STATISTICAL MODELLING AND INFERENCE FOR A CLASS OF BIVARIATE AND RELATED DISTRIBUTIONS

NG CHOUNG MIN

A submission of thesis for the fulfilment
of the degree of Doctor of Philosophy

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2010

ACKNOWLEDGEMENT

I would like to take this opportunity to convey my sincere appreciation to those who have helped and guided me throughout the process of researching and writing this thesis.

First and foremost, I would like to express my highest gratitude to my thesis supervisor, Professor Dr. Ong Seng Huat for overseeing the whole process with unstinting support and interest. Your wise counsels and insights enable me to gain much useful knowledge not only in the researched topic, but also of the essential researching skills for future work. Thank you for imparting sound advice and knowledge as well as guiding me in solving the problems that arise throughout these past few years.

Thank you to my thesis examiners for taking the time to go through this thesis and giving such constructive, helpful comments as well as suggestions to improve the thesis.

A very special thanks to my family as their indefatigable encouragement and support see me through the thick and thin during the development of my thesis.

Last but not least, I would like to extend my gratefulness to all those who have lent a helping hand in various ways, contributing to the successful completion of my thesis.

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: NG CHOUNG MIN (I.C/Passport No: 821126-13-5196)

Registration/Matric No: SHB060001

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

STATISTICAL MODELLING AND INFERENCE FOR A CLASS OF BIVARIATE AND RELATED DISTRIBUTIONS

Field of Study: STATISTICS

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature Date

Subscribed and solemnly declared before,

Witness’s Signature Date

Name:
Designation:

ABSTRACT

This thesis considers bivariate extension of the Meixner class of distributions by the method of generalized trivariate reduction so that the marginal distributions have different parameters; in particular, a new bivariate negative binomial (*BNB*) distribution is examined. Different marginal parameters allow flexibility in statistical modelling and simulation studies when different marginal distributions and a specified correlation are required. The multivariate extension of this class of distributions is also given. Specifically, various interesting properties of the proposed *BNB* distribution, such as canonical expansion and quadrant dependence are examined. In addition, potential applications of the proposed distribution, as a bivariate mixed Poisson distribution, and the computer generation of bivariate samples are discussed. Due to the complicated or intractable joint probability function (pf) for most bivariate and multivariate distributions, the popular method of maximum likelihood estimation (*MLE*) either leads to a slow parameter estimation or totally could not be employed. Furthermore, *MLE* is not robust in the presence of outliers. Alternative robust methods like minimum Hellinger distance (*MHD*) can be used but these methods may also involve complicated pf. To address this difficulty in estimation, a Hellinger type distance measure based on the probability or moment generating function is proposed as a tool for quick and robust parameter estimation. The proposed method is shown to yield consistent estimators. It is computationally much faster than *MLE* or *MHD* since the generating function required is usually much simpler compared to the corresponding pf. The distribution of the difference of two discrete random variables, particularly that of two correlated negative binomial random variables from the proposed *BNB* distribution, is also studied. The application of this distribution, which caters for non homogeneity in a group of individuals, in modelling fluctuating asymmetry based on meristic (counts) traits in organisms is discussed. A test for fluctuating asymmetry, based on a zero-inflated count model, is examined. Also, numerical illustrations are given to complement the ideas and theories put forth.

ABSTRAK

Disertasi ini mempertimbangkan pelanjutan kelas taburan Meixner ke kes bivariat dengan menggunakan kaedah penurunan trivariat teritlak supaya taburan-taburan sut akan mempunyai parameter yang berlainan; khasnya, satu taburan binomial negatif bivariat (*BNB*) yang baru telah dikaji. Parameter-parameter yang berlainan bagi taburan sut mengizinkan kelenturan dalam pemodelan berstatistik dan kajian simulasi apabila taburan-taburan sut yang berlainan serta satu korelasi yang tertentu diperlukan. Pelanjutan kelas taburan ini ke kes multivariat juga diberi. Khususnya, pelbagai ciri-ciri menarik, seperti kembangan kanonik dan kebersandaran sukuan, bagi taburan *BNB* yang dicadangkan dikaji. Selain itu, aplikasi berpotensi bagi taburan yang dicadangkan, sebagai satu taburan Poisson bercampur bivariat, dan penjanaan komputer bagi sampel-sampel bivariat dibincangkan. Disebabkan oleh fungsi kebarangkalian (fk) tercantum yang rumit bagi kebanyakan taburan bivariat dan multivariat, kaedah penganggaran kebolehjadian maksimum (*PKM*) yang popular akan membawa kepada sama ada satu penganggaran parameter yang lambat, ataupun langsung tidak dapat digunakan. Tambahan pula, *PKM* tidak teguh semasa terdapatnya outlier. Kaedah-kaedah teguh alternatif seperti jarak Hellinger minimum (*JHM*) boleh digunakan tetapi kaedah-kaedah ini mungkin juga melibatkan fk yang rumit. Untuk mengatasi kesukaran dalam penganggaran ini, satu sukatan jarak jenis Hellinger yang berdasarkan fungsi penjana kebarangkalian atau fungsi penjana momen dicadangkan sebagai satu alat untuk penganggaran parameter yang cepat dan teguh. Kaedah yang dicadangkan ini ditunjukkan menghasilkan penganggar-penganggar yang konsisten. Kaedah ini adalah lebih cepat secara pengiraan berbanding dengan *PKM* atau *JHM* memandangkan fungsi penjana yang diperlukan adalah biasanya lebih ringkas berbanding dengan fk yang sepadan. Taburan bagi perbezaan antara dua pembolehubah rawak diskrit, khasnya bagi dua pembolehubah rawak binomial negatif yang berkolerasi dari taburan *BNB* yang dicadangkan, juga dikaji. Aplikasi taburan ini, yang mengambil kira ketidakhomogenan dalam satu kumpulan individu, dalam pemodelan asimetri berfluktuasi berdasarkan ciri-ciri meristik (bilangan) pada organisma dibincangkan. Satu ujian bagi asimetri berfluktuasi berdasarkan satu model bilangan sifar-tertambah (zero-inflated) turut dikaji. Sebagai tambahan, ilustrasi-ilustrasi berangka diberikan bagi melengkapkan gagasan-gagasan dan teori-teori yang dikemukakan dalam disertasi ini.

TABLE OF CONTENTS

NOTATION AND ABBREVIATIONS

$\xrightarrow{a.s.}$	-	Almost sure convergence
$\xrightarrow{P}$	-	Convergence in probability
$\xrightarrow{d}$	-	Convergence in distribution
$(a)_n$	-	Pochhammer's symbol
${}_1F_1\ (a;b;z)$	-	Confluent hypergeometric function
${}_2F_1\ (a,b;c;z)$	-	Gaussian hypergeometric function
$m_r(x;\nu,p)$	-	rth Meixner polynomial
$m_i^*(x;\nu,p)$	-	ith orthonormal Meixner polynomial
α	-	Significance level
$\boldsymbol{\theta}=(\theta_1,\theta_2,\dots,\theta_d)$	-	d-dimensional parameter vector
$\widehat{\boldsymbol{\theta}}$, $\overline{\boldsymbol{\theta}}$	-	Estimates of $\boldsymbol{\theta}$
$\mu_{\boldsymbol{X}}^{[x_1,x_2,\dots,x_k]}$	-	Factorial moment of order $(x_1,x_2,\dots,x_k)$ for $\boldsymbol{X}$
$\rho_{\boldsymbol{X}}$	-	Correlation coefficient of $\boldsymbol{X}$
χ^2	-	Chi-square
χ_k^2	-	Chi-square with k degrees of freedom
$\Gamma(\alpha,\beta)$	-	Gamma distribution with parameters α and β
$\boldsymbol{\Gamma}$	-	Fisher information matrix
$f_{\boldsymbol{X}}$, $f_{\boldsymbol{\theta}}$ or f	-	Probability function for $\boldsymbol{X}$ with parameters $\boldsymbol{\theta}$
f_n	-	Empirical probability function
$G_{\boldsymbol{X}}$, $G_{\boldsymbol{\theta}}$ or G	-	Probability generating function for $\boldsymbol{X}$ with parameters $\boldsymbol{\theta}$
$\widehat{G}_n$	-	Empirical probability generating function
H_0	-	Null hypothesis

H_1	-	Alternative hypothesis
I	-	Information matrix for a sample
$L(\boldsymbol{\theta})$	-	Likelihood function for unknown parameters $\boldsymbol{\theta}$
MD_α	-	Minimum distance
$M_{\boldsymbol{X}}$	-	Moment generating function for $\boldsymbol{X}$
n	-	Sample size
N	-	Number of Monte Carlo samples
$N(\mu, \sigma^2)$	-	Normal distribution with mean μ and standard deviation σ
$\Pr(\cdot)$	-	Probability of an event
S_c	-	Score statistic
S_{c_1}	-	Score statistic for one-sided hypothesis test
U	-	Efficient score matrix
$\boldsymbol{X} = (X_1, X_2, \cdots, X_k)$	-	k-variate random variable
$X_1, X_2, Y_1, Y_2, W_1, W_2$	-	Random variables
$X \sim D(\boldsymbol{\theta})$	-	X distributed as a distribution D with parameters $\boldsymbol{\theta}$
d.f.	-	Degrees of freedom
fmgf	-	Factorial moment generating function
mgf	-	Moment generating function
pf	-	Probability function
pdf	-	Probability density function
pgf	-	Probability generating function
pmf	-	Probability mass function
rv	-	Random variable
$Binomial(n, p)$	-	Binomial distribution with index n and probability p
Binomial-Binomial	-	Binomial difference

BN	-	Bivariate normal
BNB	-	Bivariate negative binomial
BNNB	-	Bivariate non-central negative binomial
$BPoisson$	-	Bivariate Poisson
$B\Gamma$	-	Bivariate Gamma
DA	-	Directional asymmetry
EBNB	-	Extended bivariate negative binomial
FA	-	Fluctuating asymmetry
MBinomial	-	Multivariate binomial
MNB	-	Multivariate negative binomial
$NB(\nu, p)$	-	Negative binomial distribution with index ν and probability p
LR	-	Likelihood ratio
MHD	-	Minimum Hellinger distance
MHDE	-	Minimum Hellinger distance estimation
MGHD	-	Minimum generalized Hellinger distance
MPGHD	-	Minimum penalized generalized Hellinger distance
MLE	-	Maximum likelihood estimation
MSE	-	Mean squared error
NB-NB	-	Negative binomial difference
$Poisson(\lambda)$	-	Poisson distribution with mean λ
Poisson-Poisson	-	Poisson difference distribution
SA	-	Simulated annealing

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF APPENDICES

CHAPTER 1 : INTRODUCTION

1.0 Probability Distributions in Statistics

Probability distributions form the core of statistics as empirical models in statistical data analysis or as mathematical models to explain the random variations or disturbances in a theoretical analysis. Various systems of univariate distributions have been proposed when only one variable is of interest. For instance, for continuous random variables, we have the Pearson system, Johnson's system and the Edgeworth and Gram-Charlier expansions while, for the discrete case, the difference-equation families (Katz, 1946; Ord, 1967, 1972; Sundt & Jewell, 1981), power series distributions and Kemp families (see Johnson *et al*., 2005) are well-known. The Meixner class defined by Meixner (1934), which contains as members the binomial, Poisson, negative binomial, normal, gamma and Meixner hypergeometric distributions such that their orthogonal polynomials have generating functions of the form $G(t,x) = f(t)e^{xu(t)}$, has a number of interesting properties. Apart from the binomial distribution, members of this class are infinitely divisible. Distributions from the Meixner class are solutions to a quadratic regression problem (Laha & Lukacs, 1960). Another interesting characterization of the Meixner family is due to Morris (1982) who showed that distributions of the natural exponential family have at most a quadratic variance function of the mean if, and only if, they belong to the Meixner family. A recent addition to families of distributions is that of Jones (2004), who considered a general family based on the distribution of order statistics. For more than one variable, bivariate and multivariate extensions of these univariate distributions are required where the interdependencies among these variables are incorporated. A number of books on bivariate and multivariate distributions (Mardia, 1970; Hutchinson & Lai, 1990; Joe, 1997; Johnson *et al*., 1997; Kotz *et al*., 2000; Balakrishnan & Lai, 2009) have appeared in the literature and these have proved very useful to researchers as well as practitioners.

Various methods of constructing bivariate and multivariate distributions are available in the existing literature. Some of these are the differential equations, bivariate Edgeworth expansions and translation methods (see Mardia, 1970), which are extensions of univariate methods. Other methods such as compounding, generalizing and convolutions have also been proposed (see Hutchinson & Lai, 1990; Kocherlakota & Kocherlakota, 1992). Some specific methods of construction pertinent to this thesis will be discussed in the next section. More recent methods of construction through copulas and mixtures have been considered in order to achieve different types of dependence structure in multivariate distributions (see Joe, 1997). For bivariate generalizations of important discrete univariate distributions like the Poisson, binomial, negative binomial, logarithmic series and Neyman Type A, their formulations, statistical inference and applications in diverse areas are described in detail in Kocherlakota & Kocherlakota (1992); see also Hutchinson & Lai (1990) for analyses of discrete data sets. Bivariate and multivariate generalizations of distributions are of continuing interest as exemplified by recent works of Biswas & Hwang (2002), Jones & Larsen (2004), Kundu & Gupta (2009) and Kundu *et al*. (2010). Furthermore, the recent work of Azzalini (2005) and Arellano-Valle & Azzalini (2008) on the skew-normal (Azzalini, 1985) and other skewed distributions have attracted the attention of many researchers.

1.1 Literature Review

In this section, we present a brief review pertinent to the development of this thesis and to provide motivation for the research problems considered.

The Bernoulli trials form the building blocks in the construction of bivariate discrete distributions, leading to the development of bivariate binomial distributions as studied by Aitken & Gonin (1935), Hamdan (1972) and also, Hamdan & Jensen (1976). Edwards & Gurland (1961) and Holgate (1964) then considered the bivariate negative binomial and bivariate Poisson (which can also be obtained as a limit of the bivariate binomial distribution (Campbell, 1934)) distributions, respectively. Among others, Kibble (1941), Lancaster (1957), Gumbel (1960), Mardia (1967), Downton (1970) and Lai & Moore (1984) have contributed significantly to the development of bivariate normal, gamma and exponential distributions. The efforts by these earlier researchers have since paved the way for the construction of a multitude of bivariate and multivariate distributions such as multivariate mixed Poisson, multivariate power series, bivariate logistic, bivariate skew-normal, bivariate generalized exponential, Farlie-Gumbel-Morgenstern copula and multivariate log normal distributions (see Johnson *et al*., 1997; Kotz *et al*., 2000; Balakrishnan & Lai, 2009) to reflect realistic situations for practical applications.

A popular method of constructing bivariate distributions among the researchers is trivariate reduction or random element in common. Through trivariate reduction, many bivariate generalizations of well-known univariate distributions have been constructed (see Mardia, 1970; Kocherlakota & Kocherlakota, 1992). For the Meixner family of distributions (Meixner, 1934), Eagleson (1964) employed trivariate reduction to obtain their bivariate extensions which have the property that their joint distributions possess canonical expansions in their marginal distributions and corresponding orthogonal polynomials (see Chapter 2). However, apart from a restricted range of the correlation coefficient, bivariate distributions formed by trivariate reduction usually suffer from a lack of flexibility due to the marginal distributions having the same parameters. For instance, by the trivariate reduction, the bivariate binomial distribution has equal probability parameters (Hamdan, 1972) and the bivariate gamma distribution has equal scale parameters (Cherian, 1941 (cited in Mardia, 1970)) in the marginals.

Furthermore, in statistical modelling and simulation studies, different marginal distributions and a less restrictive correlation are often required. Most Monte Carlo simulation experiments need a varying degree of dependence (Devroye, 1986, p. 573) or a specification of different marginal distributions, especially when the bivariate structure of the distribution is not well understood. Clearly, possessing different marginal distributions makes the bivariate distribution more flexible for empirical modelling. In this connection, it is of interest to consider an extension of the existing Meixner class of bivariate distributions, such that the marginal distributions have different parameters and a wide range of the correlation coefficient.

Mixed Poisson models form a useful class of distributions in practical applications (see Johnson *et al*., 2005). These applications of bivariate and multivariate mixed Poisson distributions have been examined by Stein & Juritz (1987), Aitchison & Ho (1989), Chib & Winkelmann (2001) and Ferrari *et al*. (2004), among others. Edwards & Gurland (1961) and Subrahmaniam (1966) have also considered modelling accident proneness using their bivariate negative binomial in the context of the bivariate mixed Poisson model. However, as observed by Chib & Winkelmann (2001), there is still a lack of parametric multivariate count distributions to cater for a wide variability of correlation structures that arise in practice. It would be instructive to enhance the current class of bivariate and multivariate count distributions for statistical analysis by considering new distribution for correlated counts that arise as a member of the mixed Poisson family under different circumstances.

In this technological age, computers have made it possible to simulate diverse types of populations from a variety of distributions, and to perform fast and efficient computation of essential quantities like the distribution function. More importantly, simulated data from a specified distribution can greatly assist in statistical data analysis as well as inference; for example, in examining the properties and performance of estimators and hypothesis tests (see Zhao & Joe, 2005; Patil & Shirke, 2007). Thus, Ong (1992, 1993, 1995) has considered the computer generation and computation of bivariate distributions based on their mixture formulations. Recently, Michael & Schucany (2002) and Minhajuddin *et al*. (2004) considered the simulation of bivariate and multivariate distributions similar to the work of Ong (1993). Balakrishnan & Lai (2009) have reviewed a number of the simulation approaches found in literature for bivariate continuous distributions. As varying the degree of dependence is required in many Monte Carlo simulations, Ong (2008) gave a review of this issue in the context of mixture models.

Since most simulation studies require a specification of the dependence between random variables of interest, distributional properties such as canonical expansion and quadrant dependence must be investigated. Lancaster (1958), Eagleson (1964), Hamdan & Al-Bayyati (1971), Hamdan & Jensen (1976) and Gupta (1979) are among those who have enriched the field of study on canonical expansion for bivariate and multivariate distributions. A canonical expansion of a bivariate distribution is a single series expansion in terms of its marginal distributions and the corresponding orthogonal polynomials. It throws light on the structure of the distribution, such as the correlations between the random variables. Recently, Cuadras (2002) derived canonical expansion of bivariate distributions in terms of distribution functions.

Another interesting concept of dependence in a bivariate distribution is quadrant dependence. Quadrant dependence (Lehmann, 1966) is a useful concept of bivariate dependence since it is easier to verify than the usual linear dependence. This concept is imperative in reliability theory, where the random variables are seldom independent in practice. Joe (1997) gave a rather comprehensive description on most positive as well as negative dependence concepts found in literature. Lai & Xie (2000) have constructed a family of positive quadrant dependent bivariate distributions. Colangelo *et al*. (2006), Belzunce *et al*. (2007) and Colangelo *et al*. (2008) are among those continuing the work related to positive dependence.

The famed maximum likelihood estimation is well-known to yield estimators which are asymptotically efficient but sensitive towards outliers. On the other hand, minimum Hellinger distance estimation has been proven to work well in the presence of outliers (Beran, 1977). Due to this attractive characteristic, Tamura & Boos (1986), Simpson (1987), Lindsay (1994) and Basu *et al*. (1997) have developed minimum divergence methods to account for various circumstances of data that arise in practice. Basu (2002) further considered the corresponding tests of hypotheses for the generalized Hellinger divergence family in discrete models. However, most of these methods made use of the probability functions that are inherently intractable for most bivariate and multivariate distributions. Furthermore, a complicated probability function slows down the parameter estimation process. This is especially the case for discrete distributions. To address this problem, Kemp & Kemp (1988) have introduced a rapid estimation method for univariate distributions based upon the probability generating function but the estimation depends upon predetermined initial values of the variable in the probability generating function. A method of parameter estimation to avoid this dependence will be useful.

