometrics, 18, 354-364.

179. Martin, D.C. and Katti, S.K. (1965). Fitting of certain contagious distributions to some available data by maximum likelihood method. Biometrics, 21, 34-48.

180. McGuire, J.U., Brindley, T.A. and Bancroft, T.A. (1957). The distribution of European corn-borer larvae in field-corn. Biometrics, 13, 65-78, and correction, 14, 432-434.

181. Naylor, T.H., Balintfy, J.L., Burdick, D.S. and Chu, K. (1966). Computer Simulation Techniques, Wiley, New York.

182. Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. Computer Journal, 7, (1964/65), 308-313.

183. Nelson, D.L. (1975). Some remarks on generalizations of the negative binomial and Poisson distributions. Technometrics, 17, 135-136.

184. Neumann, J. von (1951). Various techniques used in connection with random digits. In The Monte Carlo Method, eds. A.S. Householder et al. (1951). National Bureau of Standards, Applied Mathematics Series No. 12; pp.36-38.

185. Neyman, J. (1939). On a new class of contagious distributions applicable in entomology and bacteriology. Annals of Mathematical Statistics, 10, 35-37.

186. Neyman, J. (1965). Certain chance mechanisms involving discrete distributions. Sankhyā, Series A, 27, 249-258.

187. Neyman, J. and Scott, E.L. (1964). A stochastic model of epidemiology. In Stochastic Models in Medicine and Biology, ed. J. Gurland; University of Wisconsin Press, Madison; pp.45-83.

188. Noack, A. (1950). A class of random variables with discrete distributions. Annals of Mathematical Statistics, 21, 127-132.

189. Ord, J.K. (1967a). On a system of discrete distributions. Biometrika, 54, 649-656.

190. Ord, J.K. (1967b). Graphical methods for a class of discrete distributions. Journal of Royal Statistical Society, Series A, 130, 232-238.

191. Ord, J.K. (1968). The discrete Student's t distributions. Annals of Mathematical Statistics, 39, 1513-1516.

192. Ord, J.K. (1972). Families of Frequency Distributions. Griffin, London.

193. Ord, J.K. (1975). A characterization of a dependent bivariate Poisson distribution. In Patil, G.P. et al. (eds., 1975, Vol. 3, pp.291-297).

194. Ord, J.K., Patil, G.P. and Taillie, C. (eds., 1979). Statistical Distributions in Ecological Work (Statistical Ecology Series, Vol. 4); Co-operative Publishing House, Fairland, Maryland, U.S.A.

195. Paloheimo, J.E. (1972). A spatial bivariate Poisson distribution. Biometrika, 59, 489-492.

196. Papageorgiou, H. (1977). Certain properties of bivariate generalized Poisson distributions - with special reference to the Hermite distribution. Ph.D. Thesis, Statistics, University of Bradford, England.

197. Park, S.H. (1970). Regression and correlation in a bivariate distribution with different marginal densities. Technometrics, 12, 687-691.

198. Patel, Y.C. (1976a). Even point estimation and moment estimation in Hermite distributions. Biometrics, 32, 865-873.

199. Patel, Y.C. (1976b). Estimation of the parameters of the triple and quadruple Stuttering-Poisson distributions. Technometrics, 18, 67-73.

200. Patel, Y.C. (1977). Higher moments of moment estimators and even-point estimators for the parameters of the Hermite distribution. Annals of Institute of Statistical Mathematics (Tokyo), 29, 119-130.

201. 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.

202. Patil, G.P. (1962). On certain properties of the generalized power series distributions. Annals of Institute of Statistical Mathematics (Tokyo), 14, 179-182.

203. Patil, G.P. (1964). On certain compound Poisson and compound binomial distributions. Sankhyā, Series A, 26, 293-294.

204. Patil, G.P. (ed., 1965). Classical and Contagious Discrete Distributions. Statistical Publishing Society, Calcutta, India.

205. Patil, G.P. (ed., 1970). Random Counts in Scientific Work, Volumes 1, 2 and 3; Pennsylvania State University Press, University Park, U.S.A.

206. Patil, G.P. and Bildikar, S. (1967). Multivariate logarithmic series distribution as a probability model in population and community ecology and some of its statistical properties. Journal of American Statistical Association, 62, 655-674.

207. Patil, G.P. and Joshi, S.W. (1968). A Dictionary and Bibliography of Discrete Distributions. Oliver and Boyd, London.

208. Patil, G.P., Kotz, S. and Ord, J.K. (eds., 1975). Statistical Distributions in Scientific Work, Volumes 1, 2 and 3; D. Reidel Publishing Company, Dordrecht, Holland.

209. Patil, G.P., Pielou, E.C. and Waters, W.E. (1971). Statistical Ecology Vol. I: Spatial Patterns and Statistical Distributions. Pennyslvania State University Press, U.S.A.

210. Patil, G.P. and Rao, C.R. (1977). The weighted distributions - a survey of their applications. In Applications of Statistics; ed. P.R. Krishnanah, pp.383-405. North-Holland, Amsterdam.

211. Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics, 34, 179-189.

212. Patil, G.P. and Stiteler, W.M. (1972). Statistical measurement of spatial patterns. Bulletin of International Statistical Institute, 44, 55-83.

213. Patil, G.P. and Stiteler, W.M. (1974). Concepts of aggregation and their quantification: a critical review with some new results and applications. Research on Population Ecology, 15, 238-254.

214. Pearson, K. (1895). Contributions to the mathematical theory of evolution - I: skew variation in homogeneous material. Philosophical Transactions, Series A, 186, 343-414.

215. Philipson, C. (1960). Note on the application of compound Poisson processes to sickness and accident statistics. ASTIN (Association for STudies In Non-life insurance), Bulletin, 1, 224-237.

216. Philipson, C. (1961). Some estimation problems connected with compound Poisson processes. Skandinavisk Aktuarietidskrift, 44, 240-250.

217. Pieters, E.P., Gates, C.E., Matis, J.H. and Sterling, W.L. (1977). Small sample comparison of different estimators of negative binomial parameters. Biometrics, 33, 718-728.

218. Preston, F.W. (1948). The commonness, and rarity, of species. Ecology, 29, 254-283.

219. Quenouille, M.H. (1949). A relationship between the logarithmic, Poisson and negative binomial series. Biometrics, 5, 162-164.

220. Rai, G. (1971). A mathematical model for accident proneness. Trabajos de Estadistica, 22, 207-212.

221. Rao, C.R. (1957). Theory of the method of estimation by minimum chi-square. Bulletin of International Statistical Institute, 35, Book 2, pp.25-32.

222. Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment. In Patil, G.P. (ed., 1965, pp.320-332).

223. Rao, C.R. (1973). Linear Statistical Inference and its Applications. Second Edition. John Wiley, New York.

224. Reid, D.D. (1981). The Poisson-lognormal distribution and its use as a model of plankton aggregation. In Taillie, C. et al. (eds., 1981, Vol. 6, pp.303-316).

225. Riordan, J. (1979). Combinatorial Identities. Krieger; Huntingdon, New York.

226. Ross, G.J.S. (1970). The efficient use of function minimization in nonlinear maximum likelihood estimation. Applied Statistics, 19, 205-221.

227. Sankaran, M. (1968). Mixtures by the inverse Gaussian distribution. Sankhyā, Series B, 30, 455-458.

228. Sankaran, M. (1970a). On certain properties of a class of compound Poisson distributions. Sankhyā, Series B, 32, 353-362.

229. Sankaran, M. (1970b). The discrete Poisson-Lindley distribution. Biometrics, 26, 145-149.

230. Satterthwaite, F.E. (1942). Generalized Poisson distributions. Annals of Mathematical Statistics, 13, 410-417.

231. Shaban, S.A. (1981). On the discrete Poisson inverse Gaussian distribution. Biometrical Journal, 23, 297-303.

232. Shaban, S.A. (1982). Computation of the Poisson-inverse Gaussian distribution. Communications in Statistics, Series A - Theory and Methods, A10, 1389-1399.

233. Shenton, L.R. and Bowman, K.O. (1963). Higher moments of a maximum likelihood estimate. Journal of Royal Statistical Society, Series B, 25, 305-317.

234. Shenton, L.R. and Bowman, K.O. (1967). Remarks on large sample estimators for some discrete distributions. Technometrics, 9, 587-598.

235. Shenton, L.R. and Myers, R. (1965). Comments on estimation for the negative binomial distribution. In Patil, G.P. (ed., 1965, 241-262).

236. Shenton, L.R. and Wallington, P.A. (1962). The bias of moments estimators with an application to the negative binomial. Biometrika, 49, 193-204.

237. Shenton, L.R., Bowman, K.O. and Reinfelds, J. (1967). Sampling moments of moment and maximum likelihood estimators for discrete distributions. Bulletin of the International Statistical Institute, Vol. 42, Book 1, pp.359-361.

238. Shenton, L.R., Bowman, K.O. and Sheehan, D. (1971). Sampling moments of moments associated with univariate distributions. Journal of Royal Statistical Society, Series B, 33, 444-457.

239. Shumway, R. and Gurland, J. (1960a). Fitting the Poisson-binomial distribution. Biometrics, 16, 522-533.

240. Shumway, R. and Gurland, J. (1960b). A fitting procedure for some generalized Poisson distributions. Skandinavisk Aktuariedskrift, 43, 87-108.

241. Sichel, H.S. (1971). On a family of discrete distributions particularly suited to represent long-tailed frequency data. Proceedings of 3rd Symposium on Mathematical Statistics; ed. N.F. Laubschar; C.S.I.R., Pretoria, South Africa; pp.51-97.

242. Sichel, H.S. (1973a). Statistical valuation of diamondiferous deposits. Journal of the South African Institute of Mining Metallurgy, 73, 17-25.

