

NEW STATISTICAL MODELS FOR
DISCRETE UNI- AND MULTIVARIATE
DATA SETS WITH SPECIAL
REFERENCE TO THE DIRICHLET
MULTINOMIAL DISTRIBUTION

by

Aiala Barr

A thesis submitted to the Faculty of Science, University of the
Witwatersrand, Johannesburg, in fulfilment of the requirements for
the degree of Doctor of Philosophy.

Johannesburg, 1987

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Aiale Barr

(Name of candidate)

26 day of January, 1987

ABSTRACT

In classical statistical theory the bi- and multivariate discrete distributions are largely determined by their univariate marginals and they are characterized by a much restricted correlation structure. This study proposes a new statistical model which overcomes these limitations.

The new theory for bivariate discrete distributions is based on two concepts: the ascending diagonal arrays and the third marginal distribution, which make it possible to construct a wide variety of joint discrete probability functions.

The general trivariate model is based on the same principle as the bivariate distribution, with a additional array defined as the "total of variables".

The multivariate case is treated by reducing it into a number of trivariate models.

Some families of distributions are considered for the third marginal and for the ascending diagonal arrays. The new model enables the calculation of any moments for the marginals x_1 and x_2 and of the correlations between the variables, without the need of

explicitly deriving the probability functions for the marginals.

The Compound Poisson distributions proved to be well suited for modelling the third marginal and a general investigation into their properties is undertaken.

The models are also investigated with respect to stochastic processes. The dependency of the stationary model as a function of time is explored and attention is paid to the limiting process $t \rightarrow \infty$.

In particular the Dirichlet-Multinomial distribution is investigated in view of its known low correlation structure. A wider range of correlations is obtained by applying various distributions to the total of variables.

Some existing estimation procedures are surveyed in connection with the distributions which were applied to the third marginal and the ascending diagonal arrays. Maximum likelihood estimation methods are proposed for the joint distributions.

Finally, the new model is tested on four sets of data from different application fields, which had proved to be difficult to fit by other statistical representations. The results were most encouraging.

The last chapter summarizes the developments and suggests possible avenues for further research.

To Zeev, Rimon and Iris

I am pleased to be able to share with you the report of
the children's work and the results of the study, for the
children's work and the results of the study, for the
children's work and the results of the study, for the
children's work and the results of the study, for the

I am pleased to be able to share with you the report of
the children's work and the results of the study, for the
children's work and the results of the study, for the
children's work and the results of the study, for the

ACKNOWLEDGEMENTS

I am indebted to Professor H S Sichel from the Department of Statistics, the University of the Witwatersrand, for his invaluable advice and guidance, the vital discussions and the interest shown in this research.

I am grateful to my husband and my children for their help and great support during the preparation of this thesis.

C O N T E N T S

	Page
DECLARATION	ii
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
LIST OF FIGURES	xiv
LIST OF TABLES	xvii
1 <u>INTRODUCTION</u>	1
1.1 A New Approach for Bi- and Multivariate Discrete ... Distributions	1
1.2 Univariate Compound Poisson Distributions	6
1.2.1 Negative Binomial distribution	7
1.2.2 Poisson-Lognormal distribution	8
1.2.3 Generalized Inverse Gaussian-Poisson distribution ..	9
1.3 Beta-Binomial Distribution	13
1.4 Dirichlet-Multinomial Distribution	16
1.5 Computing Work	20

2	<u>THE BIVARIATE MODEL</u>	21
2.1	The General Model	21
2.1.1	The new model	24
2.2	The Symmetrical Model	28
2.3	Special Distributions for the Ascending Diagonal ...	31
	Arrays	
2.3.1	Beta-Binomial distribution with constant parameters α_1 and α_2	32
2.3.2	Generalized Positive Hypergeometric distribution ...	36
2.3.3	Beta-Binomial distribution with α_1 and α_2 as functions of S	37
2.4	Distributions for the Third Marginal	38
2.4.1	Generalized Inverse Gaussian-Poisson distribution ..	38
2.4.2	Inverse Gaussian-Poisson distribution	39
2.4.3	Negative Binomial distribution	40
2.5	Joint Distributions	41
2.5.1	The Beta-Binomial distribution for the ascending ... diagonal arrays and any distribution for the third marginal	41
2.5.2	The Binomial c for the ascending diagonal arrays and any distribution for the third marginal	45
2.5.3	The Beta-Binomial distribution for the ascending ... diagonal arrays and the Generalized Inverse Gaussian-Poisson distribution for the third marginal	47