There are times when the distribution for the difference of two independent or correlated random variables from a bivariate case comes in handy for data analysis. In particular, this distribution of differences has found application in areas such as reliability theory, marketing, risk analysis, accident analysis and sports modelling. Surprisingly, in spite of the usefulness, the application of this distribution to paired count data still receives comparatively less attention in statistical literature than its continuous counterpart. Another situation is the analysis of fluctuating asymmetry in organisms. The idea of fluctuating asymmetry (Van Valen, 1962) has been widely researched and used as a measure of developmental stability in organisms (Palmer, 1994). Developmental stability is defined as the ability to defend against random deviations from perfect bilateral symmetry, which may be expressed as $(L - R)$, the random differences between the left (L) and right (R) sides of a particular morphological structure or trait. Recently, Graham *et al*. (2003) studied growth models as well as fluctuating asymmetry under additive and multiplicative error models, and with respect to active and inert tissue models involving the lognormal and gamma distributions. These continuous distributions arise naturally in growth (Mosimann & Campbell, 1988) processes.

Fluctuating asymmetry has also been considered based upon the difference in counts of a trait, for example, the difference in counts of *Drosophila* bristles (Mather, 1953; Woods *et al*., 1998), spots on the plumage of barn owls (Roulin *et al*., 2003) and number of pectoral fin rays in rainbow trout fry (Young *et al*., 2009). It is noted that heterogeneity in a population invariably influences the trait size. Differences in trait size may in turn affect the data collected especially for meristic counts; for example, Johnson *et al*. (2004) has shown that larger chinook salmon has significantly more traits which can be counted than smaller fish. Larger traits may also appear to be more asymmetrical (Palmer, 1994; Knierim *et al*., 2007). Hence, a joint distribution of (L, R) counts that takes into account the heterogeneity aspect of the population is of interest in the study of fluctuating asymmetry. Furthermore, analysis involving meristic counts needs care since the data may be highly skewed with a majority of perfect bilateral symmetrical data and span a narrow range of differences in left and right counts (Knierim *et al*., 2007). Since a high incidence of zeros is a natural measure of symmetry, the inflation of the zero counts in the distribution of $(L - R)$ may be proposed as a measure of developmental stability. This may be useful; for instance, in the control of moths by using egg parasitoids (see Hewa-Kapuge & Hoffmann, 2001).

1.2 Contributions of the Thesis

The main contributions of this thesis are listed below.

- An extended class of bivariate and multivariate distributions from the Meixner family (except the Meixner hypergeometric distribution) has been formulated by an extension of the trivariate reduction method. The marginals for these distributions are more flexible with different parameters.
    - A result on the existence of canonical expansion for this class of bivariate distributions has been obtained by extending the result of Eagleson (1964).
    - In particular, a new extended bivariate negative binomial distribution, which includes bivariate negative binomial of Edwards & Gurland (1961) and Subrahmaniam (1966), has been obtained by the extended trivariate reduction method. This distribution has also been shown to arise as a bivariate mixed Poisson model.
    - Explicit formulae for the distributional properties of the extended bivariate negative binomial distribution are given, including that of the canonical expansion and the information matrix.
- A rapid yet robust and consistent parameter estimation method based on the probability generating function for bivariate and multivariate distributions (MD_α estimation method) has been proposed.
    - This method does not suffer from the effects of predetermined initial values since there is no such need for selecting these values.

- A family of distributions defined by the difference of two random variables has been considered when they are (a) independent and (b) correlated, and many properties as well as computational issues have been studied.
    - A novel application of the zero-inflated model has been proposed to measure asymmetry.
    - Power of one-sided score and likelihood ratio tests for significance of zero-count inflation has been examined.

Parts of the thesis work have been published (Ong *et al.*, 2008; Sugita *et al.*, 2010). Three papers based upon Chapters 3, 4 and 5 have been submitted for publication.

1.3 Thesis Organization

Chapter 2 will serve as the preliminary chapter to the ensuing chapters by explaining briefly the needed terms and concepts. Some fundamental definitions and supporting theorems for ease of reference can be found in this chapter.

The next three chapters comprise of the main findings for this thesis. Chapter 3 contains the formulations for a class of bivariate and multivariate distributions, particularly of those which arise from the Meixner class of univariate distributions. Of a special interest here is the extended bivariate negative binomial distribution, which has marginal distributions possessing different parameters. Basic properties of this distribution such as joint probability mass function, factorial moments and conditional distributions are given. The canonical expansion and quadrant dependence of this distribution are also investigated. A result on the canonical expansion of distributions derived from the extended trivariate reduction method is given, extending the result of Eagleson (1964) for the Meixner class of distributions. On the practical side, applications of this distribution are illustrated with numerical examples. Also given are algorithms to generate bivariate samples with varying dependence according to the formulations introduced in this chapter.

In Chapter 4, a rapid parameter estimation method based on probability generating function for bivariate and multivariate distributions is proposed. Consistency of the estimators from this method will be shown. Numerical examples are given to clearly demonstrate the competency of this method.

Chapter 5 will in turn dwell upon the distribution of the difference between two discrete random variables, specifically when these random variables are correlated. This distribution serves to model the difference of a meristic trait between bilateral sides of an organism. By making use of a zero-inflated model, a statistical test to determine fluctuating asymmetry in organisms is established.

Finally, Chapter 6 gives the conclusion along with suggestions for further works.

CHAPTER 2 : PRELIMINARIES

2.0 Introduction

In this chapter, terms and concepts are presented to facilitate discussion in the ensuing chapters. Interpretation of certain notations is also explained here.

2.1 Formulation of Bivariate Distributions

Bivariate discrete and continuous distributions can be constructed in a wide variety of ways. Examples of methods of construction are trivariate reduction, mixing or compounding, and sampling. A number of these methods have been reviewed in Mardia (1970) and Kocherlakota & Kocherlakota (1992).

2.1.1 Trivariate Reduction

The method of trivariate reduction or random element in common is a popular method of construction due to its simplicity and ease of generating samples on a computer when given a univariate generator. This method is defined (Mardia, 1970) following the definition for additive property of a family of distributions.

Definition 2.1 (*Additive Property*, Arnold, 1970): Let $\{F(y;\lambda)\}_{\lambda>0}$ be a family of distributions parameterized by λ, and let h be a function mapping $\mathbb{R}^2$ into $\mathbb{R}$. $\{F(y;\lambda)\}_{\lambda>0}$ will be said to form an additive family under h if for any $\lambda_1, \lambda_2 > 0$ and Y_1, Y_2 are independent, $F_{Y_1}(y) = F(y;\lambda_1)$ and $F_{Y_2}(y) = F(y;\lambda_2)$ imply $F_{h(Y_1,Y_2)}(y) = F(y;\lambda_1 + \lambda_2)$.

Definition 2.2 (*Trivariate Reduction*): Given three independent additive random variables (rv) Y_1, Y_2 and W from the same family of distribution, a bivariate generalization (X_1, X_2) is given by the convolutions

$$X_1 = Y_1 + W \text{ and } X_2 = Y_2 + W. \tag{2.1}$$

Bivariate distributions in literature that can be formed using (2.1) include bivariate normal, gamma, Poisson and negative binomial distributions. Note that this trivariate reduction technique has been generalized by Arnold (1967) to construct bivariate and multivariate distributions which are closed under other operations (see, for example, Mardia, 1970; Kundu & Gupta, 2009).

2.1.2 Compounding

The compounding technique to produce bivariate distributions can be defined as:

Definition 2.3 (*Compounding of Uncorrelated Random Variables*): Let $X_1|\psi$ and $X_2|\psi$ be independent rv's with probability functions (pf's) $f_{X_1}(x_1|\psi)$ and $f_{X_2}(x_2|\psi)$ for a given parameter ψ. Let ψ be a value of the rv Ψ with mixing distribution having pf $g(\psi; \xi)$, where ξ is the vector of parameters. Then, the bivariate distribution (X_1, X_2) is said to be a compound distribution with pf given by

$$h(x_1, x_2; \xi) = \int f_{X_1}(x_1|\psi) f_{X_2}(x_2|\psi) g(\psi; \xi)\, d\psi. \tag{2.2}$$

Specifically, (X_1, X_2) is a mixed Poisson distribution with its joint probability mass function (pmf) given by (2.2) where $f_{X_i}(x_i|\psi)$, $i = 1,2$ is the Poisson pmf with parameter ψ regarded as a value of the rv Ψ with mixing distribution $g(\psi; \xi)$. When Ψ is taken to be the gamma rv, the bivariate negative binomial distribution or also known as bivariate compound Poisson distribution is obtained as shown in Arbous & Kerrich (1951).

In the above formulation, $X_1|\psi$ and $X_2|\psi$ are assumed to be independent. There are cases where X_1 and X_2 may be correlated given the parameter ψ. To account for this correlation, the compounding technique can be extended as follows.

Definition 2.4 (*Compounding of Correlated Random Variables*): Let $(X_1, X_2)|\psi$ be rv's with joint probability function $f_{(X_1,X_2)}(x_1, x_2|\psi)$ for a given parameter ψ. Let ψ be a value of the rv Ψ as defined in Definition 2.3. Then, the bivariate distribution (X_1, X_2) is said to be a compound distribution with pf given by

$$h(x_1, x_2; \xi) = \int f_{(X_1,X_2)}(x_1, x_2|\psi) g(\psi; \xi)\, d\psi.$$

Following this convention, Edwards & Gurland (1961) obtained a bivariate negative binomial distribution which they termed as the compound correlated bivariate Poisson distribution.

2.2 Properties of Distribution

2.2.1 Probability Generating Function and Joint Probability Function

Probability generating function (pgf) for a bivariate discrete distribution (X_1, X_2) is unique with respect to its corresponding joint probability mass function (pmf) $\Pr(X_1 = x_1, X_2 = x_2) = f(x_1, x_2)$. They are related through the equation

$$G_{(X_1,X_2)}(z_1, z_2) = E\left[z_1^{x_1} z_2^{x_2}\right] = \sum_{x_1,x_2} f(x_1, x_2) z_1^{x_1} z_2^{x_2}\,. \tag{2.3}$$

Henceforth, the pgf for (X_1, X_2) will be denoted by $G_{(X_1,X_2)}(z_1, z_2)$ or $G(z_1, z_2)$. For continuous distributions, the terms moment generating function (mgf) and joint probability density function (pdf) are used instead of pgf and pmf, respectively.

Kocherlakota & Kocherlakota (1992, p. 2) outlined two ways to obtain the pmf from its pgf as follows.

(i) The pgf is expanded in powers of z_1 and z_2, that is the form of the right-most expression in the equation (2.3). The coefficient of the term $z_1^{x_1} z_2^{x_2}$, which is $f(x_1, x_2)$, will give the pmf of the distribution at (x_1, x_2).

(ii) The pgf is differentiated repeatedly with respect to z_1 and z_2 before evaluating the result at $z_1 = 0$ and $z_2 = 0$ to obtain the pmf, $f(x_1, x_2)$. Mathematically,

$$f(x_1, x_2) = \frac{1}{x_1!\, x_2!} \frac{\partial^{x_1+x_2}}{\partial z_1^{x_1} \partial z_2^{x_2}} G(z_1, z_2)\bigg|_{z_1=0, z_2=0} .$$

2.2.2 Marginal and Conditional Distributions

Let the joint pmf of (X_1, X_2) be $f(x_1, x_2)$. Then, the pmf's of the marginal distributions X_1 and X_2 are $f_{X_1}(x_1) = \sum_{x_2} f(x_1, x_2)$ and $f_{X_2}(x_2) = \sum_{x_1} f(x_1, x_2)$ respectively. The corresponding pgf's of the marginals which can be obtained from $G_{(X_1,X_2)}(z_1, z_2)$ are $G_{X_1}(z_1) = G_{(X_1,X_2)}(z_1, 1) = \sum_{x_1} f_{X_1}(x_1) z_1^{x_1}$ and similarly, $G_{X_2}(z_2) = G_{(X_1,X_2)}(1, z_2) = \sum_{x_2} f_{X_2}(x_2) z_2^{x_2}$.

A useful Theorem 1.3.1 (due to Subrahmaniam, 1966) from Kocherlakota & Kocherlakota (1992, p. 13) regarding the pgf of the conditional distribution of (X_1, X_2) will be quoted here. The result from this theorem enables the determination of the regression of X_1 on X_2 without having to first find the conditional probability function, which may be difficult in most cases.

Theorem 2.1 (Theorem 1.3.1, Kocherlakota & Kocherlakota, 1992): Let $G(z_1, z_2)$ be the joint pgf of (X_1, X_2). Then the pgf of the conditional distribution of X_1 given $X_2 = x_2$ is $G_{X_1}(z|x_2) = \frac{G^{(0,x_2)}(z,0)}{G^{(0,x_2)}(1,0)}$, where $G^{(x_1,x_2)}(u,v) = \frac{\partial^{x_1+x_2}}{\partial z_1^{x_1} \partial z_2^{x_2}} G(z_1, z_2)\Big|_{z_1=u, z_2=v}$.

Based on this theorem, a corollary is found by Kocherlakota and Kocherlakota regarding the regression of X_1 on X_2. Again, the corollary is quoted here.

Corollary 2.1 (Corollary 1.3.1, Kocherlakota & Kocherlakota, 1992, p. 14): Regression of X_1 on X_2 is $E[X_1|X_2 = x_2] = \frac{G^{(1,x_2)}(1,0)}{G^{(0,x_2)}(1,0)}$.

2.2.3 Factorial Moments

Let the factorial moments for a bivariate discrete distribution (X_1, X_2) be $\mu_{(X_1,X_2)}^{[x_1,x_2]} = E\left[X_1^{[x_1]} X_2^{[x_2]}\right]$ where $X_i^{[x_i]} = X_i(X_i - 1) \dots (X_i - x_i + 1), i = 1, 2$. Then, the factorial moment generating function can be defined by (Kocherlakota & Kocherlakota, 1992)

$$H(t_1, t_2) = G(t_1 + 1, t_2 + 1) = \sum_{t_1, t_2} \mu_{(X_1,X_2)}^{[x_1,x_2]} \frac{t_1^{x_1}}{x_1!} \frac{t_2^{x_2}}{x_2!} \ . \qquad (2.4)$$

Another method to obtain the factorial moments is by differentiating $G(t_1 + 1, t_2 + 1)$ repeatedly and evaluating the result at $t_1 = 0$ and $t_2 = 0$, that is

$$\mu_{(X_1,X_2)}^{[x_1,x_2]} = \frac{\partial^{x_1+x_2}}{\partial t_1^{x_1} \partial t_2^{x_2}} G(t_1 + 1, t_2 + 1)\Big|_{t_1=0, t_2=0} = G^{(x_1,x_2)}(1,1). \qquad (2.5)$$

These two methods are a direct extension from the univariate discrete case of obtaining the factorial moments $\mu_X^{[x]}$ for a random variable X (see, for example, Johnson *et al*., 2005, p. 59).

Correlation between X_1 and X_2 can be found by making use of the factorial moments $\mu_{(X_1,X_2)}^{[x_1,x_2]}$ as well as the factorial moments of the marginals, $\mu_{X_1}^{[x_1]}$ and $\mu_{X_2}^{[x_2]}$ through the equation

$$\rho_{(X_1,X_2)} = \frac{\mu_{(X_1,X_2)}^{[1,1]} - \mu_{X_1}^{[1]}\mu_{X_2}^{[1]}}{\sqrt{\left(\mu_{X_1}^{[2]} + \mu_{X_1}^{[1]}\left(1 - \mu_{X_1}^{[1]}\right)\right)\left(\mu_{X_2}^{[2]} + \mu_{X_2}^{[1]}\left(1 - \mu_{X_2}^{[1]}\right)\right)}} . \tag{2.6}$$

2.2.4 Information Matrix

Let $f_{\boldsymbol{X}}(\mathbf{x}_i;\boldsymbol{\theta})$, $i = 1,2,\dots,n$ denote the pf and $L(\boldsymbol{\theta};\mathbf{x})$ denote the likelihood function, with $\boldsymbol{\theta} = (\theta_1,\theta_2,\dots,\theta_d)$ representing the vector of unknown parameters and $\mathbf{x} = (\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n)$ the n sets of observed values for the k-variate random variable $\boldsymbol{X} = (X_1,X_2,\dots,X_k)$. The information matrix for a single observation, also known as the Fisher information, is given as (Hogg & Craig, 1995, p. 372)

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \left\{E\left[\frac{\partial \ln f_{\boldsymbol{X}}(\mathbf{x}_1;\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial \ln f_{\boldsymbol{X}}(\mathbf{x}_1;\boldsymbol{\theta})}{\partial\theta_j}\right]\right\} \quad \text{or}$$

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \left\{E\left[-\frac{\partial^2 \ln f_{\boldsymbol{X}}(\mathbf{x}_1;\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]\right\},\ i,j = 1,2,\dots,d.$$

Under regularity conditions, the information matrix for the sample **x** is then

$$\mathbf{I}(\boldsymbol{\theta}) = \left\{E\left[-\frac{\partial^2 \ln L(\boldsymbol{\theta};\mathbf{x})}{\partial\theta_i\partial\theta_j}\right]\right\} = n\boldsymbol{\Gamma}(\boldsymbol{\theta}),\ i,j = 1,2,\dots,d.$$

2.3 Structure of Bivariate Distributions

2.3.1 Canonical Expansion

The canonical expansion of a bivariate probability density function is a useful tool in the study of the structure of bivariate distributions (Kotz, 1974). If a bivariate distribution can be expressed in the canonical form, regression in that distribution can take a simple form (Lancaster, 1958) and correlation coefficient in a contingency table can be estimated (Hamdan & Al-Bayyati, 1971). Also, a bivariate distribution with prescribed correlations and given marginal distributions can be constructed through the notion of canonical expansion (Lancaster, 1958).

The canonical expansion of a bivariate distribution is defined as follows.

Definition 2.5 (*Canonical Expansion*): Let $h(x_1, x_2)$ be a bivariate pdf with marginal pdf's $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ where the parameters have been suppressed for simplicity. Let $\left\{\varphi_i^{(1)}(x_1)\right\}$ and $\left\{\varphi_i^{(2)}(x_2)\right\}$ be complete sets of orthonormal functions with respect to $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ respectively. Then, $h(x_1, x_2)$ can be expanded as a double series

$$h(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \rho_{ij}\, \varphi_i^{(1)}(x_1) \varphi_i^{(2)}(x_2) \tag{2.7}$$

where $\rho_{ij} = \int\int h(x_1, x_2) \varphi_i^{(1)}(x_1) \varphi_i^{(2)}(x_2)\, dx_1\, dx_2$ and $\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \rho_{ij}$ is convergent (Lancaster, 1958). Next, let $\phi^2 = \int \left(\frac{dH(x_1,x_2)}{dF_1(x_1) dF_2(x_2)}\right)^2 dF_1(x_1) dF_2(x_2) - 1$ where $H(x_1, x_2)$, $F_1(x_1)$ and $F_2(x_2)$ are the distribution functions corresponding respectively to $h(x_1, x_2)$, $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. If ϕ^2 is bounded (Lancaster, 1958), then (2.7) can be expressed in the canonical form such that $\rho_{ij} = 0$ for $i \neq j$, that is, the coefficient matrix $[\rho_{ij}]$ is diagonal. The double series (2.7) becomes

$$h(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \sum_{i=0}^{\infty} \rho_i \varphi_i^{(1)}(x_1) \varphi_i^{(2)}(x_2)\,,\ \rho_{ii} = \rho_i \tag{2.8}$$

with $\phi^2 = \sum_{i=1}^{\infty} \rho_i^2$. The series (2.8) is the canonical expansion of the bivariate pdf $h(x_1, x_2)$ and ρ_i is known as the *i*th canonical coefficient or canonical correlation.