243. Sichel, H.S. (1973b). The density and size distributions of diamonds. Bulletin of International Statistical Institute, 45, Book 2, pp.420-427.

244. Sichel, H.S. (1974). On a distribution representing sentence-length in written prose. Journal of Royal Statistical Society, Series A, 137, 25-34.

245. Sichel, H.S. (1975). On a distribution law for word frequencies. Journal of American Statistical Association, 70, 542-547.

246. Sichel, H.S. (1982a). Asymptotic efficiencies of three methods of estimation for the inverse Gaussian-Poisson distribution. Biometrika, 69, 467-472.

247. Sichel, H.S. (1982b). Repeat-buying and the generalised inverse Gaussian-Poisson distribution. Applied Statistics, 31, 193-204.

248. Simonsen, W. (1976). On the solution of a maximum likelihood equation of the negative binomial distribution. Scandinavian Actuarial Journal, (1976), pp.220-231; with correction (1980), pp.227-228.

249. Skellam, J.G. (1952). Studies in statistical ecology - I: Spatial pattern. Biometrika, 39, 346-362.

250. Slater, L.J. (1966). Generalized Hypergeometric Functions. Cambridge University Press.

251. Sprott, D.A. (1958). The method of maximum likelihood applied to the Poisson binomial distribution. Biometrika, 14, 97-106.

252. Sprott, D.A. (1965). A class of contagious distributions and maximum likelihood estimation. In Patil, G.P. (ed., 1965, pp.337-350).

253. Stacy, E.W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, 33, 1187-1192.

254. Staff, P.J. (1964). The displaced Poisson distribution. Australian Journal of Statistics, 6, 12-20.

255. Steutel, F.W. (1968). A class of infinitely divisible mixtures. Annals of Mathematical Statistics, 39, 1153-1157.

256. Steutel, F.W. (1970). Preservation of Infinite Divisibility under Mixing. Mathematical Centre Tracts 33, Mathematisch Centrum, Amsterdam, Netherlands.

257. Steyn, H.S. (1957). On regression properties of discrete systems of probability functions. Proceedings of Koninklijke Nederlandse Akademie van Wetenchappen, Series A, 60, 119-127.

258. Subrahmaniam, K. (1966). A test for intrinsic correlation in the theory of accident proneness. Journal of Royal Statistical Society, Series B, 28, 180-189.

259. Subrahmaniam, K. and Subrahmaniam, K. (1973). On the estimation of the parameters in the negative binomial. Journal of Royal Statistical Society, Series B, 35, 131-146.

260. Taillie, C., Patil, G.P. and Baldessari, B.A. (eds., 1981). Statistical Distributions in Scientific Work, Vols. 4, 5 and 6. D. Reidel Publishing Company, Dordrecht, Holland.

261. Talwalker, S. (1975). Models in medicine and toxicology. In Patil, G.P. et al. (eds., 1975, Vol. 2, pp.263-274).

262. Teicher, H. (1960). On the mixtures of distributions. Annals of Mathematical Statistics, 31, 55-73.

263. Teicher, H. (1961). Identifiability of mixtures. Annals of Mathematical Statistics, 32, 244-248.

264. Thelander, U. (1979). A remark on 'The Poisson-exponential model and the non-central chi-squared distribution'. Scandinavian Actuarial Journal, pp.52-56.

265. Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. Biometrika, 36, 18-25.

266. Tripathi, R.C. and Gurland, J. (1977). A general family of discrete distributions with hypergeometric probabilities. Journal of Royal Statistical Society, Series B, 39, 349-356.

267. Tucker, H.G. (1961). Best one-sided bounds of infinitely divisible random variables. Sankhyā, Series A, 24, 287-396.

268. Tweedie, M.C.K. (1947). Functions of a statistical variate with given means, with special reference to the Laplacian distributions. Proceedings of Cambridge Philosophical Society, 43, 41-49.

269. Tweedie, M.C.K. (1957). Statistical properties of the inverse Gaussian distribution - Parts I and II. Annals of Mathematical Statistics, 28, 362-377, 696-705.

270. Ussell, T. and Corbin, K.W. (1971). Fitting discrete probability distributions to evolutionary events. Science, 172, 1089-1096.

271. Walker, A.J. (1974). New fast method for generating discrete random numbers with arbitrary frequency distribution. Electronics Letters, 10, 127-128.

272. Wani, J.K. (1968). On the linear exponential family. Proceedings of Cambridge Philosophical Society, 64, 481-483.

273. Warde, W.D. and Katti, S.K. (1971). Infinite divisibility of discrete distributions - II. Annals of Mathematical Statistics, 42, 1088-1090.

274. Westman, W.E. (1971). Mathematical models of contagion and their relation to density and basal area sampling techniques. In Patil, G.P. et al. (eds., 1971, pp.515-536).

Additional Reference

275. Patil, G.P., Boswell, M.T., and Friday, D.S. (1975*). Chance mechanisms in computer generation of random variables In Patil, G.P. et al. (eds., 1975, Vol. 2, pp.37-50).