2.5.4	The Beta-Binomial distribution for the ascending ... diagonal arrays and the Inverse Gaussian-Poisson distribution for the third marginal	49
2.5.5	The Beta-Binomial distribution for the ascending ... diagonal arrays and the Negative Binomial distribution for the third marginal	51
2.5.6	The Beta-Binomial distribution for the ascending ... diagonal arrays and for the third marginal	54
2.5.7	The Hypergeometric distribution for the ascending .. diagonal arrays and the Beta-Binomial distribution for the third marginal	57
2.5.8	The Binomial distribution for the ascending diagonal arrays and the Beta-Binomial distribution for the third marginal	59
2.5.9	The Binomial distribution for the ascending diagonal arrays and for the third marginal	61
2.5.10	The Beta-Binomial distribution with α_1 and α_2 as ... functions of the ascending diagonal arrays and any distribution for the third marginal	63
2.6	Regression	65
	Case 1	66
	Case 2	68
	Case 3	69
	Case 4	70
	Case 5	72

2.7	The Indicator of the Distribution on the Ascending Diagonal Arrays	74
3	<u>THE TRIVARIATE MODEL</u>	77
3.1	The General Model	77
3.2	Special Distribution for the Ascending Diagonal Arrays, the Third Marginal and the Total of Variables	82
4	<u>THE MULTIVARIATE CASE</u>	88
4.1	The General Model	88
4.2	The Discrete Dirichlet-Multinomial Distribution	89
5	<u>GENERAL PROOFS OF POISSON MIXTURES</u>	100
5.1	The Univariate Case	100
5.1.1	The coefficient of dispersion	105
5.1.2	The coefficient of variation	106
5.1.3	The coefficients of skewness and kurtosis	107
5.1.4	Yule characteristic	110
5.2	The Bivariate Case	110
6	<u>THE BEHAVIOUR OF SPECIAL COMPOUND POISSON DISTRIBUTIONS AS A FUNCTION OF TIME</u>	123
6.1	The Generalized Inverse Gaussian-Poisson Family as a Function of Time	123

6.1.1	Derivation of the Poisson distribution	126
6.1.2	Derivation of the Negative Binomial distribution ...	127
6.1.3	Derivation of the Inverse Gaussian-Poisson	129
	distribution	
6.2	A Discriminating Ratio Function	130
6.2.1	GIGP	130
6.2.2	IGP	131
6.2.3	NBD	131
6.3	Statistical Indices	132
6.3.1	Mean	132
6.3.2	Variance	134
6.3.3	Coefficients of dispersion and variation	136
6.3.4	Coefficients of skewness and kurtosis	138
6.4	The Ratio of the Coefficient of Skewness as a	141
	Function of t	
6.5	The Ratio of the Coefficients of Kurtosis as a	148
	Function of t	
6.6	Relations between the IGP Model and the NBD Model ..	154
7	<u>DEVELOPMENT OF THE BIVARIATE, TRIVARIATE AND</u>	165
	<u>MULTIVARIATE MODELS AS A FUNCTION OF TIME</u>	
7.1	The Bivariate Model	165
7.2	The Trivariate Model	169
7.3	The Multivariate Model	171

8	<u>ESTIMATION</u>	172
8.1	Methods of Estimation for the Beta Binomial Distribution	172
8.1.1	Method of moments	174
8.1.2	Method of equating the first sample proportion and the sample mean	174
8.1.3	Method of equating the first and last observed proportions	175
8.1.4	Method of maximum likelihood	176
8.2	Methods of Estimation for the Generalized Positive Hypergeometric Distribution	176
8.3	Method of Estimation for the Parameters of the Ascending Diagonal Arrays Distributions	176
8.4	Methods of Estimation for the Negative Binomial Distribution	177
8.4.1	Method of moments	178
8.4.2	Method of mean and zero proportion	178
8.4.3	Method of maximum likelihood	179
8.5	Zero-Augmented NBD	179
8.6	Inverse Gaussian-Poisson Distribution	181
8.6.1	Method of moments	182
8.6.2	Method of zero-proportion and mean in sample	182
8.6.3	Maximum likelihood method	183

8.7	Generalized Inverse Gaussian-Poisson Distribution ..	183
8.8	Methods of Estimation using Joint Distributions	184
8.8.1	The NBD - Binomial model	184
8.8.2	The NBD - BBD model	187
8.9	The Expected Fisher Information Matrix	190
8.9.1	The NBD - Binomial model	191
8.9.2	The NBD - BBD model	192
9	<u>APPLICATIONS OF THE NEW STATISTICAL MODELS</u>	194
9.1	The Data Set from the Accident Field, used by	195
	Bates and Neyman (1952)	
9.2	The Data Set from the Children Accident Field,	211
	used by Mellinger et al (1965)	
9.3	The Data Set from the Library Circulation Field, ...	218
	used by Chen (1976)	
9.3.1	The NBD - GPHD model	220
9.3.2	The NBD - Binomial model	226
9.4	The Data Set from the Library Circulation Field, ...	236
	used by Cane and Burell (1982)	
10	<u>SUMMARY</u>	245
	<u>REFERENCES</u>	252