Eagleson (1964) has shown that for the Meixner class of distributions, the bivariate distributions obtained from trivariate reduction (2.1) have canonical expansions.

Theorem 2.2 (Eagleson, 1964, p. 1211): If, for a particular distribution,

(i) the orthogonal polynomials are generated by a function of the form $f(t)e^{xu(t)}$ where $f(t)$ is a power series in t with $f(0) = 1$, and $u(t)$ is a power series in t with $u(0) = 0$ and $u'(0) = 1$,

(ii) the distribution is additive and

(iii) a bivariate distribution is generated by using the additive property (2.1),

then the matrix of correlations of the pairs of orthonormal polynomials on the marginals is diagonal. Further, $\rho_{ii} = \rho_i$ depends only on the normalising factor of the ith orthogonal polynomial.

2.3.2 Quadrant Dependence

Joe (1997) has discussed various forms of bivariate dependence which can be used in statistical analysis. One of them is positive quadrant dependence, introduced by Lehmann (1966). This dependence as well as negative quadrant dependence is a very useful measure as it is usually simpler and easier to establish than the other concepts of dependence (Lai & Xie, 2000). Positive and negative quadrant dependences are defined as follows.

Definition 2.6 (*Positive (Negative) Quadrant Dependence*): Two random variables X_1 and X_2 are said to be positive (negative) quadrant dependent if

$$\Pr(X_1 \leq x_1, X_2 \leq x_2) \geq (\leq) \Pr(X_1 \leq x_1) \Pr(X_2 \leq x_2), \forall x_1, x_2. \tag{2.9}$$

Jensen (1971) has extended (2.9) to regions other than quadrant with the concept of positive dependence when the marginal distributions are identical. Jensen's definition of positive dependence is as follows.

Definition 2.7 (*Jensen's Positive Dependence*): Two random variables X_1 and X_2 are said to be positively dependent if $\Pr(X_1 \in A, X_2 \in A) \geq Pr(X_1 \in A) Pr(X_2 \in A)$, for every measurable set A with respect to the marginal measure.

2.4 Parameter Estimation

There are several widely used parameter estimation methods. Among them are the classical method of moments, method of even-points and zero-zero cell frequency technique. The more recent estimation methods include the *M*-estimation, expectation-maximization (EM) and minimum divergence estimations. Above all, the method of maximum likelihood under regularity conditions has proven to be the most preferred method as the estimators are asymptotically unbiased, consistent and efficient. Unfortunately, this method of estimation fails to produce satisfactory estimates in the presence of outliers in the data. Minimum Hellinger distance estimation as well as several other related minimum divergence methods has been proposed as a method which not only overcomes this weakness, but also retains the desirable properties of asymptotic unbiasedness, consistency and efficiency.

2.4.1 Maximum Likelihood Estimation

Let $L(\boldsymbol{\theta};\mathbf{x})$ denote the likelihood function with $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ representing the vector of unknown parameters and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ is the n sets of observed values for the k-variate random variable $\boldsymbol{X} = (X_1, X_2, \dots, X_k)$. The method of maximum likelihood estimation (*MLE*) obtains the estimates of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_d)$, by maximizing $L(\boldsymbol{\theta};\mathbf{x})$ with respect to $\boldsymbol{\theta}$. In other words, the likelihood equations

$$\frac{\partial \ln L(\boldsymbol{\theta};\mathbf{x})}{\partial \theta_i} = 0, \; i = 1,2, \dots, d \tag{2.10}$$

are solved for $\theta_1, \theta_2, \dots, \theta_d$ as $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_d$.

Assuming the following regularity conditions on $L(\boldsymbol{\theta};\mathbf{x})$ (Cox & Hinkley, 1974)

(i) the parameter space Ω has finite dimension, is closed and compact, and the true parameter value is in the interior to Ω,

(ii) the probability distributions defined by any two different values of $\boldsymbol{\theta}$ are distinct, and

(iii) for almost all $\mathbf{x}$, the derivatives $\frac{\partial \ln L(\boldsymbol{\theta};\mathbf{x})}{\partial \theta_i}$, $\frac{\partial^2 \ln L(\boldsymbol{\theta};\mathbf{x})}{\partial \theta_i^2}$ and $\frac{\partial^3 \ln L(\boldsymbol{\theta};\mathbf{x})}{\partial \theta_i^3}$ exist for every θ_i, $i = 1,2, \dots, d$ belonging to a non-degenerate interval A

are satisfied, there exists a sequence of roots of (2.10) that is consistent. However, if there is more than one root of (2.10), it is not known which roots are consistent. Wald (1949) has shown that given certain conditions, the global maximum of (2.10) is consistent.

2.4.2 Minimum Hellinger Distance Estimation

Minimum Hellinger distance estimation (*MHDE*) was first introduced by Beran (1977) and followed up by Tamura & Boos (1986) and Simpson (1987, 1989) among others. The more recent studies on *MHDE* and its related methods include Lindsay (1994), Basu *et al*. (1997) and Basu (2002). *MHDE* has been shown to yield estimators which are asymptotically efficient and relative to *MLE*, attractively robust. *MHD* estimators are also consistent under the correct conditions (Tamura & Boos, 1986; Simpson, 1987).

Let the L^2 norm be denoted by $\|h(t)\|_2 = (\int |h(t)|^2 dt)^{1/2}$. Also, let $f_\theta(\mathbf{x})$, $\mathbf{x} = (x_1, x_2, \dots, x_k)$ denote the pf for a k-variate parametric family $F_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \Omega$ and $\Omega \in \mathbb{R}^d$, where Ω is the parameter space. Let $f_n(\mathbf{x})$ denote the nonparametric density estimate obtained from a random sample $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ of k-vectors. The *MHD* estimate of $\boldsymbol{\theta}$, denoted by $\overline{\boldsymbol{\theta}}$ minimizes the Hellinger distance measure $\left\|f_n^{1/2}(\mathbf{x}) - f_{\boldsymbol{\theta}}^{1/2}(\mathbf{x})\right\|_2^2$. Succinctly,

$$\overline{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta} \in \Omega} \left\|f_n^{1/2}(\mathbf{x}) - f_{\boldsymbol{\theta}}^{1/2}(\mathbf{x})\right\|_2^2 . \tag{2.11}$$

The ease to search for optimum of (2.11) depends on the complexity of the parametric pf involved.

The *MHDE* has been generalized by Basu *et al*. (1997) to obtain the minimum generalized Hellinger divergence (*MGHD*) estimate of $\boldsymbol{\theta}$, denoted by $\widetilde{\boldsymbol{\theta}}$, which minimizes the divergence measure $K_\alpha(1 - \sum_{\mathbf{x}} f_n^\alpha(\mathbf{x}) f_{\boldsymbol{\theta}}^{1-\alpha}(\mathbf{x}))$, $\alpha \in (0,1)$ for count data models. K_α is a nonnegative standardizing constant. Minimizing this divergence measure is equivalent to maximizing the term $\sum_{\mathbf{x}} f_n^\alpha(\mathbf{x}) f_{\boldsymbol{\theta}}^{1-\alpha}(\mathbf{x})$ over $\boldsymbol{\theta} \in \Omega$. This estimation method is also asymptotically efficient and robust.

However, both *MHDE* and *MGHD* perform poorly relative to *MLE* when the sample sizes are small. Basu *et al*. (1997) pointed out that this is mainly due to large weights being accorded to the empty cells in the data set. Thus, they suggested a penalized version of *MGHD* (*MPGHD*) which minimizes the divergence measure

$$D_\alpha = \sum_{\mathbf{x}:f_n(\mathbf{x})\neq 0} \left\{ \frac{f_n(\mathbf{x})}{\alpha(\alpha-1)} \left[\left(\frac{f_n(\mathbf{x})}{f_{\boldsymbol{\theta}}(\mathbf{x})} \right)^{\alpha-1} - 1 \right] + \frac{f_{\boldsymbol{\theta}}(\mathbf{x}) - f_n(\mathbf{x})}{\alpha} \right\} + \sum_{\mathbf{x}:f_n(\mathbf{x})=0} f_{\boldsymbol{\theta}}(\mathbf{x}) \qquad (2.12)$$

over $\boldsymbol{\theta} \in \Omega$. They showed that the *MPGHD* performs just as well as, if not better than, *MHDE* empirically. For bivariate and multivariate discrete cases, there are invariably many empty cells in the data set and thus, *MPGHD* is expected to be a more appropriate method of estimation.

2.4.3 Simulated Annealing

In principle, the problem of determining the globally optimum estimate when there are multiple roots of a given objective function may be resolved by simply finding all the roots and then, choosing the root corresponding to the optimum value of the function. Nevertheless, depending on the complexity of the objective function, this approach may take a considerable amount of time even to find one root, and there is a chance of missing out the root that corresponds to the global optimum. Furthermore, in the case of maximum likelihood estimation, there may be an infinite number of roots of (2.10) (Barnett, 1966).

In the light of the preceding discussion, the simulated annealing (SA) method is employed to estimate the d-dimensional unknown parameter vector $\boldsymbol{\theta}$. This is a discrete optimization method developed in the early 1980s by Kirkpatrick *et al*. (1983) based on a set of ideas put forth by Metropolis *et al*. (1953) in finding an equilibrium point that minimizes the total energy in a system of particles undergoing a change in temperature using statistical thermodynamics. Essentially, this stochastic global search technique approximates the minimum of the objective function f.

SA operates iteratively by choosing a new set of parameter values $\boldsymbol{\theta}'$ from the neighbourhood of the present set of $\boldsymbol{\theta}$ at a temperature T which follows a specified cooling schedule. This $\boldsymbol{\theta}'$ is accepted if the value of f is smaller (downhill move). Uniquely, SA does not reject outright $\boldsymbol{\theta}'$ if the value of f is worse (uphill move). It allows $\boldsymbol{\theta}'$ to be accepted with a positive probability, usually given by the Metropolis acceptance probability (Fouskakis & Draper, 2002)

$$\Pr(\boldsymbol{\theta}, \boldsymbol{\theta}', T) = \begin{cases} 1, & \text{if } f(\boldsymbol{\theta}') \geq f(\boldsymbol{\theta}) \\ e^{\left(\frac{f(\boldsymbol{\theta}') - f(\boldsymbol{\theta})}{T}\right)}, & \text{if } f(\boldsymbol{\theta}') < f(\boldsymbol{\theta}) \end{cases}.$$

This flexibility enables the algorithm to escape from getting stuck in local minima.

Usually, the temperature T is decreased after every m iterations. At higher temperature, the system accepts moves almost randomly, regardless of whether they are uphill or downhill. As the temperature is lowered, the probability of making uphill moves drops and eventually, the system may achieve a globally minimum state when no further moves are accepted. Although this method is initially introduced to solve discrete problems, the SA method has been adapted to solve continuous problems as well (Brooks & Morgan, 1995; Parker, 2000).

The classical SA sometimes suffers a very slow and inefficient convergence to the optimum state. To overcome this, modifications to the cooling schedule and acceptance probability or hybrids with other search methods among other solutions have been proposed by researchers including Szu & Hartley (1987), Ingber (1992), Brooks (1995) and Mendonca & Caloba (1997). In this thesis, the SA combined with the downhill simplex method of Nelder and Mead in the subroutine `amebsa` proposed by

Press *et al*. (1992) is used for all the required function minimizations. The efficiency of this SA is not affected in the narrow valleys of optimization (Press *et al*., 1992). The SA algorithm is outlined as follows.

Algorithm 2.1 (Subroutine `amebsa`, Press *et al*., 1992, p. 445): Outline of SA Combined with Downhill Simplex Method

---

(1) Initialize a $(d+1) \times d$ input matrix **p** of parameter vectors $\boldsymbol{\theta_0}$. The $(d+1)$ rows of d-dimensional vector $\boldsymbol{\theta_0}$ are the vertices of the starting simplex.

(2) Compute the corresponding objective function f for each $(d+1)$ vectors of $\boldsymbol{\theta_0}$ and assign the values to a vector **y** of length $(d+1)$.

(3) Select the initial temperature T_0.

(4) Set $\boldsymbol{\theta}_{optimal} = \mathbf{p}(1,:)$, $y_{optimal} = \mathbf{y}(1)$, $T = T_0$ and stopping criterion = false.

(5) **while** (stopping criterion $\neq$ false)

  a) **do**

    i. Add a random fluctuation to **y** and assign to **yt**.

    ii. Determine the point in **p** with highest (worst), second highest and lowest (best) values in **yt**.

    iii. Compute the fractional range from the highest to lowest value in **yt**.

    iv. **if** satisfactory,

      then put the best point and function value in slot 1 of arrays **p** and **y**;

      RETURN

v. Extrapolate by a factor -1 through the face of the simplex from the highest point, that is, reflect the simplex from the highest point to a new point **ptry** of length d. If the corresponding f value assigned to $ytry$ improves, replace $\boldsymbol{\theta}_{optimal} =$ **ptry** and $y_{optimal} = ytry$.

vi. **if** $ytry$ is smaller than the lowest value in **yt**,

then try an additional extrapolation by a factor of 2 to a new point **ptry**;

if the corresponding f value assigned to $ytry$ improves,

then replace $\boldsymbol{\theta}_{optimal} =$ **ptry** and $y_{optimal} = ytry$;

else if $ytry$ is worse than the second highest value in **yt**,

then look for an intermediate lower point, that is, do a one-dimensional contraction to a new point **ptry**;

if the corresponding f value assigned to $ytry$ improves,

then replace $\boldsymbol{\theta}_{optimal} =$ **ptry** and $y_{optimal} = ytry$;

if $ytry$ is still worse than the highest value in **yt**,

then contract around the lowest point and replace **p**;

Compute the corresponding **y** for the new **p**.

end do

b) **if** stopping criterion is met,

then stopping criterion = true;

else decrease T according to a selected temperature cooling schedule.

end while

(6) $\boldsymbol{\theta}_{optimal}$ is the optimal estimate for $\boldsymbol{\theta}$ with corresponding optimal objective function value of $y_{optimal}$.

2.4.4 Consistency of Estimators

One of the basic large sample properties of estimators is the consistency of the estimators (Newey & McFadden, 1994). An estimator $\widehat{\boldsymbol{\theta}}$ is said to be consistent for the parameter $\boldsymbol{\theta}$ if $\widehat{\boldsymbol{\theta}}$ converges in probability (weak consistency) or if with probability 1 or convergence almost surely (strong consistency) to the true value of the parameter $\boldsymbol{\theta}_0$, that is $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ or $\widehat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as the data sample size $n \to \infty$.

Definition 2.8 (*Uniform Convergence in Probability*): Let $\hat{Q}_n(\boldsymbol{\theta})$ be an objective function that converges uniformly in probability to $Q_0(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Omega$. This implies $\sup_{\boldsymbol{\theta}\in\Omega}\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right| \xrightarrow{P} 0$.

The following is a basic consistency theorem from Newey & McFadden (1994). Strong consistency result holds when $\sup_{\boldsymbol{\theta}\in\Omega}\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0$.

Theorem 2.3 (Newey & McFadden, 1994, p. 2121): If there is a function $Q_0(\boldsymbol{\theta})$ such that

(i) $Q_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$,

(ii) Ω is compact,

(iii) $Q_0(\boldsymbol{\theta})$ is continuous, and

(iv) $\hat{Q}_n(\boldsymbol{\theta})$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$,

then $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$.

The following theorem is needed to establish consistency of the estimators in Chapter 4.

Theorem 2.5 (*Continuous Mapping Theorem*, Athreya & Lahiri, 2006, p. 305): Let $\{X_n\}_{n\geq 1}$, X be random variables such that $X_n \xrightarrow{d} X$. Let $f: \mathbb{R} \to \mathbb{R}$ be Borel measurable such that $\Pr\left(X \in D_f\right) = 0$, where D_f is the set of discontinuities of f. Then, $f(X_n) \xrightarrow{d} f(X)$. In particular, this holds if $f: \mathbb{R} \to \mathbb{R}$ is continuous.

2.5 Hypothesis Testing

This section will briefly review the Pearson's χ^2 goodness-of-fit test and two commonly used parametric tests of hypotheses, namely the likelihood ratio (*LR*) and score tests. In large samples, these tests are asymptotically equivalent. Rao (1973) has examined the problem of constructing these tests for simple and composite hypotheses. Specifically, one-sided *LR* and score tests are described.

2.5.1 Pearson's Chi-Square Goodness-of-Fit Test

Let the cell probabilities be specified functions of $\pi_1(\boldsymbol{\theta}), \pi_2(\boldsymbol{\theta}), \ldots, \pi_k(\boldsymbol{\theta})$ of $d < k-1$ unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$, which is estimated by $\widehat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_d\right)$. Further, the estimates of the cell probabilities will simply be denoted as $\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_k$. Then, the test statistic for the goodness-of-fit test is

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i} = \sum_{i=1}^{k} \frac{\left(\text{Observed}_i - \text{Expected}_i\right)^2}{\text{Expected}_i} \tag{2.13}$$

where n_i is the observed frequency of cell i and $n = \sum_{i=1}^{k} n_i$. Asymptotically, this test statistic is distributed as $\chi^2(k-1-d)$.

The expected frequency in any cell should take a minimum value of 5 to ensure that the test statistic is asymptotically χ^2 distributed. This is done by grouping the expected values of several cells together. However, in large samples, this minimum value can be taken to be as small as one.

2.5.2 Likelihood Ratio and Score Tests

Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ of k-vectors with sample size n be as defined in Section 2.4.1. Let H_0 and H_1 denote null and alternative hypotheses respectively. Then, it is of interest to test

$$H_0: \ \theta_1 = \theta_1^0, \theta_2 = \theta_2^0, \dots, \theta_k = \theta_k^0; \ \theta_{k+1}, \theta_{k+2}, \dots, \theta_d \text{ unspecified}$$

against the two-sided alternative

$$H_1: \ \boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d) \text{ unspecified,}$$

or against the one-sided alternative

$$H_1: \ \theta_1 \geq \theta_1^0, \theta_2 \geq \theta_2^0, \dots, \theta_k \geq \theta_k^0; \ \theta_{k+1}, \theta_{k+2}, \dots, \theta_d \text{ unspecified.}$$

- *Likelihood Ratio (LR) Test*

For both one-sided and two-sided tests, the *LR* incorporates the information from both the null and alternative models for the data **x** and gives

$$\lambda = \frac{\sup_{\boldsymbol{\theta} \in \Omega_{H_0}} L(\boldsymbol{\theta}; \mathbf{x})}{\sup_{\boldsymbol{\theta} \in \Omega_{H_1}} L(\boldsymbol{\theta}; \mathbf{x})}$$

where the supremums in the numerator and denominator are evaluated under the parameter spaces of H_0 and H_1, denoted by Ω_{H_0} and Ω_{H_1}, respectively (Rao, 1973).

The test statistic for *LR* test is then taken as

$$LR_T = -2 \ln \lambda = -2 \ln \left(\frac{L(\widehat{\boldsymbol{\theta}}^*; \mathbf{x})}{L(\widehat{\boldsymbol{\theta}}; \mathbf{x})} \right) \tag{2.14}$$

where $\widehat{\boldsymbol{\theta}}^* = \left(\theta_1^0, \theta_2^0, \ldots, \theta_k^0, \hat{\theta}_{k+1}, \hat{\theta}_{k+2}, \ldots, \hat{\theta}_d\right)$ and $\widehat{\boldsymbol{\theta}}$ are the maximum likelihood estimates under H_0 and H_1, respectively. For a two-sided alternative, the given statistic (2.14) has, asymptotically, a χ^2 distribution with k degrees of freedom. This distribution will concisely be denoted as χ_k^2. For a one-sided alternative, Self & Liang (1987) gave the asymptotic null distribution for the test statistic under certain regularity conditions (see Theorem 3, Self & Liang, 1987, p. 607). In particular, the asymptotic null distribution for the statistic $-2 \ln \lambda$ is a 50:50 mixture of χ_0^2 and χ_1^2 when $k = 1$ (see Case 5, Self & Liang, 1987, p. 608). Here, the convention that the central χ_0^2 is identically zero resulting in all probability mass at zero is adopted. When $L(\widehat{\boldsymbol{\theta}}; \mathbf{x})$ under the one-sided H_1 is maximized at $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^*$, we will have the test statistic $LR_T = 0$, yielding a p-value of 1.00.

- *Score Test*

The score test statistic is slightly different between one-sided and two-sided cases. Let the ith efficient score of Rao (1973) be

$$u_i(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \frac{\partial \ln L(\boldsymbol{\theta}; \mathbf{x})}{\partial \theta_i}, \; i = 1,2, \ldots, d$$

and n is the sample size of data $\mathbf{x}$.

The score test for the two-sided composite hypothesis is given by

$S_c = \mathbf{U}^{*T} (\boldsymbol{\Gamma}^*)^{-1} \mathbf{U}^*$, where

$$\mathbf{U}^{*T} = [u_1(\boldsymbol{\theta}), u_2(\boldsymbol{\theta}), \ldots, u_d(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^*} \text{ and } \boldsymbol{\Gamma}^* = \left\{\Gamma_{ij}^*\right\} = \frac{1}{n}\left\{E\left[-\frac{\partial^2 \ln L(\boldsymbol{\theta}; \mathbf{x})}{\partial \theta_i \partial \theta_j}\right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^*}\right\}$$

with $i, j = 1,2, \ldots, d$ being the information matrix of a single observation for $\boldsymbol{\theta}$ evaluated under $\widehat{\boldsymbol{\theta}}^*$, which is the maximum likelihood estimates under H_0. This test statistic is also asymptotically distributed as a χ_k^2.

For one-sided test, let $\vartheta_i = \theta_i - \theta_i^0$, $i = 1,2,\dots,k$. Then, H_0 and H_1 become

H_0 : $\vartheta_1 = 0, \vartheta_2 = 0, \dots, \vartheta_k = 0$; $\theta_{k+1}, \theta_{k+2}, \dots, \theta_d$ unspecified,

H_1 : $\vartheta_1 \geq 0, \vartheta_2 \geq 0, \dots, \vartheta_k \geq 0$; $\theta_{k+1}, \theta_{k+2}, \dots, \theta_d$ unspecified.

Let $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\varphi})$ with $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_k)$ and $\boldsymbol{\varphi} = (\theta_{k+1}, \theta_{k+2}, \dots, \theta_d)$, the score statistic is then defined as (Silvapulle & Silvapulle, 1995)

$$S_{c_1} = \mathbf{U}_\vartheta^{*T} (\boldsymbol{\Gamma}_{\vartheta\vartheta}^*)^{-1} \mathbf{U}_\vartheta^* - \inf_{\mathbf{b} \in \Omega_{H_1}} \{(\mathbf{U}_\vartheta^* - \mathbf{b})^T (\boldsymbol{\Gamma}_{\vartheta\vartheta}^*)^{-1} (\mathbf{U}_\vartheta^* - \mathbf{b})\} \tag{2.15}$$

where $\mathbf{U}_\vartheta^* = [u_1(\boldsymbol{\theta}), u_2(\boldsymbol{\theta}), \dots, u_k(\boldsymbol{\theta})]_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_0^*}$ and $\boldsymbol{\Gamma}_{\vartheta\vartheta}^* = \{\Gamma_{ij}^*\}$, $i, j = 1,2,\dots,k$ with $\widehat{\boldsymbol{\theta}}_0^* = (\boldsymbol{\vartheta}_0, \widehat{\boldsymbol{\varphi}})$, $\boldsymbol{\vartheta}_0 = \mathbf{0}$ and $\widehat{\boldsymbol{\varphi}}$ is the *ML* estimate of $\boldsymbol{\varphi}$ under H_0. This statistic has asymptotically a chi-bar-squared $\bar{\chi}^2$ distribution which is a mixture of χ^2 distributions. Readers are referred to the paper by Shapiro (1988) for the detailed discussion on this asymptotic distribution. Nevertheless, it is sufficient for the purpose of this thesis to state that when there is only one specified parameter $(k = 1)$ under H_0, the *p*-value of the test is

$$\Pr(S_{c_1} \geq sc) = \frac{1}{2}\Pr(\chi_0^2 \geq sc) + \frac{1}{2}\Pr(\chi_1^2 \geq sc) \tag{2.16}$$

with sc being the observed value for S_{c_1}. Again, the convention that χ_0^2 has all probability mass at zero is adopted. If the *ML* estimate of $\boldsymbol{\vartheta}$ under unconstrained parameterization is $\widehat{\boldsymbol{\vartheta}} \leq \boldsymbol{\vartheta}_0$, then the infimum of the second term on the right hand side of (2.15) is achieved when $\mathbf{b} = \mathbf{0}$, yielding $S_{c_1} = 0$ and a *p*-value of 1.00. Silvapulle & Silvapulle (1995) have shown that the large-sample *p*-value of the test is bounded by $\frac{1}{2}\Pr(\chi_1^2 \geq sc) \leq p\text{-value} \leq \frac{1}{2}\Pr(\chi_{k-1}^2 \geq sc) + \frac{1}{2}\Pr(\chi_k^2 \geq sc)$.

CHAPTER 3 : EXTENSION OF A CLASS OF BIVARIATE MEIXNER DISTRIBUTIONS

3.0 Introduction

Univariate binomial, gamma, Meixner hypergeometric, negative binomial, normal and Poisson distributions are members of the Meixner class of distributions (Meixner, 1934). The Meixner class of distributions has a number of interesting characterizations and properties as described in Chapter 1. In particular, joint distributions of the Meixner class formed by random elements in common have canonical expansions and this property has been used to characterize the Meixner class of distributions (Eagleson & Lancaster, 1967). Lancaster (1975) has examined the convolutions of these joint Meixner distributions. The bivariate counterparts of some of the distributions in this Meixner class have also been researched by others (see Mardia, 1970; Kocherlakota & Kocherlakota, 1992). However, most of these bivariate distributions have been formulated in such a way that the marginal distributions have at least one parameter in common. For example, the popular bivariate negative binomial distribution of Edwards & Gurland (1961), formulated as a bivariate mixed Poisson distribution, has the probability generating function (pgf) for the joint random variables (rv's) (X_1, X_2) given by

$$G_{(X_1,X_2)}(z_1, z_2) = \left(\frac{\Theta}{1 - \theta_1 z_1 - \theta_2 z_2 - \theta_3 z_1 z_2}\right)^{\nu}, \quad \Theta = 1 - \theta_1 - \theta_2 - \theta_3$$

with marginal pgf's as

$$G_{X_i}(z) = \left(\frac{1 - p_i}{1 - p_i z}\right)^{\nu}, \qquad i = 1,2$$

where $p_1 = \frac{\theta_1+\theta_3}{1-\theta_2}$ and $p_2 = \frac{\theta_2+\theta_3}{1-\theta_1}$. Note that the marginals have the same index parameter, ν.

In this chapter, an extension of the Meixner class of bivariate distributions of Eagleson (1964) to bivariate distributions having marginal distributions with different parameters is introduced. Some of these distributions are known with the exception of the extended bivariate negative binomial distribution. This extended bivariate negative binomial distribution will be considered in detail. Specifically, a sufficient condition for the joint probability function to have a canonical expansion is given. Also, the application of this extended class in the computer generation of bivariate samples given the marginal distributions and correlation are examined. Multivariate extensions of the distributions have also been derived. Numerical illustrations are given at the end of the chapter to demonstrate the viability of this family of distributions.

Note that, for the entire thesis, $X \sim D(\boldsymbol{\theta})$ will refer to the rv X being distributed as a distribution D with parameter vector $\boldsymbol{\theta}$.

3.1 Formulation of Bivariate Distributions

3.1.1 Extended Trivariate Reduction

The method of trivariate reduction may be extended in the following manner.

Definition 3.1 (*Extension of Trivariate Reduction Method*): Given independent Y_1 and Y_2 rv's, consider

$$X_1 = Y_1 + W_1 \text{ and } X_2 = Y_2 + W_2 \tag{3.1}$$

where (W_1, W_2) is a pair of randomly correlated elements independent of Y_1 and Y_2. Then, the joint rv's (X_1, X_2) has a distribution formed by the extended trivariate reduction method.

The common random element W in the usual trivariate reduction is now replaced by the joint rv's (W_1, W_2). Lai (1995) has examined the situation where $W_1 = I_1 W$ and $W_2 = I_2 W$, with I_1 and I_2 being indicator rv's such that (I_1, I_2) has a joint probability distribution $(p_{00}, p_{01}, p_{10}, p_{11})$.

For the genesis by (3.1), the general form of the bivariate pgf is given by

$$G_{(X_1,X_2)}(z_1, z_2) = G_{Y_1}(z_1; \boldsymbol{\theta}_1) G_{Y_2}(z_2; \boldsymbol{\theta}_2) G_{(W_1,W_2)}(z_1, z_2; \boldsymbol{\theta}) \quad (3.2)$$

where G_{Y_1}, G_{Y_2}, and $G_{(W_1,W_2)}$ are the corresponding pgf of Y_1, Y_2 and (W_1, W_2) with parameter vectors $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}$ respectively. Y_1, Y_2, W_1 and W_2 are from the same family of univariate discrete distributions with (W_1, W_2) being jointly distributed. For continuous cases, the pgf is replaced by the moment generating function (mgf). Note that if (W_1, W_2) has a distribution formed by trivariate reduction, (X_1, X_2) again has a distribution formed by the usual trivariate reduction.

The model (3.1) can be easily applied to simulate samples with specified marginals and correlation from (X_1, X_2) as follows.

(1) Generate y_1 and y_2 from specified distributions Y_1 and Y_2.

(2) Use known joint distribution (W_1, W_2) as well as correlation to generate (w_1, w_2).

(3) $x_1 = y_1 + w_1$ and $x_2 = y_2 + w_2$.

a) Bivariate Binomial Distribution

The formulation (3.1) has been used by Hamdan & Jensen (1976) to generalize the then existing bivariate binomial ($BBinomial$) distribution to obtain one with different index and probability parameters. They let $Y_1 \sim Binomial(n_1 - k, p_1)$, $Y_2 \sim Binomial(n_2 - k, p_2)$ and $(W_1, W_2) \sim BBinomial(k, p_{00}, p_{01}, p_{10}, p_{11})$ be the bivariate binomial random variables with joint pgf $(p_{00} + p_{10}z_1 + p_{01}z_2 + p_{11}z_1z_2)^k$, where k is an integer such that $k < \min(n_1, n_2)$. Then, by (3.2), the joint pgf for bivariate binomial (X_1, X_2) distribution is

$$G_{(X_1,X_2)}(z_1, z_2) = (q_1 + p_1z_1)^{n_1-k}(q_2 + p_2z_2)^{n_2-k}(p_{00} + p_{10}z_1 + p_{01}z_2 + p_{11}z_1z_2)^k$$

with the marginal pgf's

$$G_{X_i}(z) = (q_i + p_iz)^{n_i},\ i = 1,2$$

where $p_1 = p_{10} + p_{11},\ p_2 = p_{01} + p_{11},\ q_i = 1 - p_i,\ i = 1,2$ and $p_{00} + p_{01} + p_{10} + p_{11} = 1$. Also, $0 \leq p_1, p_2, p_{00}, p_{01}, p_{10}, p_{11} \leq 1$. Thus, the marginals are $X_i \sim Binomial(n_i, p_i),\ i = 1,2$. This distribution has been considered by Hamdan & Jensen (1976) with applications to statistical quality control.

b) Bivariate Poisson Distribution

Bivariate Poisson ($BPoisson$) distribution obtained through the usual trivariate reduction (2.1) already results in a distribution with different marginal parameters (Holgate, 1964). Now, let $Y_1 \sim Poisson(\lambda_1)$ and $Y_2 \sim Poisson(\lambda_2)$. As noted in Section 3.1.1, (X_1, X_2) from formulation (3.1) with $(W_1, W_2) \sim BPoisson(\lambda_3, \lambda_4, \lambda_5)$ formed from (2.1) will also have a bivariate Poisson distribution from the usual trivariate reduction and the joint pgf is

$$G_{(X_1,X_2)}(z_1, z_2) = e^{\left((\lambda_1+\lambda_3)(z_1-1)+(\lambda_2+\lambda_4)(z_2-1)+\lambda_5(z_1-1)(z_2-1)\right)}$$

with the marginal pgf's

$$G_{X_1}(z) = e^{(\lambda_1+\lambda_3)(z-1)} \text{ and } G_{X_2}(z) = e^{(\lambda_2+\lambda_4)(z-1)}.$$

Thus, $X_1 \sim Poisson(\lambda_1 + \lambda_3)$ and $X_2 \sim Poisson(\lambda_2 + \lambda_4)$.

c) Bivariate Gamma Distribution

The bivariate gamma ($B\Gamma$) distribution arising from the usual trivarite reduction method has been introduced by Cherian (1941) (cited in Mardia, 1970). This $B\Gamma$ distribution has different index parameters but the same scale parameter of unity. On the other hand, a $B\Gamma$ distribution with different parameters has been obtained by Gupta (1979) through a more mathematically involved formulation. Gupta obtained the distribution by first generalizing an 'indirect method' of Bennett & Rice (1934) (cited in Gupta, 1979) using the Fourier transform method and then, specializing two arbitrary functions (instantaneous nonlinearities) to Dirac delta functionals given the first result.

Alternatively, Gupta's (1979) generalization of the $B\Gamma$ distribution may be obtained by using the simpler formulation (3.1). Let $Y_1 \sim \Gamma(\alpha_1, \beta_1)$, $Y_2 \sim \Gamma(\alpha_2, \beta_2)$ and $(W_1, W_2) \sim B\Gamma(\nu, \beta_1, \beta_2, \rho)$ of Wicksell-Kibble (Kibble, 1941). Then, the joint mgf of bivariate gamma (X_1, X_2) is

$$M_{(X_1,X_2)}(t_1, t_2) = \left(1 - \frac{t_1}{\beta_1}\right)^{-\alpha_1} \left(1 - \frac{t_2}{\beta_2}\right)^{-\alpha_2} \left[\left(1 - \frac{t_1}{\beta_1}\right)\left(1 - \frac{t_2}{\beta_2}\right) - \frac{\rho t_1 t_2}{\beta_1 \beta_2}\right]^{-\nu}.$$

The marginals which have different index and scale parameters are $X_i \sim \Gamma(\alpha_i + \nu, \beta_i)$, $i = 1,2$ with mgf's as $M_{X_i}(t) = (1 - t_i/\beta_i)^{-(\alpha_i + \nu)}$, $i = 1,2$.

d) Bivariate Normal Distribution

Pearson (1897) has used the method of trivariate reduction (2.1) to obtain the bivariate normal (BN) distribution. Let $Y_1 \sim N(\mu_1, \sigma_1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$ and $(W_1, W_2) \sim BN(\mu_3, \mu_4, \sigma_3^2, \sigma_4^2, \rho)$. Then, the joint mgf for bivariate normal (X_1, X_2) distribution by (3.1) is

$$M_{(X_1,X_2)}(t_1, t_2) = e^{\left((\mu_1+\mu_3)t_1+(\mu_2+\mu_4)t_2+\frac{1}{2}[(\sigma_1^2+\sigma_3^2)t_1^2+2\rho\sigma_3\sigma_4 t_1 t_2+(\sigma_2^2+\sigma_4^2)t_2^2]\right)}$$

where ρ is the correlation between X_1 and X_2. The mgf's for the marginals are

$$M_{X_1}(t) = e^{\left((\mu_1+\mu_3)t+\frac{1}{2}(\sigma_1^2+\sigma_3^2)t^2\right)} \text{ and } M_{X_2}(t) = e^{\left((\mu_2+\mu_4)t+\frac{1}{2}(\sigma_2^2+\sigma_4^2)t^2\right)}.$$

Therefore, $X_1 \sim N(\mu_1 + \mu_3, \sigma_1^2 + \sigma_3^2)$ and $X_2 \sim N(\mu_2 + \mu_4, \sigma_2^2 + \sigma_4^2)$.

e) Bivariate Negative Binomial Distribution

- **(W_1, W_2) distributed as bivariate negative binomial (*BNB*) distribution of Edwards & Gurland (1961), Subrahmaniam (1966)**

Let $Y_1 \sim NB(\alpha_1, p_1)$, $Y_2 \sim NB(\alpha_2, p_2)$ and $(W_1, W_2) \sim BNB(\nu, \theta_1, \theta_2, \theta_3)$ be the Edwards and Gurland's *BNB* or compound correlated bivariate Poisson distribution with joint pgf

$$G_{(W_1,W_2)}(z_1, z_2) = \left(\frac{\Theta}{1 - \theta_1 z_1 - \theta_2 z_2 - \theta_3 z_1 z_2}\right)^{\nu}, \quad \Theta = 1 - \theta_1 - \theta_2 - \theta_3.$$

Note that this *BNB* distribution has a correlation coefficient, $\rho_{(W_1,W_2)}$ in the range

$$0 \le \rho_{(W_1,W_2)} = \frac{(\theta_3 + \theta_1\theta_2)}{\sqrt{(1-\theta_1)(1-\theta_2)(\theta_1+\theta_3)(\theta_2+\theta_3)}} \le 1.$$

Then, (X_1, X_2) is an extended bivariate negative binomial (*EBNB*) with joint pgf

$$G_{(X_1,X_2)}(z_1, z_2) = \left(\frac{q_1}{1 - p_1 z_1}\right)^{\alpha_1} \left(\frac{q_2}{1 - p_2 z_2}\right)^{\alpha_2} \left(\frac{\Theta}{1 - \theta_1 z_1 - \theta_2 z_2 - \theta_3 z_1 z_2}\right)^{\nu} \tag{3.3}$$

where $p_1 = \frac{\theta_1+\theta_3}{1-\theta_2}$, $p_2 = \frac{\theta_2+\theta_3}{1-\theta_1}$, $q_i = 1 - p_i$, $i = 1,2$ and $\Theta = 1 - \theta_1 - \theta_2 - \theta_3$. The marginal pgf's of X_1 and X_2 are given by

$$G_{X_i}(z) = \left(\frac{q_i}{1 - p_i z_i}\right)^{\alpha_i + \nu}, i = 1,2.$$

That is, $X_1 \sim NB(\alpha_1 + \nu, p_1)$ and $X_2 \sim NB(\alpha_2 + \nu, p_2)$. The distribution of (X_1, X_2) with pgf (3.3) will be denoted as *EBNB-I*. This *EBNB-I* has a correlation in the range $0 \leq \rho_{(X_1,X_2)} \leq 1$ as seen from equation (3.10) because the term $(\theta_3 + \theta_1\theta_2)$ is nonnegative.