LIST OF FIGURES

	Page
Figure 2.1 The bivariate table with x_1 marginal, x_2 marginal and the third marginal	23
Figure 2.2 The bivariate table with symmetry along the .. main diagonal	29
Figure 2.3 The indicator function	76
Figure 6.1 The coefficients of skewness ratio for IGP ... with $b = 0,1$ and $c = 0,4$	144
Figure 6.2 The coefficients of skewness ratio for IGP ... with $b = 0,1$ and $c = 0,04$	145
Figure 6.3 The coefficients of skewness ratio for IGP ... with $b = 0,1$ and $c = 0,004$	146
Figure 6.4 The coefficients of skewness ratio for GIGP .. with $\gamma = 0$, $b = 0,1$ and $c = 0,04$	147
Figure 6.5 The coefficients of kurtosis ratio for IGP ... with $b = 0,1$ and $c = 0,4$	150
Figure 6.6 The coefficients of kurtosis ratio for IGP ... with $b = 0,1$ and $c = 0,04$	151
Figure 6.7 The coefficients of kurtosis ratio for IGP ... with $b = 0,1$ and $c = 0,004$	152

Figure 6.8	The coefficients of kurtosis ratio for GIGP ..	153
	with $\gamma = 0$, $b = 0,1$ and $c = 0,04$	
Figure 6.9	The NBD and the IGP with equal means and	157
	variances for $k = 0,9$, $\theta_{nb} = 0,985$, $\alpha_I = 10,35$	
	and $\theta_I = 0,992$	
Figure 9.1	Estimates of α_1 as a function of S	202
Figure 9.2	Estimates of α_2 as a function of S	203
Figure 9.3	The x_1 marginal, data of Bates and Neyman	205
	(1952)	
Figure 9.4	The x_2 marginal, data of Bates and Neyman	206
	(1952)	
Figure 9.5	The third marginal, data of Bates and Neyman .	207
	(1952)	
Figure 9.6	The regression curve $E(x_2 x_1)$, Bates and Neyman	208
	(1952)	
Figure 9.7	The x_1 marginal, data of Mellinger (1965)	214
Figure 9.8	The x_2 marginal, data of Mellinger (1965)	215
Figure 9.9	The third marginal, data of Mellinger (1965) .	216
Figure 9.10	The regression curve $E(x_2 x_1)$, data of	217
	Mellinger (1965)	
Figure 9.11	The x_1 marginal, data of Chen (1976), the	221
	NBD-GPHD model	
Figure 9.12	The x_2 marginal, data of Chen (1976), the	222
	NBD-GPHD model	
Figure 9.13	The third marginal, data of Chen (1976), the .	223
	NBD-GPHD model	

Figure 9.14	The regression curve $E(x_2 x_1)$, data of Chen (1976), the NBD-GPHD model	.. 225
Figure 9.15	The x_1 marginal, data of Chen (1976), the univariate NBD model	 227
Figure 9.16	The x_2 marginal, data of Chen (1976), the univariate NBD model	 228
Figure 9.17	The means of the ascending diagonal arrays ($\hat{p} = 0,505$) Chen (1976)	... 229
Figure 9.18	The x_1 marginal, data of Chen (1976), the NBD-Binomial model	 232
Figure 9.19	The x_2 marginal, data of Chen (1976), the NBD-Binomial model	 233
Figure 9.20	The third marginal, data (1976), the NBD-Binomial model	. 234
Figure 9.21	The regression curve $E(x_2 x_1)$, data of Chen (1976), the NBD-Binomial model	.. 235
Figure 9.22	Estimates of α_1 and α_2 as a function of S, data of Cane and Burrell (1982)	... 240
Figure 9.23	The x_1 marginal, data of Cane and Burrell (1982)	 241
Figure 9.24	The x_2 marginal, data of Cane and Burrell (1982)	 242
Figure 9.25	The third marginal, data of Cane and Burrell (1982)	. 243
Figure 9.26	The regression curve $E(x_2 x_1)$, Cane and Burrell (1982)	244