- **(W_1, W_2) distributed as *BNB* distribution of Mitchell & Paulson (1981)**

Let $Y_1 \sim NB\left(\alpha_1, \frac{\phi_1}{1+\phi_1}\right)$ and $Y_2 \sim NB\left(\alpha_2, \frac{\phi_2}{1+\phi_2}\right)$. Also, let Mitchell and Paulson's *BNB* be denoted as $(W_1, W_2) \sim BNB(\nu, \theta_1, \theta_2, a, b, c)$ with joint pgf

$$G_{(W_1,W_2)}(z_1, z_2) = \left[\left(1 + \tau_1(1 - z_1)\right)\left(1 + \tau_2(1 - z_2)\right) - d\right]^{-\nu} \cdot \left(a + \frac{b}{1 + \phi_1(1 - z_1)} + \frac{c}{1 + \phi_2(1 - z_2)}\right)^{\nu}$$

where $\tau_i = \theta_i/(1 - \theta_i)$, $i = 1,2$, $\phi_1 = \tau_1/(a + c)$, $\phi_2 = \tau_2/(a + b)$ and $d = 1 - a - b - c$ with each parameter being nonnegative, $b + d < 1$ and $c + d < 1$. Mitchell & Paulson (1981) have introduced this generalization of the *BNB* so that its correlation coefficient

$$\rho_{(W_1,W_2)} = \frac{ad - bc}{1 - d}\left(\frac{\phi_1\phi_2}{(1 + \phi_1)(1 + \phi_2)}\right)^{\frac{1}{2}}$$

covers the full range, giving $-1 \leq \rho_{(W_1,W_2)} \leq 1$.

By (3.2) this extended *EBNB* distribution has pgf

$$G_{(X_1,X_2)}(z_1, z_2) = \left(1 + \phi_1(1 - z_1)\right)^{-\alpha_1}\left(1 + \phi_2(1 - z_2)\right)^{-\alpha_2} \cdot \left[\left(1 + \tau_1(1 - z_1)\right)\left(1 + \tau_2(1 - z_2)\right) - d\right]^{-\nu} \cdot \left(a + \frac{b}{1 + \phi_1(1 - z_1)} + \frac{c}{1 + \phi_2(1 - z_2)}\right)^{\nu} \tag{3.4}$$

The marginal pgf's of X_1 and X_2 are given by

$$G_{X_i}(z) = \left(1 + \phi_i(1 - z)\right)^{-(\alpha_i + \nu)},\ i = 1,2.$$

Therefore, $X_1 \sim NB(\alpha_1 + \nu, \phi_1/(1 + \phi_1))$ and $X_2 \sim NB(\alpha_2 + \nu, \phi_2/(1 + \phi_2))$. This distribution of (X_1, X_2) will be denoted as *EBNB-II*. The correlation of *EBNB-II* is $-1 \leq \rho_{(X_1,X_2)} \leq 1$ because $\rho_{(W_1,W_2)}$ can take any value in the range $[-1, 1]$ in equation (3.10).

As a special case, when $b = c = 0$, (W_1, W_2) is distributed as the *BNB* distribution of Edwards & Gurland (1961) which has been considered earlier. In addition, when $\alpha_1 = \alpha_2 = b = c = 0$, (X_1, X_2) is then the *BNB* of Edwards and Gurland.

3.1.2 Mixed Poisson Formulation

An extension to the compounding technique elucidated in Definition 2.3 is by having

$$h(x_1, x_2; \xi) = \iint f_{X_1}(x_1|\psi_1) f_{X_2}(x_2|\psi_2) g(\psi_1, \psi_2; \xi) d\psi_1 d\psi_2 \qquad (3.5)$$

where ψ_1 and ψ_2 are jointly distributed with pdf $g(\psi_1, \psi_2; \xi)$ (see, for example, Ong, 1993). The *EBNB* distribution can be formulated as a mixed Poisson distribution given by (3.5).

The bivariate gamma distribution considered by Gupta (1979) has mgf

$$M_{(U,V)}(t_1, t_2) = \left(\frac{\beta_1}{\beta_1 - t_1}\right)^{\alpha_1 - \nu} \left(\frac{\beta_2}{\beta_2 - t_2}\right)^{\alpha_2 - \nu} \left[\left(\frac{\beta_1 - t_1}{\beta_1}\right)\left(\frac{\beta_2 - t_2}{\beta_2}\right) - \frac{\rho^2 t_1 t_2}{\beta_1 \beta_2}\right]^{-\nu} \qquad (3.6)$$

where $0 \leq \rho^2 \leq 1$ and $\nu < \min(\alpha_1, \alpha_2)$. It may be formulated by the extended trivariate reduction (3.1) where $Y_1 \sim \Gamma(\alpha_1 - \nu, \beta_1)$ and $Y_2 \sim \Gamma(\alpha_2 - \nu, \beta_2)$ are independent gamma rv's and $(W_1, W_2) \sim B\Gamma(\nu, \beta_1, \beta_2, \rho)$ has the Wicksell-Kibble's bivariate gamma distribution (Kibble, 1941).

The mixed Poisson formulation is as follows:

Suppose $X_1|U \sim Poisson(U)$ and $X_2|V \sim Poisson(V)$ where U and V have a joint bivariate gamma distribution given by (3.6). Then, the unconditional (X_1, X_2) has the *EBNB* distribution.

The formulation is easily proved by using the following relation between the mgf of the mixing distribution and the pgf of the mixed distribution (see Ong, 1990)

$$G_{(X_1,X_2)}(z_1, z_2) = M_{(U,V)}(z_1 - 1, z_2 - 1).$$

This leads to the *EBNB-I* pgf

$$G_{(X_1,X_2)}(z_1, z_2) = \left(\frac{q_1}{1 - p_1 z_1}\right)^{\alpha_1 - \nu} \left(\frac{q_2}{1 - p_2 z_2}\right)^{\alpha_2 - \nu} \left(\frac{\Theta}{1 - \theta_1 z_1 - \theta_2 z_2 - \theta_3 z_1 z_2}\right)^{\nu} \tag{3.7}$$

where $\nu < \min(\alpha_1, \alpha_2)$, $p_i = 1/(1 + \beta_i)$, $q_i = 1 - p_i$, $i = 1,2$,

$$\theta_1 = \frac{p_1(1 - \rho^2 p_2)}{1 - \rho^2 p_1 p_2}, \theta_2 = \frac{p_2(1 - \rho^2 p_1)}{1 - \rho^2 p_1 p_2}, \theta_3 = \frac{-p_1 p_2(1 - \rho^2)}{1 - \rho^2 p_1 p_2}$$

and $\Theta = 1 - \theta_1 - \theta_2 - \theta_3 = q_1 q_2/(1 - \rho^2 p_1 p_2)$. Note that $-1 < \theta_3 < 0$ with $\theta_3 + \theta_1\theta_2 > 0$. Rewriting in terms of θ_1, θ_2 and θ_3, $p_1 = \frac{\theta_1 + \theta_3}{1 - \theta_2}$ and $p_2 = \frac{\theta_2 + \theta_3}{1 - \theta_1}$. The marginal pgf's of X_1 and X_2 are given by

$$G_{X_1}(z) = \left(\frac{q_1}{1 - p_1 z_1}\right)^{\alpha_1} \text{ and } G_{X_2}(z) = \left(\frac{q_2}{1 - p_2 z_2}\right)^{\alpha_2}.$$

That is, $X_1 \sim NB(\alpha_1, p_1)$ and $X_2 \sim NB(\alpha_2, p_2)$. As a special case of this *EBNB-I*, the *BNB* is included with (U, V) having the Wicksell-Kibble's bivariate gamma distribution (see Ong, 1990), that is when $\alpha_1 = \alpha_2 = \nu$ in (3.6) and (3.7).

3.2 Extension to Multivariate Distributions

The formulation $X_1 = Y_1 + W_1$ and $X_2 = Y_2 + W_2$ can be extended to develop multivariate distributions. By taking $X_1 = Y_1 + W_1,\ X_2 = Y_2 + W_2, \dots, X_k = Y_k + W_k$, multivariate distributions with different parameters for the marginals can be obtained. Shown below are two examples of multivariate distributions formed by this method.

- **Multivariate negative binomial distribution (*MNB*)**

Let $X_1 = Y_1 + W_1,\ X_2 = Y_2 + W_2, \dots, X_k = Y_k + W_k$ where $Y_1, Y_2, \dots, Y_k$ are independent *NB* rv's with pgf's

$$G_{Y_i}(z_i) = [q_i/(1 - p_i z_i)]^{\alpha_i}, \qquad i = 1, 2, \dots, k$$

and $\mathbf{W} = (W_1, W_2, \dots, W_k)$ is distributed as *MNB* rv's with pgf

$$G_{\mathbf{W}}(z_1, z_2, \dots, z_k) = \left[\Theta \middle/ \left(1 - \sum_{i=1}^{k} \theta_i z_i - \sum_{1 \le i < j \le k} \theta_{ij} z_i z_j - \cdots - \theta_{12\dots k} z_1 z_2 \dots z_k\right)\right]^{\nu}$$

where $\Theta = 1 - \sum_{i=1}^{k} \theta_i - \sum_{1 \le i < j \le k} \theta_{ij} - \cdots - \theta_{12\dots k}$. Thus, $\mathbf{X} = (X_1, X_2, \dots, X_k)$ is distributed as *MNB* with pgf

$$G_{\mathbf{X}}(z_1, z_2, \dots, z_k) = \prod_{i=1}^{k} \left[\frac{q_i}{1 - p_i z_i}\right]^{\alpha_i}$$
$$\cdot \left[\Theta \middle/ \left(1 - \sum_{i=1}^{k} \theta_i z_i - \sum_{1 \le i < j \le k} \theta_{ij} z_i z_j - \cdots - \theta_{12\dots k} z_1 z_2 \dots z_k\right)\right]^{\nu}$$

with the marginals given by

$$G_{X_i}(z) = [q_i/(1-p_i z)]^{\alpha_i+\nu}$$

where $p_i = \phi_i/(\Theta + \phi_i)$, $q_i = 1 - p_i$, $i = 1,2,\dots,k$ and

$$\phi_i = \theta_i + \sum_{\substack{j=1 \\ j\neq i}}^{k} \theta_{ij} + \sum_{\substack{1\le r<s\le k \\ r\neq i, s\neq i}} \theta_{irs} + \cdots + \theta_{12\dots k}\,.$$

Therefore, each of the marginals $X_i \sim NB(\alpha_i+\nu, p_i)$, $i = 1,2,\dots,k$ will have different parameters.

- **Multivariate Binomial Distribution (*MBinomial*)**

Let $Y_1, Y_2, \dots, Y_k$ be independent binomial rv's with pgf's

$$G_{Y_i}(z_i) = (q_i + p_i z_i)^{n_i - n},\ \ i = 1,2,\dots,k, n < \min(n_1, n_2, \dots, n_k)$$

and $\mathbf{W} = (W_1, W_2, \dots, W_k)$ is distributed as *MBinomial* rv's with pgf

$$G_{\mathbf{W}}(z_1, z_2, \dots, z_k) = \left(\Theta + \sum_{i=1}^{k} \theta_i z_i + \sum_{1\le i<j\le k} \theta_{ij} z_i z_j + \cdots + \theta_{12\dots k} z_1 z_2 \dots z_k\right)^n$$

where $\Theta = 1 - \sum_{i=1}^{k} \theta_i - \sum_{1\le i<j\le k} \theta_{ij} - \cdots - \theta_{12\dots k}$. Thus, $\mathbf{X} = (X_1, X_2, \dots, X_k)$ is distributed as *MBinomial* with pgf

$$G_{\mathbf{X}}(z_1, z_2, \dots, z_k) = \prod_{i=1}^{k} (q_i + p_i z_i)^{n_i - n} \cdot \left(\Theta + \sum_{i=1}^{k} \theta_i z_i + \sum_{1\le i<j\le k} \theta_{ij} z_i z_j + \cdots + \theta_{12\dots k} z_1 z_2 \dots z_k\right)^n$$

with the marginals given by

$$G_{X_i}(z) = (q_i + p_i z)^{n_i}$$

where $p_i = \phi_i/(\Theta + \phi_i)$, $q_i = 1 - p_i$, $i = 1,2,\dots,k$ and

$$\phi_i = \theta_i + \sum_{\substack{j=1 \\ j\neq i}}^{k} \theta_{ij} + \sum_{\substack{1\leq r<s\leq k \\ r\neq i, s\neq i}} \theta_{irs} + \cdots + \theta_{12\dots k}\ .$$

Therefore, each of the marginals $X_i \sim Binomial(n_i, p_i)$, $i = 1,2,\dots,k$ will have different parameters.

3.3 Canonical Expansion of Bivariate Distributions Formed by Extended Trivariate Reduction

Members of the Meixner class of distributions, considered as weights for orthogonal polynomials, have generating functions for their corresponding orthogonal polynomials of the form $G(t,x) = f(t)e^{xu(t)}$. Eagleson (1964) has shown that their bivariate distributions obtained from trivariate reduction (2.1) have canonical expansions (Barrett & Lampard, 1955; Lancaster, 1958). The following result extends Eagleson's result to bivariate distributions which can be formed by the extended trivariate reduction explained in Section 3.1.1.

Result 3.1: If (W_1, W_2) has a bivariate distribution with canonical expansion in terms of orthogonal polynomials, then another bivariate distribution (X_1, X_2) generated using the additive property $X_1 = Y_1 + W_1$ and $X_2 = Y_2 + W_2$, where Y_1 and Y_2 are independent, also has a canonical expansion in terms of orthogonal polynomials.

Proof:

Let $\xi_r^*(u)$ denote the rth orthonormal polynomial of the rth orthogonal polynomial $\xi_r(u)$ on a distribution U and $c_r^{(U)} = \int_{-\infty}^{\infty} \xi_r^2(u)dF(u)$ where $F(u)$ is the distribution function of U. Also, let $\rho_{rs}^{(U,V)}$ denote the correlation of a (r,s) pair of such orthonormal polynomials on the marginals of the bivariate distribution (U, V), that is $\rho_{rs}^{(U,V)} = E[\xi_r^*(u)\xi_s^*(v)]$. Then, by extending Theorem 2.2,

$$\left[c_r^{(X_1)} c_s^{(X_2)}\right]^{\frac{1}{2}} \rho_{rs}^{(X_1,X_2)}$$

$$= \int \xi_r(x_1)\xi_s(x_2)dF(x_1, x_2)$$

$$= \iiint \left(\sum_{i=0}^{r} \binom{r}{i} \xi_i(y_1)\xi_{r-i}(w_1)\right)\left(\sum_{j=0}^{s} \binom{s}{j} \xi_j(y_2)\xi_{s-j}(w_2)\right) dF_1(y_1)dF_2(y_2)dF_{12}(w_1, w_2)$$

$$= \sum_{i=0}^{r}\sum_{j=0}^{s} \binom{r}{i}\binom{s}{j} \left[\iint \xi_i(y_1)\xi_j(y_2)dF_1(y_1)dF_2(y_2)\right]\left[\int \xi_{r-i}(w_1)\xi_{s-j}(w_2)dF_{12}(w_1, w_2)\right].$$

Using a corollary from Lancaster (1963, p. 535) [which states that a necessary and sufficient condition for independence of the marginal variables of a bivariate statistical distribution is that $\rho_{ij} = 0$, for $i > 0$ and $j > 0$] and Theorem A also from Lancaster (1963, p. 532) [which states $\int \xi_i(u)\, \xi_j(v)dF(u, v) = 0$ if $i \neq j$], then

$$\left[c_r^{(X_1)} c_s^{(X_2)}\right]^{\frac{1}{2}} \rho_{rs}^{(X_1,X_2)} = \int \xi_r(w_1)\xi_s(w_2)dF_{12}(w_1, w_2)$$

$$= \delta_{rs} \left[c_r^{(W_1)} c_s^{(W_2)}\right]^{\frac{1}{2}} \rho_{rs}^{(W_1,W_2)}$$

where δ_{rs} is the Kronecker's delta, indicating that the matrix of correlations is diagonal.

■

Remark: In general, the existence of the canonical expansion of a bivariate distribution may be proved by using the criterion of Brown (1958) which requires that the conditional moments $E[X^n|y]$ and $E[Y^n|x]$ must be polynomials with degree less than or equal to n. This criterion may not be easy to apply.

3.4 A General Form of Extended Bivariate Negative Binomial

In this section, the *EBNB* distribution will be described in further detail. In the mixed Poisson formulation, it is found that $-1 < \theta_3 < 0$ whereas by the extended trivariate reduction formulation, $0 < \theta_3 < 1$ for the *EBNB-I* distribution. Hence, a general *EBNB-I* distribution with $-1 < \theta_3 < 1$ can be now defined.

Definition 3.2 (*Extended Bivariate Negative Binomial*): The joint pgf of the *EBNB-I* is

$$G_{(X_1,X_2)}(z_1,z_2) = \left(\frac{q_1}{1-p_1z_1}\right)^{\alpha_1}\left(\frac{q_2}{1-p_2z_2}\right)^{\alpha_2}\left(\frac{\Theta}{1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2}\right)^{\nu} \tag{3.8}$$

with the parameters $p_1 = \frac{\theta_1+\theta_3}{1-\theta_2}$, $p_2 = \frac{\theta_2+\theta_3}{1-\theta_1}$, $q_i = 1-p_i$, $i = 1,2$ and $\Theta = 1-\theta_1-\theta_2-\theta_3$, and the restrictions $0 < p_1, p_2, \theta_1, \theta_2, \Theta < 1$, $-1 < \theta_3 < 1$, $\theta_3+\theta_1 > 0$, $\theta_3+\theta_2 > 0$, $\theta_3+\theta_1\theta_2 > 0$ and $\alpha_1, \alpha_2, \nu > 0$.

This is obtained by combining (3.3) and (3.7) (substitute $\alpha_1-\nu$ with α_1 and $\alpha_2-\nu$ with α_2). Furthermore, note that the marginal distributions $X_1 \sim NB(\alpha_1+\nu, p_1)$ and $X_2 \sim NB(\alpha_2+\nu, p_2)$ have different parameters. The correlation is in the range [0,1] as indicated by equation (3.10).

The following discussion focuses on the distributional properties for the *EBNB-I* distribution.

3.4.1 Joint Probability Mass Function

To obtain the *EBNB-I* distribution's pmf from the pgf, equation (3.8) is expanded in powers of z_1 and z_2. Then, the coefficient for the term $z_1^{x_1}z_2^{x_2}$ will give the pmf $\Pr(X_1 = x_1, X_2 = x_2) = f(x_1, x_2)$ as

$$f(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)\left(\frac{\Theta}{q_1q_2}\right)^{\nu}\sum_{r=0}^{x_1}\sum_{s=0}^{x_2}\left[\frac{(\alpha_1)_{x_1-r}}{(\alpha_1+\nu)_{x_1}}\binom{x_1}{r}\left(\frac{\theta_1}{p_1}\right)^{r}\frac{(\alpha_2)_{x_2-s}}{(\alpha_2+\nu)_{x_2}}\binom{x_2}{s}\left(\frac{\theta_2}{p_2}\right)^{s}\right.$$

$$\left.\cdot\sum_{i=0}^{\min(r,s)}(\nu)_{r+s-i}\binom{r}{i}\binom{s}{i}i!\left(\frac{\theta_3}{\theta_1\theta_2}\right)^{i}\right]$$

where $f_{X_i}(x_i)=(\alpha_i+\nu)_{x_i}p_i^{x_i}\,q_i^{\alpha_i+\nu}/x_i!$, $i=1,2$ are the pmf's of the marginal distributions X_1 and X_2 and the parameters are as defined in Section 3.1.1 e). Another method to obtain this pmf is described in Section 2.2.1.

On the other hand, for the case when ν is a non-negative integer, the pmf for *EBNB-II* is given below as

$\Pr(X_1=x_1,X_2=x_2)$

$$=f_{X_1}(x_1)f_{X_2}(x_2)\sum_{\substack{\alpha,\beta,\gamma\geq 0\\ \alpha+\beta+\gamma=\nu}}\frac{\nu!}{\alpha!\,\beta!\,\gamma!}\lambda_1^{\alpha}\lambda_2^{\beta}\lambda_3^{\gamma}[(1-d)(1-\theta_1)(1-\theta_2)]^{\nu}$$

$$\cdot\sum_{r=0}^{x_1}\sum_{s=0}^{x_2}\frac{(\alpha_1)_{x_1-r}}{(\alpha_1+\nu)_{x_1}}\binom{x_1}{r}\frac{(\alpha_2)_{x_2-s}}{(\alpha_2+\nu)_{x_2}}\binom{x_2}{s}\sum_{i=0}^{r}\sum_{j=0}^{s}\left[(\nu)_{r-i}(\nu)_i\binom{r}{i}\left(\theta_1/\left(\frac{\phi_1}{1+\phi_1}\right)\right)^{i}\right.$$

$$\left.\cdot\ (\nu)_{s-j}(\nu)_j\binom{s}{j}\left(\theta_1/\left(\frac{\phi_1}{1+\phi_1}\right)\right)^{j}\ {}_2F_1\ \big(\nu+i,\nu+j;\nu;d(1-\theta_1)(1-\theta_2)\big)\right]$$

where the parameters are as defined in Section 3.1.1 e), $\lambda_1=\frac{a}{1-d}$, $\lambda_2=\frac{b}{1-d}$, $\lambda_3=\frac{c}{1-d}$ and $f_{X_i}(x_i)=(\alpha_i+\nu)_{x_i}\left(\frac{\phi_i}{1+\phi_i}\right)^{x_i}\left(\frac{1}{1+\phi_i}\right)^{\alpha_i+\nu}/x_i!$, $i=1,2$ are the pmf's of the marginal distributions X_1 and X_2.

3.4.2 Factorial Moments and Correlation

By differentiating the pgf (3.8) repeatedly,

$$G^{(x_1,x_2)}(z_1,z_2)$$

$$= G_{(X_1,X_2)}(z_1,z_2)\sum_{r=0}^{x_1}\sum_{s=0}^{x_2}\left[\frac{(\alpha_1)_{x_1-r}}{(x_1-r)!}\left(\frac{p_1}{1-p_1z_1}\right)^{x_1-r}\frac{(\alpha_2)_{x_2-s}}{(x_2-s)!}\left(\frac{p_2}{1-p_2z_2}\right)^{x_2-s}\right.$$

$$\cdot\sum_{i=0}^{\min(r,s)}\frac{(\nu)_{r+s-i}}{i!\,(r-i)!\,(s-i)!}\left(\frac{\theta_1+\theta_3z_2}{1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2}\right)^{r-i}\left(\frac{\theta_2+\theta_3z_1}{1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2}\right)^{s-i}$$

$$\left.\cdot\left(\frac{\theta_3}{1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2}\right)^{i}\right]x_1!\,x_2!$$

and with equation (2.5), this gives the factorial moment of order (x_1,x_2) as

$$\mu_{(X_1,X_2)}^{[x_1,x_2]} = \mu_{X_1}^{[x_1]}\mu_{X_2}^{[x_2]}\sum_{r=0}^{x_1}\sum_{s=0}^{x_2}\left[\frac{(\alpha_1)_{x_1-r}}{(\alpha_1+\nu)_{x_1}}\binom{x_1}{r}\frac{(\alpha_2)_{x_2-s}}{(\alpha_2+\nu)_{x_2}}\binom{x_2}{s}\right.$$

$$\left.\cdot\sum_{i=0}^{\min(r,s)}(\nu)_{r+s-i}\binom{r}{i}\binom{s}{i}i!\left(\frac{\Theta\theta_3}{(\theta_1+\theta_3)(\theta_2+\theta_3)}\right)^{i}\right] \tag{3.9}$$

where $\mu_{X_i}^{[x_i]} = (\alpha_i+\nu)_{x_i}(p_i/q_i)^{x_i}$, $i = 1,2$, are the factorial moments of the marginal distributions X_1 and X_2, of order x_1 and x_2, respectively.

From (2.6), (3.9) and $\mu_{X_i}^{[x_i]}$, $i = 1,2$, the correlation coefficient is found to be

$$\rho_{(X_1,X_2)} = \frac{\nu(\theta_3+\theta_1\theta_2)}{\sqrt{(\alpha_1+\nu)(\alpha_2+\nu)(1-\theta_1)(1-\theta_2)(\theta_1+\theta_3)(\theta_2+\theta_3)}}$$

$$= \frac{\nu}{\sqrt{(\alpha_1+\nu)(\alpha_2+\nu)}}\rho_{(W_1,W_2)} \tag{3.10}$$

where $\rho_{(W_1,W_2)}$ is the correlation coefficient of (W_1,W_2).

3.4.3 Conditional Distributions and Regressions

From Theorem 2.1, the pgf of the conditional distribution of X_1 given $X_2 = x_2$ is

$$G_{X_1}(z|x_2) = \left(\frac{1-p_1}{1-p_1 z}\right)^{\alpha_1} \sum_{s=0}^{x_2} \frac{\Pr(Y_2 = x_2 - s)\Pr(W_2 = s)}{\Pr(X_2 = x_2)} \cdot \left(\frac{\theta_2 + \theta_3 z}{\theta_2 + \theta_3}\right)^{s} \left(1 - \frac{\theta_1}{1-\theta_1}(z-1)\right)^{-(\nu+s)} \tag{3.11}$$

From (3.11), the conditional distribution of X_1 given $X_2 = x_2$ is observed to be the convolution of V_1 and V_2 where $V_1 \sim NB(\alpha_1, p_1)$ and V_2 is a finite mixture of convolutions $(U_{1s} + U_{2s})$, $s = 0,1,\ldots,x_2$ with $U_{1s} \sim Binomial\left(s, \frac{\theta_3}{\theta_2+\theta_3}\right)$ and $U_{2s} \sim NB(\nu + s, \theta_1)$ when $0 < \theta_3 < 1$. When $-1 < \theta_3 < 0$, V_2 is a mixture of convolutions of pseudo-binomial and negative binomial random variables as described in Kemp (1979).

Hence, the regression of X_1 on X_2 according to Corollary 2.1 is

$$E[X_1|X_2 = x_2] = E[V_1] + E[V_2]$$

$$= \frac{\alpha_1 p_1}{q_1} + \sum_{s=0}^{x_2} \frac{\Pr(Y_2 = x_2 - s)\Pr(W_2 = s)}{\Pr(X_2 = x_2)} \left(E[U_{1s}] + E[U_{2s}]\right)$$

$$= \frac{\alpha_1 p_1}{q_1} + \frac{\nu\theta_1}{1-\theta_1} + \left(\frac{\theta_3 + \theta_1\theta_2}{(1-\theta_1)(\theta_2+\theta_3)}\right) \sum_{s=0}^{x_2} \frac{\Pr(Y_2 = x_2 - s)\Pr(W_2 = s)}{\Pr(X_2 = x_2)}.$$

Note that V_1 is equivalent to $Y_1 \sim NB(\alpha_1, p_1)$ and V_2 gives the conditional distribution of W_1 given $W_2 = s$.

Similarly, the pgf of the conditional distribution of X_2 given $X_1 = x_1$ and the regression of X_2 on X_1 are obtained as

$$G_{X_2}(z|x_1) = \left(\frac{1-p_2}{1-p_2z}\right)^{\alpha_2} \sum_{s=0}^{x_1} \frac{\Pr(Y_1 = x_1 - s)\Pr(W_1 = s)}{\Pr(X_1 = x_1)}$$

$$\cdot \left(\frac{\theta_1 + \theta_3 z}{\theta_1 + \theta_3}\right)^s \left(1 - \frac{\theta_2}{1-\theta_2}(z-1)\right)^{-(\nu+s)}$$

and

$$E[X_2|X_1 = x_1] = \frac{\alpha_2 p_2}{q_2} + \frac{\nu\theta_2}{1-\theta_2} + \left(\frac{\theta_3 + \theta_1\theta_2}{(1-\theta_2)(\theta_1+\theta_3)}\right) \sum_{s=0}^{x_1} \frac{\Pr(Y_1 = x_1 - s)\Pr(W_1 = s)}{\Pr(X_1 = x_1)}.$$

Furthermore,

$$E\left[X_1^k \middle| X_2 = x_2\right] = \sum_{i=0}^{k}\sum_{s=0}^{x_2} \binom{k}{i} \frac{\Pr(Y_2 = x_2 - s)\Pr(W_2 = s)}{\Pr(X_2 = x_2)} E\left[Y_1^{k-i}\right] E\left[W_1^i \middle| W_2 = s\right].$$

3.4.4 Canonical Expansion

Result 3.1 shows that the *EBNB-I* distribution has a canonical expansion in terms of orthogonal polynomials, which is derived next.

The factorial moment generating function (fmgf) for $f(x)\, m_r(x;\nu,p)/(\nu)_r$ is

$$\sum_{x=0}^{\infty} (1+t)^x f(x)\, m_r(x;\nu,p)/(\nu)_r = (-p)^{-r} \left(\frac{tp}{q}\right)^r \left(1 - \frac{tp}{q}\right)^{-(\nu+r)} \tag{3.12}$$

where $m_r(x;\nu,p) = (\nu)_r \cdot {}_2F_1\ (-r, -x; \nu; 1 - 1/p)$ is the rth Meixner polynomial and $f(x) = (\nu)_x p^x\, q^\nu / x!$ is the *NB* pmf .

The fmgf of *EBNB-I* distribution from (3.3) is

$$H(t_1, t_2) = (1 - A_1t_1)^{-(\alpha_1+\nu)}(1 - A_2t_2)^{-(\alpha_2+\nu)} \left(1 - \frac{(A_3 + A_1A_2)t_1t_2}{(1 - A_1t_1)(1 - A_2t_2)}\right)^{-\nu}$$

where $A_1 = p_1/q_1$, $A_2 = p_2/q_2$ and $A_3 = \theta_3/\Theta$. Following the same technique in Kocherlakota and Kocherlakota (1992, p. 135), $H(t_1, t_2)$ is expanded to give

$$H(t_1,t_2) = \sum_{i=0}^{\infty} \frac{(\nu)_i}{i!} \left(\frac{\rho_{(W_1,W_2)}\sqrt{(1-\theta_1)(1-\theta_2)}}{\Theta\sqrt{A_1A_2}} \right)^i$$

$$\cdot (A_1t_1)^i(1-A_1t_1)^{-(\alpha_1+\nu+i)}(A_2t_2)^i(1-A_2t_2)^{-(\alpha_2+\nu+i)}. \tag{3.13}$$

Using the relation in (3.12), the canonical expansion for the pmf of *EBNB-I* distribution is found to be

$$f(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \sum_{i=0}^{\infty} \left(\frac{\rho_{(W_1,W_2)}\sqrt{(1-\theta_1)(1-\theta_2)}}{\Theta\sqrt{A_1A_2}} \right)^i \frac{(\nu)_i(p_1p_2)^i}{i!}$$

$$\cdot \frac{m_i(x_1;\alpha_1+\nu,p_1)}{(\alpha_1+\nu)_i} \frac{m_i(x_2;\alpha_2+\nu,p_2)}{(\alpha_2+\nu)_i}.$$

Then,

$$f(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \sum_{i=0}^{\infty} \frac{(\nu)_i\rho^i_{(W_1,W_2)}}{\sqrt{(\alpha_1+\nu)_i(\alpha_2+\nu)_i}} m_i^*(x_1;\alpha_1+\nu,p_1)m_i^*(x_2;\alpha_2+\nu,p_2) \tag{3.14}$$

where $f_{X_j}(x_j) = (\alpha_j+\nu)_{x_j} p_j^{x_j} q_j^{\alpha_j+\nu}/x_j!$, $j = 1,2$ are the marginal pmf's of X_1 and X_2 and $m_i^*(x_j;\alpha_j+\nu,p_j) = m_i(x_j;\alpha_j+\nu,p_j)/\sqrt{(\alpha_j+\nu)_i i!\, p_j^{-i}}$, $j = 1,2$ is the ith orthonormal Meixner polynomial.

3.4.5 Quadrant Dependence

It is easy to show that the *EBNB-I* distribution is positively quadrant dependent from the canonical expansion of its joint pmf given by (3.14) when the marginals are identical. Next, it is shown that *EBNB-I* and *EBNB-II* distributions are quadrant dependent when the marginal parameters are *different*.

Result 3.2: The *EBNB-II* (based on Mitchell and Paulson's *BNB*) distribution with joint pgf (3.4) is positive quadrant dependent when $0 \leq \rho_{(W_1,W_2)} \leq 1$ and negative quadrant dependent when $-1 \leq \rho_{(W_1,W_2)} \leq 0$.

Proof:

Note that $G_{(X_1,X_2)}(z_1,z_2) \geq (\leq) G_{X_1}(z_1)G_{X_2}(z_2)$ implies $\Pr(X_1 \leq x_1, X_2 \leq x_2) \geq (\leq)\Pr(X_1 \leq x_1)\Pr(X_2 \leq x_2), \forall x_1, x_2$, that is, positive (negative) quadrant dependence (2.9). This follows by extracting the (x_1,x_2)-th term from the pgf.

Rewriting equation (3.4),

$$G_{(X_1,X_2)}(z_1,z_2)$$

$$= \left(1+\phi_1(1-z_1)\right)^{-(\alpha_1+\nu)}\left(1+\phi_2(1-z_2)\right)^{-(\alpha_2+\nu)} \cdot \left(\frac{1+\tau_1(1-z_1)+\tau_2(1-z_2)+\tau_1\tau_2(1-z_1)(1-z_2)-d}{1+\tau_1(1-z_1)+\tau_2(1-z_2)+a\phi_1\phi_2(1-z_1)(1-z_2)-d}\right)^{-\nu}$$

$$= \left(1+\phi_1(1-z_1)\right)^{-(\alpha_1+\nu)}\left(1+\phi_2(1-z_2)\right)^{-(\alpha_2+\nu)} \cdot \left(1-\frac{(ad-bc)\phi_1\phi_2(1-z_1)(1-z_2)}{\Phi}\right)^{-\nu}$$

$$= \left(1+\phi_1(1-z_1)\right)^{-(\alpha_1+\nu)}\left(1+\phi_2(1-z_2)\right)^{-(\alpha_2+\nu)} \cdot \left(1-\frac{\rho_{(W_1,W_2)}(1-d)\sqrt{\phi_1(1+\phi_1)\phi_2(1+\phi_2)}(1-z_1)(1-z_2)}{\Phi}\right)^{-\nu} \quad (3.15)$$

$$= \left(1+\phi_1(1-z_1)\right)^{-(\alpha_1+\nu)}\left(1+\phi_2(1-z_2)\right)^{-(\alpha_2+\nu)} \cdot \left\{1+\sum_{i=1}^{\infty}\frac{(\nu)_i}{i!}\left(\frac{\rho_{(W_1,W_2)}(1-d)\sqrt{\phi_1(1+\phi_1)\phi_2(1+\phi_2)}(1-z_1)(1-z_2)}{\Phi}\right)^{i}\right\} \quad (3.16)$$

where $\Phi = 1-d+\tau_1(1-z_1)+\tau_2(1-z_2)+a\phi_1\phi_2(1-z_1)(1-z_2)$.

Since $|z_i| \leq 1$, $i = 1{,}2$, the terms $1 - z_i$ and $(1 - z_1)(1 - z_2)$ as well as $\tau_i(1 - z_i)$ are nonnegative. Also, $1 - d > 0$. When $0 \leq \rho_{(W_1,W_2)} \leq 1$, the infinite series in braces in (3.16) is positive. Hence, (3.16) implies that $G_{(X_1,X_2)}(z_1, z_2) \geq G_{X_1}(z_1)G_{X_2}(z_2)$. When $-1 \leq \rho_{(W_1,W_2)} \leq 0$, rewrite (3.15) as

$$G_{(X_1,X_2)}(z_1,z_2) = \big(1 + \phi_1(1 - z_1)\big)^{-(\alpha_1+\nu)}\big(1 + \phi_2(1 - z_2)\big)^{-(\alpha_2+\nu)} \cdot \left\{\left(1 + \frac{\big(-\rho_{(W_1,W_2)}\big)(1 - d)\sqrt{\phi_1(1 + \phi_1)\phi_2(1 + \phi_2)}(1 - z_1)(1 - z_2)}{\Phi}\right)^{-\nu}\right\}$$

showing that the last term in brackets is greater than 1. Hence, the expression in braces is less than 1 implying that $G_{(X_1,X_2)}(z_1, z_2) \leq G_{X_1}(z_1)G_{X_2}(z_2)$. From the remark at the beginning of the proof, it is concluded that the *EBNB-II* is positive quadrant dependent when $0 \leq \rho_{(W_1,W_2)} \leq 1$ and negative quadrant dependent when $-1 \leq \rho_{(W_1,W_2)} \leq 0$. ■

Corollary 3.1: The *EBNB-I* distribution with joint pgf (3.3) is positive quadrant dependent when $0 \leq \rho_{(W_1,W_2)} \leq 1$.

Proof:

From (3.16) with $b = c = 0$ or by rewriting equation (3.3) using the relation of (2.4) in (3.13), it is obtained

$$G_{(X_1,X_2)}(z_1,z_2) = \big(1 + A_1(1 - z_1)\big)^{-(\alpha_1+\nu)}\big(1 + A_2(1 - z_2)\big)^{-(\alpha_2+\nu)} \cdot \left\{1 + \sum_{i=1}^{\infty} \frac{(\nu)_i}{i!} \left(\frac{\rho_{(W_1,W_2)}\sqrt{(1 - \theta_1)(1 - \theta_2)}}{\Theta\sqrt{A_1A_2}}\right)^i \left(\frac{A_1A_2(1 - z_1)(1 - z_2)}{\big(1 + A_1(1 - z_1)\big)\big(1 + A_2(1 - z_2)\big)}\right)^i\right\} \tag{3.17}$$

where $A_1 = p_1/q_1$ and $A_2 = p_2/q_2$.

Similar to the arguments in Result 3.2, when $0 \leq \rho_{(W_1,W_2)} \leq 1$, the infinite series in braces in (3.17) is positive. Hence, (3.17) implies that $G_{(X_1,X_2)}(z_1,z_2) \geq G_{X_1}(z_1)G_{X_2}(z_2)$. Thus, *EBNB-I* is positive quadrant dependent when $0 \leq \rho_{(W_1,W_2)} \leq 1$.

■

Remarks: (1) Quadrant dependence of the *BNB* distributions of Mitchell & Paulson (1981), Edwards & Gurland (1961) and Subrahmaniam (1966) follow on setting $\alpha_1 = \alpha_2 = 0$.

(2) For the extended trivariate reduction (3.1), it is easy to show that if (W_1, W_2) is positive (negative) quadrant dependent, then (X_1, X_2) is also positive (negative) quadrant dependent.

3.4.6 Partial Derivates and Information Matrix

Taking natural logarithm of equation (3.8) as

$$\ln\big(G(z_1,z_2)\big) = \alpha_1[\ln(1-p_1) - \ln(1-p_1z_1)] + \alpha_2[\ln(1-p_2) - \ln(1-p_2z_2)] \\ + \nu[\ln\Theta - \ln(1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2)]$$

with $\theta_1 = \big(p_1(1-p_2) - \theta_3(1-p_1)\big)/(1-p_1p_2)$,

$\theta_2 = \big(p_2(1-p_1) - \theta_3(1-p_2)\big)/(1-p_1p_2)$ and

$\Theta = (1+\theta_3)(1-p_1)(1-p_2)/(1-p_1p_2)$,

the first order differentiations of the pgf with respect to its parameters are obtained.