LIST OF TABLES

	Page
Table 6.1 The NBD model with parameters $k(1) = 1/2$ and $\theta_{nb}(1) = 1/2$ and the IGP model with parameters $\gamma_I(1) = -1/2$, $\theta_I(1) = 2/3$ and $\alpha_I(1) = \sqrt{3}/2$ at $t = 1$	159
Table 6.2 The NBD model with parameters $k(2) = 1/2$ and $\theta_{nb}(2) = 2/3$ and the IGP model with parameters $\gamma_I(2) = -1/2$, $\theta_I(2) = 4/5$ and $\alpha_I(2) = \sqrt{5}/2$ at $t = 2$	160
Table 6.3 The NBD model with parameters $k(3) = 1/2$ and $\theta_{nb}(3) = 3/4$ and the IGP model with parameters $\gamma_I(3) = -1/2$, $\theta_I(3) = 6/7$ and $\alpha_I(3) = \sqrt{7}/2$ at $t = 3$	161
Table 9.1 The data set of Bates and Neyman (1952) with $\hat{\gamma} = 1,54$, $\hat{\theta} = 0,8136$, $\hat{\alpha}_1 = 1,73$ and $\hat{\alpha}_2 = 6,79$	200
Table 9.2 Estimates of the parameters α_1 and α_2 for the ascending diagonal arrays	201
Table 9.3 The χ^2 results obtained from Table 9.1	204
Table 9.4 The fit of the univariate NBD model to the marginals	209

Table 9.5	The data set of Mellinger et al. (1965) with $\hat{\gamma} = 3,587$, $\hat{\theta} = 0,406$, $\hat{\alpha}_1 = 12,27$ and $\hat{\alpha}_2 = 11,48$	212
Table 9.6	The χ^2 results obtained from Table 9.5	213
Table 9.7	The data set of Chen (1976), the NBD-GPHD model with $\hat{\gamma} = 3,91$, $\hat{\theta} = 0,56$, $\hat{\alpha}_1 = -39,32$ and $\hat{\alpha}_2 = -38,867$	219
Table 9.8	The χ^2 results obtained from Table 9.7	220
Table 9.9	The fit of the univariate NBD model to the marginals	224
Table 9.10	The data set of Chen (1976), the NBD-Binomial model with $\hat{\gamma} = 3,91$, $\hat{\theta} = 0,56$ and $\hat{p} = 0,505$	230
Table 9.11	The χ^2 results obtained from Table 9.10	231
Table 9.12	The data set of Cane and Burrell (1982)	237
Table 9.13	The χ^2 results obtained from Table 9.12	239

CHAPTER 1

INTRODUCTION

This study is concerned with the development of a new theory for bi- and multivariate discrete distributions. An attempt is made to investigate the general properties of Compound Poisson distributions and special attention is given to the Dirichlet distribution as a particular case of the general model.

The multivariate Dirichlet model was the subject of continuous investigations for over two decades. Aitchinson (1982) mentioned the particular interest in the model by stating: *"The realization that the Dirichlet class leans so heavily towards independence has prompted a number of authors [Connor and Mosimann (1969), Darroch and James (1974), Mosimann (1975), James and Mosimann (1980), James (1981)] to search for generalizations of the Dirichlet class with less structure"*.

1.1 A New Approach for Bi- and Multivariate Discrete Distributions

The existing theory for bi- and multivariate discrete data sets deals largely with the following two situations:

CHAPTER 1

INTRODUCTION

This study is concerned with the development of a new theory for bi- and multivariate discrete distributions. An attempt is made to investigate the general properties of Compound Poisson distributions and special attention is given to the Dirichlet distribution as a particular case of the general model.

The multivariate Dirichlet model was the subject of continuous investigations for over two decades. Aitchinson (1982) mentioned the particular interest in the model by stating: *"The realization that the Dirichlet class leans so heavily towards independence has prompted a number of authors [Connor and Mosimann (1969), Darroch and James (1974), Mosimann (1975), James and Mosimann (1980), James (1981)] to search for generalizations of the Dirichlet class with less structure".*

1.1 A New Approach for Bi- and Multivariate Discrete Distributions

The existing theory for bi- and multivariate discrete data sets deals largely with the following two situations:

- (a) The correlations between any two random variables are determined by the parameters of the marginal distributions; and
- (b) The correlations are zero for independent variables, specifically if the marginals are independent, or very low correlations exist in the Dirichlet case where the variables p_i ($i = 1, \dots, k$) are subject to the constraint $\sum_{i=1}^k p_i = 1$. For a medium number of variables in the Dirichlet case we have "almost zero" correlations between any two variables.

We note that multivariate discrete distributions are characterized by restricted correlation structures. Taillie et al. (1979) state that "the problem of finding good families of multivariate distributions is both a challenging and vexing one. An ideal multivariate family should have enough flexibility to describe widely different patterns of covariation". By deriving multivariate analogues of the univariate distributions, Taillie et al. (1979) state that "the parameters of these distributions are completely determined by their univariate marginals. This implies that the joint distribution is also determined by the marginals. In other words, these distributions have no flexibility whatsoever with regard to covariation". In the same paper Taillie et al. determine the properties which should be the guide for any statistician in creating an ideal multivariate family. According to them a multivariate family must have the following properties:

"(1) the ability to describe wide patterns of covariation,

(2) parameters that are easily interpreted in relation to that covariation; and

(3) a parametric structure permitting many and relatively simple inferential procedures".

So far no existing family of multivariate discrete distributions complies with all the three above-mentioned properties. It is with this in mind that we are searching for a new approach to handling multivariate discrete data sets. In this study we are concerned with the building of a new model, which is based on principles that are different from the classical methodology.

In the classical bivariate discrete distributions we discuss horizontal and vertical arrays. The summation of these horizontal and vertical arrays leads to the customary marginal x_1 and x_2 distributions.

In the approach set out in this investigation we introduce a third marginal S in addition to the first two marginals x_1 and x_2 . The model is based also on a new type of arrays, namely the ascending diagonal arrays, which are at a right angle to the main diagonal.

The "third marginal" is the distribution of a variable S defined as the summation of x_1 and x_2 , i.e. $S = x_1 + x_2$.

The probability law of the ascending diagonal array is the distribution of x_1 given S , where $x_1 = 0, 1, 2, \dots, S$.

The summation of the ascending diagonal arrays leads to the

probabilities in the third marginal. For the third marginal we allow any discrete distribution $f(S)$ and for the ascending diagonal arrays we consider special families of distributions

$$\phi(x_1|S), \text{ where } \sum_{x_1=0}^S \phi(x_1|S) = 1.$$

This new approach permits the creation of any bivariate and trivariate distribution. The customary x_1 and x_2 marginals are derived from the inside of the table. It is not always possible to express the theoretical distributions of x_1 and x_2 in a closed form. This does not inhibit however the calculations of the means, variances, correlations and regressions.

The multivariate model is reduced to trivariate models.

In general the new model will allow more flexibility and will not be constrained by the creation of a bivariate or multivariate version of a specific univariate distribution.

Kemp and Papageorgiu (1982), in considering alternative forms of the bivariate Hermite distribution, state that "various authors have discussed the problem of constructing meaningful and useful bivariate versions of a given univariate distribution, the main difficulty being the impossibility of producing a standard set of criteria that can always be applied to produce a unique distribution which could unequivocally be called the bivariate version".

Marshall and Olkin (1985) were aware of the problem stated above and derived a family of bivariate distributions, which were generated by the bivariate Bernoulli distribution as an extension of the univariate case. They extended their work also to some bivariate continuous distributions such as the bivariate exponential, gamma and normal distributions.

In addition to the new theory developed we wish to investigate what happens if these models are applied to stochastic processes. We are interested in the behaviour of the variables x_1 and x_2 , or S and x_1 , in terms of the exposure period t . This dependency of the stationary model as a function of time is explored and attention is paid to the limiting process $t \rightarrow \infty$.

Certain models which assume special distributions on the ascending diagonal arrays and on the third marginal are considered.

On the ascending diagonal arrays we deal with the same type of distributions and two cases are identified:

- (a) distributions which are independent of S ; and
- (b) distributions which are dependent on S .

For certain fields of applicability of the general multivariate model, it is necessary to consider the Poisson mixtures as the distribution of the summation of all the variables.

1.2 Univariate Compound Poisson Distributions

The Poisson distribution was initially used to discuss rare events and appeared in the statistical literature for the first time describing death due to mule kicks in the Prussian army. Its derivation was published by Poisson in 1837. Since then it was under the investigation of many statisticians. Real data does not meet the stringent assumptions of the distribution in numerous occasions, a fact which brought new attempts to relax the Poisson model assumptions.

The Compound Poisson model is a specific way to create families of distributions which can satisfy the real data. It is obtained from the ordinary Poisson distribution for which the parameter λ is made to vary in accordance with any mixing distribution as follows:

$$P(X=x) = \int_0^{\infty} P(X=x|\lambda) \varphi(\lambda) d\lambda, \quad (1.2.1)$$

$$\text{where } P(X=x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \text{ for } x = 0, 1, \dots \quad (1.2.2)$$

is the Poisson probability distribution function and $\varphi(\lambda)$ is the probability law of the mixing distribution when $\lambda \in (0, \infty)$.