First, let $G_{(U_1,U_2)}(z_1,z_2) = \Theta/(1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2)$ be a Edwards and Gurland's *BNB* distribution with the index parameter, $\nu = 1$ and the corresponding pmf, $U(u_1,u_2)$. Also, let $\ln(1-\theta_1z_1-\theta_2z_2-\theta_3z_1z_2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} V(i,j)z_1^iz_2^j$ where

$iV(i,j) = -\left(\theta_1 U(i-1,j) + \theta_3 U(i-1,j-1)\right)/(1-\theta_1-\theta_2-\theta_3)$. Then, the first order differentiations are as follows.

$$\frac{dG}{dp_1} = \sum_{x=1}^{\infty}\sum_{y=0}^{\infty}\sum_{i=0}^{x-1} \alpha_1 p_1^i \Pr(x-i-1,y)\, z_1^x z_2^y - \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \frac{\alpha_1}{1-p_1} \Pr(x,y) z_1^x z_2^y$$

$$+ \frac{\nu}{(1-p_1)(1-p_1p_2)} \left[\sum_{x=1}^{\infty}\sum_{y=0}^{\infty}\sum_{i=0}^{x-1}\sum_{j=0}^{y} U(x-i-1,y \right.$$

$$- j)\Pr(i,j) z_1^x z_2^y - \sum_{x=0}^{\infty}\sum_{y=1}^{\infty}\sum_{i=0}^{x}\sum_{j=0}^{y-1} p_2 U(x-i,y-j-1)\Pr(i,j) z_1^x z_2^y$$

$$\left. - \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} (1-p_2)\Pr(x,y) z_1^x z_2^y \right]$$

$$\frac{dG}{dp_2} = \sum_{x=0}^{\infty}\sum_{y=1}^{\infty}\sum_{i=0}^{y-1} \alpha_2 p_2^i \Pr(x,y-i-1)\, z_1^x z_2^y - \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \frac{\alpha_2}{1-p_2} \Pr(x,y) z_1^x z_2^y$$

$$+ \frac{\nu}{(1-p_2)(1-p_1p_2)} \left[-\sum_{x=1}^{\infty}\sum_{y=0}^{\infty}\sum_{i=0}^{x-1}\sum_{j=0}^{y} p_1 U(x-i-1,y \right.$$

$$- j)\Pr(i,j) z_1^x z_2^y + \sum_{x=0}^{\infty}\sum_{y=1}^{\infty}\sum_{i=0}^{x}\sum_{j=0}^{y-1} U(x-i,y-j-1)\Pr(i,j) z_1^x z_2^y$$

$$\left. - \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} (1-p_1)\Pr(x,y) z_1^x z_2^y \right]$$

$$\frac{dG}{d\theta_3} = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \frac{\nu}{(1+\theta_3)} \Pr(x,y) z_1^x z_2^y$$

$$+ \sum_{x=1}^{\infty}\sum_{y=1}^{\infty}\sum_{i=0}^{x-1}\sum_{j=0}^{y-1} \frac{\nu}{\Theta} U(x-i-1,y-j-1)\Pr(i,j) z_1^x z_2^y$$

$$\frac{dG}{d\alpha_1} = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \ln(1-p_1)\Pr(x,y) z_1^x z_2^y - \sum_{x=1}^{\infty}\sum_{y=0}^{\infty}\sum_{i=0}^{x-1} \frac{p_1^{i+1}}{(i+1)} \Pr(x-i-1,y) z_1^x z_2^y$$

$$\frac{dG}{d\alpha_2} = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \ln(1-p_2)\Pr(x,y)z_1^x z_2^y - \sum_{x=0}^{\infty}\sum_{y=1}^{\infty}\sum_{i=0}^{y-1} \frac{p_2^{i+1}}{(i+1)}\Pr(x,y-i-1)z_1^x z_2^y$$

$$\frac{dG}{d\nu} = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \ln(\Theta)\Pr(x,y)z_1^x z_2^y - \sum_{x=0}^{\infty}\sum_{y=0}^{\infty}\sum_{i=0}^{x}\sum_{j=0}^{y} V(x-i,y-j)\Pr(i,j)z_1^x z_2^y$$

From the fact that $\frac{dG}{d\vartheta} = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty}\frac{d\Pr(x,y)}{d\vartheta} z_1^x z_2^y$, the first order differentiations for the corresponding pmf of the *EBNB-I* distribution are

$$\frac{d\Pr(x,y)}{dp_1} = -\left(\frac{\alpha_1}{1-p_1} + \frac{\nu(1-p_2)}{(1-p_1)(1-p_1p_2)}\right)\Pr(x,y) + \sum_{i=0}^{x-1}\alpha_1 p_1^i \Pr(x-i-1,y)$$
$$+\frac{\nu}{(1-p_1)(1-p_1p_2)}\left[\sum_{i=0}^{x-1}\sum_{j=0}^{y} U(x-i-1,y-j)\Pr(i,j) - \sum_{i=0}^{x}\sum_{j=0}^{y-1} p_2 U(x-i,y-j-1)\Pr(i,j)\right],$$

$$\frac{d\Pr(x,y)}{dp_2} = -\left(\frac{\alpha_2}{1-p_2} + \frac{\nu(1-p_1)}{(1-p_2)(1-p_1p_2)}\right)\Pr(x,y) + \sum_{i=0}^{y-1}\alpha_2 p_2^i \Pr(x,y-i-1)$$
$$+\frac{\nu}{(1-p_2)(1-p_1p_2)}\left[-\sum_{i=0}^{x-1}\sum_{j=0}^{y} p_1 U(x-i-1,y-j)\Pr(i,j) + \sum_{i=0}^{x}\sum_{j=0}^{y-1} U(x-i,y-j-1)\Pr(i,j)\right],$$

$$\frac{d\Pr(x,y)}{d\theta_3} = \frac{\nu}{(1+\theta_3)}\Pr(x,y) + \sum_{i=0}^{x-1}\sum_{j=0}^{y-1}\frac{\nu}{\Theta}U(x-i-1,y-j-1)\Pr(i,j)\ ,$$

$$\frac{d\Pr(x,y)}{d\alpha_1} = \ln(1-p_1)\Pr(x,y) - \sum_{i=0}^{x-1}\frac{p_1^{i+1}}{(i+1)}\Pr(x-i-1,y)\ ,$$

$$\frac{d\Pr(x,y)}{d\alpha_2} = \ln(1-p_2)\Pr(x,y) - \sum_{i=0}^{y-1}\frac{p_2^{i+1}}{(i+1)}\Pr(x,y-i-1)\ \text{, and}$$

$$\frac{d\Pr(x,y)}{d\nu} = \ln(\Theta)\Pr(x,y) - \sum_{i=0}^{x}\sum_{j=0}^{y} V(x-i, y-j)\Pr(i,j)\,.$$

The information matrix is then obtained as explained in Section 2.2.4.

3.4.7 Computer Generation of Bivariate Samples

Here, the algorithms to generate random samples from *EBNB-I* distribution are given. These algorithms are also applicable to generate random samples for extended bivariate binomial and gamma distributions.

a) Mixture Method

By the formulation $X_1 = Y_1 + W_1$ and $X_2 = Y_2 + W_2$, the general form of pgf for *EBNB* distributions is given as in (3.2). It is found in Section 3.4.2 that the correlation for *EBNB-I* distribution has the form

$$\rho_{(X_1,X_2)} = \frac{\nu}{\sqrt{(\alpha_1+\nu)(\alpha_2+\nu)}}\rho_{(W_1,W_2)}$$

where α_1, α_2 and ν are the corresponding index parameters for Y_1, Y_2 and (W_1, W_2) distributions and $\rho_{(W_1,W_2)}$ is the correlation of the bivariate distribution of (W_1, W_2). Extended bivariate binomial and gamma distributions can also be shown to have the same form of correlation relation.

For any of the bivariate negative binomial, binomial and gamma distributions, given the marginals $X_1 \sim g(\alpha, \theta_1)$ and $X_2 \sim g(\beta, \theta_2)$ as well as the correlation $\rho_{(X_1,X_2)}$, it can be deduced that $Y_1 \sim g(\alpha - \nu, \theta_1)$, $Y_2 \sim g(\beta - \nu, \theta_2)$, $W_1 \sim g(\nu, \theta_1)$, $W_2 \sim g(\nu, \theta_2)$ and $\rho_{(W_1,W_2)} = \frac{\sqrt{(\alpha_1+\nu)(\alpha_2+\nu)}}{\nu}\rho_{(X_1,X_2)}$ with $g(\cdot)$ being one of the corresponding univariate distribution. Ong (1990, 1992) has given several mixture models as well as algorithms for computer generation for these bivariate distributions of (W_1, W_2) with given marginals and correlation. From this and utilising the formulation (3.1), an algorithm to generate bivariate data from one of the three bivariate distributions with different marginals is given below.

Algorithm 3.1: Outline of Computer Generation of Bivariate Samples using Mixture Method

---

(1) Set $0 < \nu < \min(\alpha, \beta)$, $\alpha_1 = \alpha - \nu$, $\alpha_2 = \beta - \nu$ and

$\rho_{(W_1,W_2)} = \frac{\sqrt{(\alpha_1+\nu)(\alpha_2+\nu)}}{\nu} \rho_{(X_1,X_2)}$.

(2) Generate $y_1 \sim g(\alpha_1, \theta_1)$ and $y_2 \sim g(\alpha_2, \theta_2)$.

(3) Use known marginals $W_1 \sim g(\nu, \theta_1)$, $W_2 \sim g(\nu, \theta_2)$ and $\rho_{(W_1,W_2)}$ in algorithm *BNB* from Ong (1992) to generate (w_1, w_2).

(4) $x_1 = y_1 + w_1$ and $x_2 = y_2 + w_2$.

---

b) Conditional Distribution Technique

To simplify explanation, the following will focus on the implementation of conditional distribution technique to generate *EBNB-I* data.

The conditional distribution of X_1 given $X_2 = x_2$ is the convolution of V_1 and V_2 as given in Section 3.4.3. Now, given the marginals $X_1 \sim NB(\alpha, p_1)$, $X_2 \sim NB(\beta, p_2)$ and $\rho_{(X_1,X_2)}$. When $0 < \theta_3 < 1$, it is found that $V_1 \sim NB(\alpha - \nu, p_1)$, $U_{1s} \sim Binomial\left(s, \frac{\theta_3}{\theta_2+\theta_3}\right)$ and $U_{2s} \sim NB(\nu + s, \theta_1)$, $s = 0,1,\dots,x_2$, $Y_2 \sim NB(\beta - \nu, p_2)$ and $W_2 \sim NB(\nu, p_2)$. When $-1 < \theta_3 < 0$, V_2 is a mixture of convolutions between a pseudo-binomial and a negative binomial rv's which can be easily generated using the standard inverse transform method.

Algorithm 3.2: Outline of Computer Generation of Bivariate Samples using Conditional Distribution Technique

---

(1) Set $0 < \nu < \min(\alpha, \beta)$, $\alpha_1 = \alpha - \nu$, $\alpha_2 = \beta - \nu$.

(2) Set θ_1, θ_2 and θ_3 such that

$$\rho_{(X_1,X_2)} = \frac{\nu(\theta_3 + \theta_1\theta_2)}{\sqrt{(\alpha_1 + \nu)(\alpha_2 + \nu)(1 - \theta_1)(1 - \theta_2)(\theta_1 + \theta_3)(\theta_2 + \theta_3)}}.$$

(3) Generate $x_2 \sim NB(\beta, p_2)$ and $v_1 \sim NB(\alpha_1, p_1)$.

(4) Set $v_2 = 0$. For $s = 0$ to x_2,

a) When $0 < \theta_3 < 1$,

(i) Generate $u_{1s} \sim Binomial\left(s, \frac{\theta_3}{\theta_2+\theta_3}\right)$ and $u_{2s} \sim NB(\nu + s, \theta_1)$.

(ii) $v_2 = v_2 + \left(\frac{\Pr(Y_2 = x_2 - s)\Pr(W_2 = s)}{\Pr(X_2 = x_2)}\right)(u_{1s} + u_{2s})$.

b) When $-1 < \theta_3 < 0$, generate v_2 using the inverse transform method based on the probabilities given in Kemp (1979).

(5) $x_1 = v_1 + v_2$.

3.5 Applications and Numerical Illustrations

In this section, the application of the two formulations of the *EBNB* distribution for accident data is illustrated, bearing in mind that it is also applicable in other contexts.

3.5.1 Extended Trivariate Reduction

Suppose that accidents or injuries are due to (a) individual characteristics and (b) environmental factors (Arbous & Kerrich, 1951). Let Y_1 and Y_2 represent the number of accidents due to (a) at two different time periods. Suppose that accidents due to cause (b) vary from one time period to another as a pair of correlated random variables (W_1, W_2). The total number of accidents in each period will be given by

$$X_1 = Y_1 + W_1 \text{ and } X_2 = Y_2 + W_2.$$

If Y_1 and Y_2 are assumed to be Poisson-distributed but due to individual characteristics, accident proneness varies from individual to individual as a gamma distribution, then Y_1 and Y_2 will have the *NB* distributions. A similar reasoning for accidents due to environmental factors lead to the assumption that (W_1, W_2) has the *BNB* of Edward & Gurland (1961). Hence accidents in the two time periods will have the *EBNB-I* distribution.

3.5.2 Mixed Poisson Formulation

Let X_1 and X_2 represent the number of accidents or injuries sustained by a group of individuals in two different time periods, each of unit length, with Poisson distributions $Poisson(\lambda_1)$ and $Poisson(\lambda_2)$ respectively. Suppose that the population consists of individuals where the proneness of each individual to accidents varies from individual to individual (see Edwards & Gurland, 1961; Subrahmaniam, 1966), that is, $\lambda_i,\ i = 1,2$ differs from individual to individual. If λ_1 and λ_2 have a joint bivariate gamma distribution given by (3.6), we get the *EBNB-I* distribution.

3.5.3 Examples

Two examples of fits of the *EBNB-I* distribution to a simulated data set and the rain-forest data set (see Holgate, 1966) are considered in this section. The parameters have been estimated by maximum likelihood estimation (*MLE*) and the fits are compared with Edwards and Gurland's *BNB* distribution. The log likelihood function is maximized using the numerical method of simulated annealing to obtain globally optimum parameter estimates. Suitable bounds are set for the unbounded parameters α_1, α_2 and ν to assist in the numerical parameter searches. Bounds for the parameters p_1, p_2 and θ_3 are as given in Section 3.4.

Example 1. A sample of size 500 is simulated from the *EBNB-I* distribution with $p_1 = 0.4$, $p_2 = 0.5$, $\theta_3 = 0.3$, $\alpha_1 = 0.5$, $\alpha_2 = 2.5$ and $\nu = 1.0$, where the marginals $X_1 \sim NB(0.4, 1.5)$ and $X_2 \sim NB(0.5, 3.5)$ clearly have different index parameters. Simulation is done according to the Algorithm 3.1. Observed frequencies for the data are shown in the following Table 3.1.

The *EBNB-I* and Edwards and Gurland's *BNB* distributions are fitted to the data with grouping of frequencies at the cell (16,8). The comparison of the fittings is made based on the chi-square, χ^2 goodness-of-fit statistic (2.13). The parameter estimates and corresponding χ^2 values as well as degrees of freedom (d.f.) are given in Table 3.2. The expected frequencies for these two distributions are then given in Table 3.3.

It is obvious from the χ^2 values in Table 3.2 that *BNB* could not give a satisfactory fit (*p*-value = 0.09) when the index parameters for the marginals are different as compared to *EBNB-I* (*p*-value = 0.56).

Table 3.1

Simulated Sample of Size 500 from *EBNB-I* Distribution with

$p_1 = 0.4$, $p_2 = 0.5$, $\theta_3 = 0.3$, $\alpha_1 = 0.5$, $\alpha_2 = 2.5$ and $\nu = 1.0$

x_2	x_1								
	0	1	2	3	4	5	6	7	8+
0	35	10	3	2	0	0	1	0	0
1	53	16	1	1	1	0	0	1	0
2	56	27	12	2	0	0	0	0	0
3	33	17	9	8	1	0	0	0	0
4	26	25	12	3	3	0	0	0	0
5	20	12	9	4	1	0	0	0	0
6	17	13	5	4	2	1	0	0	0
7	7	3	1	4	4	0	0	0	0
8	3	5	2	2	1	0	0	0	0
9	2	3	2	4	1	0	0	0	0
10	1	1	1	1	1	0	1	0	0
11	0	0	0	0	0	0	0	0	0
12	0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	1	0	0	0
15	0	0	0	1	0	0	0	0	0
16+	0	0	0	0	0	0	0	0	1

Note: The dotted lines indicate grouping of the data for the χ^2 goodness-of-fit test to yield a minimum expected frequency of 1.