Johnson and Kotz (1969) denote the above model as:

$$X \sim P(\lambda) \underset{\lambda}{\wedge} \varphi(\lambda|\theta), \quad (1.2.3)$$

where θ is the parameter (or the vector of parameters) of the mixing distribution.

We discuss only a few compound Poisson families by considering

some well known mixing distributions.

1.2.1 Negative Binomial distribution

The Negative Binomial distribution was developed by Montmort in 1713. In more modern times Greenwood and Yule (1920) used the compound Poisson distribution for the first time in statistical literature in accident-proneness' field. For the mixing they used the Gama distribution given by the probability law:

$$p(\lambda) = \frac{1}{\Gamma(\gamma)} \left[\frac{1-\theta}{\theta} \right]^\gamma \lambda^{\gamma-1} e^{-\left[\frac{1-\theta}{\theta} \right] \lambda}, \quad (1.2.4)$$

where $\gamma > 0$, $0 < \theta < 1$ and $\lambda \in (0, \infty)$.

The solution of the convolution integral leads to the Negative Binomial Distribution, for short called the NBD, given by

$$\begin{aligned} P(X=x) &= \int_0^\infty e^{-\lambda} \frac{\lambda^x}{x!} \frac{1}{\Gamma(\gamma)} \left[\frac{1-\theta}{\theta} \right]^\gamma \lambda^{\gamma-1} e^{-\left[\frac{1-\theta}{\theta} \right] \lambda} d\lambda \\ &= \frac{(1-\theta)^\gamma}{\Gamma(\gamma)} \frac{\Gamma(x+\gamma)}{\Gamma(x+1)} \theta^x, \end{aligned} \quad (1.2.5)$$

where $x = 0, 1, \dots$.

The above process is just one of several possible theoretical approaches which are known to lead to an NBD.

The general history of the Negative Binomial Distribution was summarized in reviews by Gurland (1954) and Bartko (1961).

The study of accident data led to an exceedingly powerful discovery, namely the Multivariate Negative Binomial Distribution. The basic mathematical properties of the bivariate NBD were

established by Arbous and Kerrich (1951). An explicit treatment of the multivariate NBD was presented by Bates and Neyman (1952) in the accident-proneness field.

The Negative Binomial Distribution has been widely used for modelling in various fields such as:

- (a) psychology by Sichel (1951);
- (b) market research by Ehrenberg (1959, 1968 and 1972), Chatfield (1966), Morrison and Schmittlein (1981) and Grahn (1969); and
- (c) accident proneness by Greenwood and Yule (1920), Bates and Neyman (1952), Lundberg (1940) and Mellinger et al. (1965).

1.2.2 Poisson-Lognormal distribution

An early study to create long tail distributions which are also capable of having a mode was made by Grundy (1951). He considered the distribution generated from the compounding between the Poisson and Lognormal distribution, i.e.

$$\text{Poisson } (\lambda) \wedge \text{Lognormal } (m, \tau).$$

It is assumed that the parameter λ has a Lognormal distribution, which is a function of m and τ . Consequently $\ln(\lambda)$ has a Normal distribution with mean ξ and variance σ^2 . The following relations exist between the parameters of Lognormal and Normal distribution:

$$m = \exp(\xi + \sigma^2/2), \text{ and} \quad (1.2.6)$$

$$r = \exp(\sigma^2). \quad (1.2.7)$$

The probability law of the Poisson Lognormal is given by:

$$P(x) = \frac{m^x r^{x(x-1)/2}}{x! \sqrt{2\pi \ln(r)}} \int_{-\infty}^{+\infty} \exp\left[-\frac{u^2}{2 \ln(r)} - m r^{x-1/2} e^u\right] du, \quad (1.2.8)$$

where $x = 0, 1, 2, \dots$, $0 < m < \infty$ and $r > 1$.

This is a skewed distribution. It is known however for its intractability, because of the convolution integral which cannot be expressed in a closed form.

Grundy (1951) also developed a usable approximation of Poisson-Lognormal frequency function.

This distribution was termed by Anscombe (1950) as the discrete lognormal.

In more recent times the distribution was used by Bulmer (1974) in the field of ecology.

1.2.3 Generalized Inverse Gaussian-Poisson distribution

In many fields the observed frequency distributions representing discrete data display long upper tails, which are difficult to reproduce by mathematical models. Previous attempts by Fisher (1943) lead to the Logarithmic Series Distribution, for short LSG, which has only one parameter and starts at $x = 1$. It can only reproduce a mode at its origin and is incapable of modal values

for $x > 1$.

Patil and Shorrock (1966) developed the modified logarithmic series distribution, which has two parameters with possible modes at $x = 0$ or $x = 1$.

The NBD itself, which generally starts at $x = 0$, cannot reproduce a long tail as well as a mode at $x \geq 1$ at the same time.

In order to model discrete distributions with extremely long upper tails, Sichel (1971) mixed the Poisson with the Generalized Inverse Gaussian, for short GIG distribution (Jorgenson, 1982), which lead to a new discrete distribution, called the Generalized Inverse Gaussian-Poisson, for short GIGP. Several contemporary writers are now calling the GIGP the Sichel distribution.

As the mixing distribution GIG lies somewhere between Pearson type III and type V distributions, it leads to compound Poisson distributions which are more flexible than the other compoundings.

The probability law of the GIGP is obtained from the compounding between the GIG given by:

$$P(\lambda | \theta, \gamma, \alpha) = \frac{(2\sqrt{1-\theta}/\alpha\theta)^\gamma}{2 K_\gamma(\alpha\sqrt{1-\theta})} \lambda^{\gamma-1} e^{-\left[\frac{1-\theta}{\theta}\right]\lambda - \frac{\alpha^2\theta}{4\lambda}}, \quad (1.2.9)$$

where $\lambda \in (0, \infty)$, $\alpha > 0$, $0 < \theta < 1$, $-\infty < \gamma < \infty$ and $K_\gamma(z)$ is the modified Bessel function of the second kind of order γ with argument z ;

and the Poisson distribution.

The convolution integral leads to the following probability distribution function:

$$P(x|\alpha, \theta, \gamma) = \frac{(\sqrt{1-\theta})^\gamma}{K_\gamma(\alpha\sqrt{1-\theta})} \frac{(\alpha\theta/2)^x}{x!} K_{x+\gamma}(\alpha), \quad (1.2.10)$$

for $x = 0, 1, 2, \dots$.

As stated by Sichel (1975), *"the choice of the mixing distribution ... is crucial. The analytical form of the mixing distribution ... is essentially based on an empirical choice. The justification for its use lies in its capability of satisfactorily representing a large body of observed data. These remarks do in fact apply to the genesis of most compound distributions quoted in the statistical literature"*.

The GIGP distribution is a three-parameter family. Due to the third parameter involved the GIGP distribution enables greater power for modelling in comparison with the customary two parameters distributions. A number of well known distribution functions such as the Poisson, negative binomial, geometric, Fisher's logarithmic series and others can be derived by a limiting process from the GIGP.

A specific form of the GIGP is obtained for $\gamma = -1/2$. This is called the Inverse Gaussian-Poisson, with the probability law given as:

$$P(x|\alpha, \theta, \gamma = -1/2) = \left[\frac{2\alpha}{\pi} \right]^{1/2} e^{\alpha\sqrt{1-\theta}} \frac{(\alpha\theta/2)^x}{x!} K_{x-1/2}(\alpha),$$

for $x = 0, 1, 2, \dots$. (1.2.11)

The Inverse Gaussian-Poisson distribution is obtained by compounding the Poisson with the Inverse Gaussian distribution. The latter was introduced to statistics by Tweedie (1957) although it was already known to eminent physicists such as Einstein and Schroedinger.

The GIGP and IGP have been used successfully in different data fields such as:

- (a) market research (repeat buying), Sichel (1982);
- (b) linguistics, Sichel (1971, 1974, 1975), Atkinson and Lam Yeh (1982);
- (c) mining, Sichel (1973a, 1973b);
- (d) ecology, Sichel (1985); and
- (e) bibliometry, Sichel (1985).

The behaviour of "compound Poisson" distributions is studied in this thesis in general and special attention is given to the relation between the compound and the mixing distribution. Among other remarkable relations, it is worth mentioning that the second moment about the mean for the compound Poisson distribution

$$\mu_2(x) = \mu_2(\lambda) + \mu_1'(\lambda) \quad (1.2.12)$$

is expressed in terms of the variance of the mixing distribution and the average of variances of the Poisson distributions. The above relation can be interpreted in terms of the Sum of Squares equation in a one-way analysis of variance as following:

$\mu_2(\lambda)$ may be looked upon as the variance between the treatments (Poisson distributions) and

$\mu'_1(\lambda)$ could be regarded as the variance within treatments (Poisson distributions).

In what follows the moments, coefficients of variation and dispersion and the coefficients of skewness and kurtosis are all expressed in terms of the exposure period t . Some interesting results are obtained in the limiting process of $t \rightarrow \infty$.

For the development of the new theory for bi- and multivariate discrete distributions we need to discuss certain probability laws of discrete random variables. In particular we need to specify the ascending diagonal arrays. To start with we consider the Beta-Binomial distribution family.

1.3 Beta-Binomial Distribution

The Beta-Binomial distribution arises in various contexts. Polya (1931) obtained this probability law by studying a particular urn problem. Skellam (1948) derived the distribution, discussed the moment and the maximum likelihood estimation and showed its use in applications. Gart (1958) obtained the Beta-Binomial distribution in studying survival curves. Darwin (1960) used the Beta-Binomial in studying species-frequency data.

Ishii and Hayakawa (1960) presented a particularly clear explanation of Beta-Binomial. They dealt also with the multivariate aspect of Beta-Binomial distribution.

The necessity for compounding a Binomial distribution was already suggested by Fisher in "Statistical methods for research workers", in his discussion about "Comparison of Sex Ratio in Human Families with Binomial Distribution". The following is a quotation from Fisher, page 66:

"If a family of 8 is regarded as a random sample of 8 from the general population, the number of boys in such families should be distributed in the binomial $(r+p)^8$ where p is the proportion of boys. If, however, families differ not only by chance, but by a tendency on the part of some parents to produce males or females, then the distribution of the number of boys should show an excess of unequally divided families, and a deficiency of equally or nearly equally divided families."

It is assumed therefore that p is not a constant, but a random variable which follows a certain probability law.

The distribution, which is obtained from the ordinary Binomial distribution where the parameter p is made to vary in accordance with the Beta distribution, was termed by Ishii and Hayakawa (1960) the Beta-Binomial distribution, or for short EBD. Johnson and Kotz (1969) denoted the Beta-Binomial model as follows:

$$\text{Binomial } (n, p) \wedge \text{Beta } p \quad (1.3.1)$$

In the statistical literature this model is also known as the Polya-Eggenberger distribution.

The general compound Binomial is obtained from the convolution integral:

$$f(x|S) = \int_0^1 \binom{S}{x} p^x (1-p)^{S-x} \psi(p) dp, \quad (1.3.2)$$

where $\psi(p)$ is the "mixing" distribution of parameter p in the Binomial with $0 < p < 1$. If we substitute the Beta distribution

$$\psi(p) = \frac{1}{B(\alpha_1, \alpha_2)} p^{\alpha_1-1} (1-p)^{\alpha_2-1} \quad (1.3.3)$$

into (1.3.2) we obtain the following:

$$f(x|S, \alpha_1, \alpha_2) = \binom{S}{x} \frac{B(x+\alpha_1, S-x+\alpha_2)}{B(\alpha_1, \alpha_2)}, \quad \text{where} \quad (1.3.4)$$

$x = 0, 1, 2, \dots, S$; $S = 1, 2, \dots$ and $\alpha_1 > 0$, $\alpha_2 > 0$.

The resulting BBD has two parameters α_1 and α_2 , both greater than zero. The shape of the distribution is determined by these two parameters. With $\alpha_1 < 1$ and $\alpha_2 \geq 1$ the distribution has a mode at zero (reverse J-shaped), whilst with $\alpha_1 < 1$ and $\alpha_2 < 1$ the distribution has modes at zero and at S (U-shaped). This is considered a necessary but not a sufficient condition.

Another form of BBD is given below:

$$f(x|S, \alpha_1, \alpha_2) = \frac{\binom{-\alpha_1}{x} \binom{-\alpha_2}{S-x}}{\binom{-\alpha_1-\alpha_2}{S}}, \quad \text{where} \quad (1.3.5)$$

$x = 0, 1, 2, \dots, S$, $S = 1, 2, \dots$, $\alpha_1 > 0$, and $\alpha_2 > 0$.

If α_1 and α_2 are taken as negative, the latter expression gives:

$$f(x|S, \alpha_1, \alpha_2) = \frac{\binom{\alpha_1}{x} \binom{\alpha_2}{S-x}}{\binom{\alpha_1+\alpha_2}{S}}, \quad \text{where} \quad (1.3.6)$$

$x = 0, 1, 2, \dots, S$, $S = 1, 2, \dots$, $\alpha_1 > 0$, and $\alpha_2 > 0$.

The last function is called the Generalized Positive

Author Barr Aiala

Name of thesis New Statistical Models For Discrete Uni- And Multivariate Data Sets With Special Reference To The Dirichlet Multinomial Distribution. 1987

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.