Table 3.2

Parameter Estimates and χ^2 Values for *EBNB-I* and *BNB* Distributions

Distribution	*ML* Estimates	χ^2
EBNB-I	$\hat{p}_1 = 0.415158$, $\hat{p}_2 = 0.506243$, $\hat{\theta}_3 = 0.350500$, $\hat{\alpha}_1 = 0.443790$, $\hat{\alpha}_2 = 2.492638$, $\hat{\nu} = 0.825251$	51.84 d.f. = 54
BNB	$\hat{p}_1 = 0.245742$, $\hat{p}_2 = 0.546149$, $\hat{\theta}_3 = 3.16 \times 10^{-8}$, $\hat{\nu} = 2.823620$	70.81 d.f. = 56

Table 3.3

Expected Frequencies when *EBNB-I* (*BNB*) Distribution is Fitted to the Simulated Data

x_2	x_1								
	0	1	2	3	4	5	6	7	8+
0	37.90 (36.40)	6.98 (13.24)	2.09 (3.26)	0.71 (0.68)	0.25 (0.13)	0.09 (0.02)	0.04 (0.00)	0.01 (0.00)	0.01 (0.00)
1	52.70 (48.90)	20.67 (24.09)	4.93 (7.48)	1.59 (1.87)	0.56 (0.41)	0.20 (0.08)	0.08 (0.02)	0.03 (0.00)	0.02 (0.00)
2	49.13 (44.48)	26.00 (27.64)	9.34 (10.37)	2.50 (3.04)	0.84 (0.77)	0.30 (0.17)	0.11 (0.04)	0.04 (0.01)	0.03 (0.00)
3	38.47 (34.03)	23.94 (25.53)	11.20 (11.22)	3.91 (3.77)	1.11 (1.07)	0.38 (0.27)	0.14 (0.06)	0.05 (0.01)	0.03 (0.00)
4	27.32 (23.57)	18.78 (20.72)	10.36 (10.44)	4.58 (3.96)	1.57 (1.25)	0.47 (0.35)	0.16 (0.09)	0.06 (0.02)	0.04 (0.01)
5	18.23 (15.31)	13.40 (15.43)	8.24 (8.77)	4.31 (3.70)	1.83 (1.29)	0.62 (0.39)	0.19 (0.11)	0.07 (0.03)	0.04 (0.01)
6	11.66 (9.50)	8.99 (10.79)	5.95 (6.83)	3.50 (3.17)	1.76 (1.21)	0.73 (0.40)	0.24 (0.12)	0.08 (0.03)	0.04 (0.01)
7	7.22 (5.70)	5.77 (7.21)	4.03 (5.02)	2.57 (2.55)	1.46 (1.05)	0.71 (0.38)	0.29 (0.12)	0.10 (0.03)	0.05 (0.01)
8	4.37 (3.33)	3.58 (4.64)	2.61 (3.53)	1.77 (1.95)	1.10 (0.87)	0.61 (0.33)	0.29 (0.11)	0.11 (0.03)	0.06 (0.01)
9	2.59 (1.90)	2.17 (2.90)	1.63 (2.40)	1.16 (1.42)	0.77 (0.68)	0.47 (0.28)	0.25 (0.10)	0.12 (0.03)	0.07 (0.01)
10	1.52 (1.07)	1.29 (1.77)	1.00 (1.58)	0.73 (1.00)	0.51 (0.51)	0.33 (0.22)	0.20 (0.08)	0.10 (0.03)	0.07 (0.01)
11	0.87 (0.59)	0.76 (1.06)	0.60 (1.01)	0.45 (0.69)	0.33 (0.37)	0.22 (0.17)	0.14 (0.07)	0.08 (0.03)	0.07 (0.01)
12	0.50 (0.33)	0.44 (0.62)	0.35 (0.63)	0.27 (0.46)	0.20 (0.26)	0.15 (0.13)	0.10 (0.05)	0.06 (0.02)	0.06 (0.01)
13	0.28 (0.18)	0.25 (0.36)	0.20 (0.39)	0.16 (0.30)	0.12 (0.18)	0.09 (0.09)	0.06 (0.04)	0.04 (0.02)	0.05 (0.01)
14	0.16 (0.10)	0.14 (0.21)	0.12 (0.24)	0.09 (0.19)	0.07 (0.12)	0.06 (0.07)	0.04 (0.03)	0.03 (0.01)	0.04 (0.01)
15	0.09 (0.05)	0.08 (0.12)	0.07 (0.14)	0.05 (0.12)	0.04 (0.08)	0.03 (0.05)	0.03 (0.02)	0.02 (0.01)	0.03 (0.01)
16+	0.11 (0.06)	0.10 (0.14)	0.08 (0.20)	0.07 (0.19)	0.06 (0.14)	0.05 (0.09)	0.04 (0.05)	0.03 (0.02)	0.06 (0.02)

Example 2. Abundance of two different plant species in the rain-forest data (Holgate, 1966) can be due to individual growth factor and environmental factors such as climate and space. This set of data can be fitted by the *EBNB-I* and Edwards and Gurland's *BNB* distributions. The *MLE* is carried out and the parameter estimates are obtained as:

(i) *EBNB-I*

$\hat{p}_1 = 0.308341, \hat{p}_2 = 0.245690, \hat{\theta}_3 = 0.225283, \hat{\alpha}_1 = 1.463361,$

$\hat{\alpha}_2 = 1.300684, \hat{v} = 0.638345$

with marginals $X_1 \sim NB(0.3083, 2.1017)$ and $X_2 \sim NB(0.2457, 1.9390)$.

(ii) Edwards and Gurland's *BNB*

$\hat{p}_1 = 0.288067, \hat{p}_2 = 0.208444, \hat{\theta}_3 = 0.003166, \hat{v} = 2.336087$

with marginals $X_1 \sim NB(0.2881, 2.3361)$ and $X_2 \sim NB(0.2084, 2.3361)$.

The expected frequencies obtained from both distributions are shown in Table 3.4. Expected frequencies from the Type II bivariate non-central *NB* (*BNNB*) distribution fit from Ong & Lee (1986) are also given for comparison. Again, the comparison of the fittings is made based on the χ^2 statistic (2.13).

Note that for the rain-forest data the marginal distributions for *EBNB-I* and the *BNB* are similar. As expected, the fit by *EBNB-I* yields a smaller χ^2 value as compared to *BNB* since more flexibility is allowed for the marginals. This χ^2 value is also smaller than the χ^2 value obtained from the Type II *BNNB* distribution as shown in Table 3.4.

Table 3.4

Observed and Expected Frequencies for Rain-forest Data

x_1	x_2	Observed	Expected		
			EBNB-I	*BNB*	Type II *BNNB*
0	0	34	31.93	30.19	28.80
1	0	12	16.10	16.94	17.64
2	0	4	6.35	6.79	7.53
{ 3	0	5	2.30	2.36	2.66
4	0	2	0.80	0.75	0.84
0	1	8	10.20	10.94	11.14
1	1	13	9.74	8.99	9.52
2	1	3	4.73	4.75	5.05
{ 3	1	3	1.87	2.04	2.11
4	1	0	0.68	0.78	0.76
0	2	3	2.88	2.84	3.00
1	2	6	2.92	3.07	3.19
{ 2	2	1	2.28	2.01	2.00
3	2	2	1.14	1.04	0.96
4	2	0	0.46	0.46	0.39
{ 0	3	1	0.78	0.64	0.67
1	3	1	0.81	0.85	0.84
2	3	0	0.67	0.67	0.61
3	3	1	0.51	0.40	0.33
4	3	0	0.27	0.21	0.15
$x_1 \leq 4$	$x_2 \geq 4$	0	1.18	0.99	0.75
$x_1 \geq 5$		1	1.39	1.31	1.04
χ^2			13.16	14.28	13.44
d.f.			6	8	9

CHAPTER 4 : PARAMETER ESTIMATION BASED ON PROBABILITY GENERATING FUNCTION

4.0 Introduction

Generating functions has been considered for statistical inference by many researchers; for example, Press (1972) and recently, Meintanis & Swanepoel (2007) as well as the references therein. For count variables, in particular, the probability generating function (pgf) has been proposed for testing goodness-of-fit and parameter estimation. The motivation to use pgf is that it is usually much simpler than the corresponding probability mass function (pmf), and this leads to simpler inference procedures. This is especially true when dealing with multivariate discrete distributions. For example, Edwards & Gurland's (1961) bivariate negative binomial (*BNB*) distribution has pgf of the form

$$G_{(X_1,X_2)}(z_1, z_2) = \left(\frac{\Theta}{1 - \theta_1 z_1 - \theta_2 z_2 - \theta_3 z_1 z_2}\right)^{\nu}, \tag{4.1}$$

where $0 \leq \theta_i, \Theta \leq 1$, $i = 1,2,3$ and $\Theta = 1 - \theta_1 - \theta_2 - \theta_3$, as opposed to its joint pmf

$$f(x_1, x_2) = \Theta^{\nu} \sum_{i=0}^{\min(x_1,x_2)} (\nu)_{x_1+x_2-i} \binom{x_1}{i}\binom{x_2}{i} \frac{i!}{x_1!\,x_2!} \theta_1^{x_1-i} \theta_2^{x_2-i} \theta_3^{i} \;. \tag{4.2}$$

Evidently, the popular maximum likelihood estimation (*MLE*) utilising this pmf will be computationally involved despite its many appealing properties such as yielding efficient estimators. Furthermore, *MLE* performs badly in the presence of outliers in the data. Hence, a method which is simpler, faster and robust against outliers is much desired.

Kemp & Kemp (1988) introduced a rapid estimation method for univariate discrete distributions based on the pgf. Dowling & Nakamura (1997) further derived asymptotic properties for these estimators. To address the problem of outliers in parameter estimation, Simpson (1987) outlined the advantage of minimum Hellinger distance (*MHD*) estimation in reducing the effects of outliers on the parameter estimates as well as giving interesting properties, including asymptotic efficiency and the breakdown point of the method. *MHD* has since been generalized to other robust minimum divergence estimation methods such as minimum generalized Hellinger distance (*MGHD*) and the penalized version of *MGHD* (*MPGHD*) methods by Basu *et al*. (1997). These methods have been described briefly in Section 2.4.

By combining the idea of *MHD* estimation method with the pgf method, a new minimum distance parameter estimation method for bivariate and multivariate discrete distributions is introduced here to obtain a rapid estimation method, which is both consistent and robust against outliers. This method will be shown to be of great utility in the multivariate case. Since the focus is on the multivariate case, only an example will be given in Section 4.5.1 to show the usefulness of this new method in parameter estimation for univariate distributions with complicated pmf.

4.1 MD_α Estimation Method

Kemp & Kemp (1988) proposed a fast estimation method for discrete distributions by solving the simultaneous equations obtained from the relation of the empirical probability generating function (epgf), $\hat{G}_n(z)$ to its theoretical pgf, $G(z)$ given by

$$G(z) = \hat{G}_n(z) = \sum_{i=0}^{k} p_i z^i$$

using two predetermined z values, where $p_i = n_i/n$, n_i is the frequency for the data value i with sample size n and $\sum_{i=0}^{k} p_i = 1$. A drawback of this method is the need to select z values and then to determine the effects of different combinations of these z values. To avoid this, a pgf-based minimum distance (MD_α) estimation method which takes into account all z values in the range $[0, 1]$ is proposed as follows.

Let the L^2 norm be denoted by $\|h(z)\|_2 = (\int |h(z)|^2 dz)^{1/2}$. Also, let $G_{\boldsymbol{\theta}}(\mathbf{z})$, $\mathbf{z} = (z_1, z_2, \dots, z_k)$ denote the pgf for a k-variate parametric family $F_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \Omega$ and $\Omega \in \mathbb{R}^d$, where Ω is the parameter space. Let $\hat{G}_n(\mathbf{z})$ denote the epgf obtained from a random sample $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n)$ of k-vectors. The MD_α estimate of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, minimizes the distance measure $\left\|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right\|_2^2$ over the region $\mathbf{z} \in [0,1]^k$ for $\alpha > 0$. Mathematically,

$$\widehat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}\in\Omega}\left\|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right\|_2^2 = \min_{\boldsymbol{\theta}\in\Omega}\left(\int_0^1 \int_0^1 \dots \int_0^1 \left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right|^2 dz_1 dz_2 \dots dz_k\right). \tag{4.3}$$

When $\alpha = 1$, MD_1 estimate of $\boldsymbol{\theta}$ is $\widehat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}\in\Omega}\left\|\hat{G}_n(\mathbf{z}) - G_{\boldsymbol{\theta}}(\mathbf{z})\right\|_2^2$. Rueda & O'Reilly (1999) have used the measure $\left\|\hat{G}_n(z) - G_\theta(z)\right\|_2^2$ as a goodness-of-fit test of a univariate Poisson distribution but not for parameter estimation. When $\alpha = 1/2$, we have $MD_{1/2}$ estimate, $\widehat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}\in\Omega}\left\|\hat{G}_n^{1/2}(\mathbf{z}) - G_{\boldsymbol{\theta}}^{1/2}(\mathbf{z})\right\|_2^2$. This measure is investigated since it is similar to *MHD* measure of $\left\|f_n^{1/2}(z) - f_{\boldsymbol{\theta}}^{1/2}(z)\right\|_2^2$ where f_n is the empirical density function and $f_{\boldsymbol{\theta}}$ is the density function of a univariate $F_{\boldsymbol{\theta}}$. The *MHD* measure has the desirable property of being robust to outliers.

4.2 Consistency of Estimators

The following Lemmas are needed to establish the consistency of the estimators obtained by the MD_α estimation method.

Lemma 4.1: If $\left|\hat{G}_n(\mathbf{z}) - G_{\boldsymbol{\theta}_0}(\mathbf{z})\right| \xrightarrow{a.s.} 0$, then $\left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})\right| \xrightarrow{a.s.} 0$ as $n \to \infty$.

Proof:

From the Strong Law of Large Numbers, $\hat{G}_n(\mathbf{z}) \xrightarrow{a.s.} G_{\boldsymbol{\theta}_0}(\mathbf{z})$ when $n \to \infty$. By the continuous mapping theorem (Theorem 2.5), $\hat{G}_n^\alpha(\mathbf{z}) \xrightarrow{a.s.} G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})$. This implies that $\left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})\right| \xrightarrow{a.s.} 0$ as $n \to \infty$. (See also Proposition 3.1 in Remillard & Theodorescu, 2000.) ■

Lemma 4.2: Let $\hat{Q}_n(\boldsymbol{\theta}) = -\left\|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right\|_2^2$ and $Q_0(\boldsymbol{\theta}) = -\left\|G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right\|_2^2$, where $\boldsymbol{\theta} \in \Omega$. Assume that the parameter space Ω is compact. Then, $\left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})\right| \xrightarrow{a.s.} 0$ implies that $\sup_{\boldsymbol{\theta}\in\Omega}\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0$ as $n \to \infty$.

Proof:

$$\sup_{\boldsymbol{\theta}\in\Omega}\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right|$$

$$= \sup_{\boldsymbol{\theta}\in\Omega}\left|\int \left(\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right)^2 - \left(G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right)^2 d\mathbf{z}\right|$$

$$\leq \sup_{\boldsymbol{\theta}\in\Omega}\int \left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})\right|\left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z}) + 2\left[G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right]\right| d\mathbf{z}$$

$$\leq \int \left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})\right| \sup_{\boldsymbol{\theta}\in\Omega}\left(\left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z})\right| + 2\left|G_{\boldsymbol{\theta}_0}^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right|\right) d\mathbf{z}$$

on expanding and simplifying the terms under the integral sign. Since $\left|\hat{G}_n^{\alpha}(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^{\alpha}(\mathbf{z})\right| \xrightarrow{a.s.} 0$ (Lemma 4.1) and Ω is compact, $\sup_{\boldsymbol{\theta}\in\Omega}\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0$ as $n \to \infty$. ■

Proposition: Let $\hat{Q}_n(\boldsymbol{\theta}) = -\left\|\hat{G}_n^{\alpha}(\mathbf{z}) - G_{\boldsymbol{\theta}}^{\alpha}(\mathbf{z})\right\|_2^2$ and $Q_0(\boldsymbol{\theta}) = -\left\|G_{\boldsymbol{\theta}_0}^{\alpha}(\mathbf{z}) - G_{\boldsymbol{\theta}}^{\alpha}(\mathbf{z})\right\|_2^2$, and $\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}\in\Omega}\left(-\hat{Q}_n(\boldsymbol{\theta})\right) = \max_{\boldsymbol{\theta}\in\Omega}\hat{Q}_n(\boldsymbol{\theta})$. Assume that the parameter space Ω is compact. Then, $\left|\hat{G}_n^{\alpha}(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^{\alpha}(\mathbf{z})\right| \xrightarrow{a.s.} 0$ implies $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$.

Proof:

$Q_0(\boldsymbol{\theta})$ achieves a unique maximum of 0 at $\boldsymbol{\theta}_0$ and $Q_0(\boldsymbol{\theta})$ is a continuous function on Ω. From Lemma 4.1 and 4.2, $\left|\hat{G}_n^{\alpha}(\mathbf{z}) - G_{\boldsymbol{\theta}_0}^{\alpha}(\mathbf{z})\right| \xrightarrow{a.s.} 0$ implies $\sup_{\boldsymbol{\theta}\in\Omega}\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right| \xrightarrow{a.s.} 0$ as $n \to \infty$. That is, $\hat{Q}_n(\boldsymbol{\theta})$ converges uniformly almost surely to $Q_0(\boldsymbol{\theta})$ for large enough n. Hence, by Theorem 2.3, $\hat{\boldsymbol{\theta}} \xrightarrow{a.s.} \boldsymbol{\theta}_0$, that is, $\hat{\boldsymbol{\theta}}$ is strongly consistent. ■

4.3 Design of Simulation Study

The simulation study of MD_{α} estimation is developed for the *BNB* distribution using the FORTRAN programming language on computers (3GB RAM) running on Windows Vista. As many as 500 simulated *BNB* samples are taken to estimate the parameters with each sample size being either n = 100 or n = 500. The samples are simulated using the mixture formulation model *BNB-II* in Ong (1992), which is given below for convenience.

Let $X_1|k{\sim}Binomial(k,\phi_1)$ and $X_2|k{\sim}Binomial(k,\phi_2)$ where k is a value of the random variable $K{\sim}NB(\nu,\phi)$. Then, $(X_1,X_2){\sim}BNB(\nu,\theta_1,\theta_2,\theta_3)$ with the pgf given in (4.1) and $\theta_1 = \phi\phi_1(1-\phi_2)/\delta$, $\theta_2 = \phi\phi_2(1-\phi_1)/\delta$, $\theta_3 = \phi\phi_1\phi_2/\delta$ and $\delta = 1 -$

$\phi(1-\phi_1)(1-\phi_2)$. The marginals are $X_1 \sim NB(\nu, p_1)$ and $X_2 \sim NB(\nu, p_2)$ with $p_1 = (\theta_1 + \theta_3)/(1-\theta_2)$ and $p_2 = (\theta_2 + \theta_3)/(1-\theta_1)$.

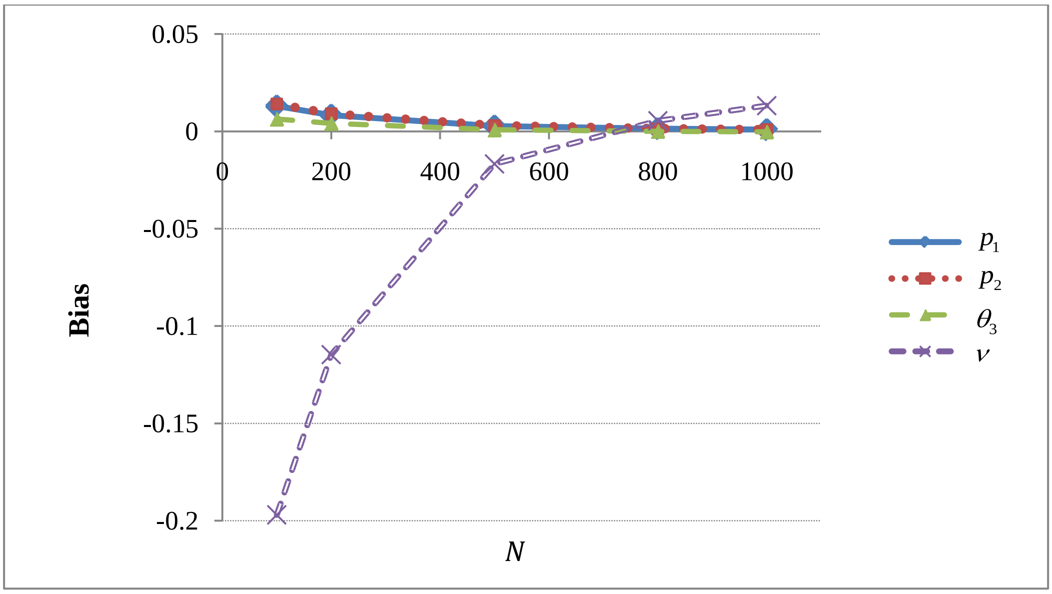


Figure 4.1 (Data in Table A1)

Bias Measures in Parameter from Simulations with ***N*** Monte Carlo Samples

To determine a feasible number of simulated samples for the simulation study, simulation runs with 100, 200, 500, 800 and 1000 Monte Carlo samples have been performed with a set of selected parameter values, $p_1 = 0.4$, $p_2 = 0.5$, $\theta_3 = 0.3$ and $\nu = 4.0$, and sample size, n = 500 for *BNB* distribution. It is found that 500 Monte Carlo samples are sufficient. For comparison, *MLE* is considered for parameter estimation in the simulations. The *ML* estimates are found to have stabilized with small parameter biases at about 500 Monte Carlo samples as shown in Figure 4.1. Table 4.1 below gives the computation time required for each of the simulation with different number of Monte Carlo samples. It is observed that the biasness in parameter estimates decreases as the number of samples increases. However, the computation time also increases. Although the biases are closest to zero with 800 Monte Carlo samples, it is decided to take 500 Monte Carlo samples for the simulation study as a trade off between computation time and a very slight difference in accuracy.

Table 4.1

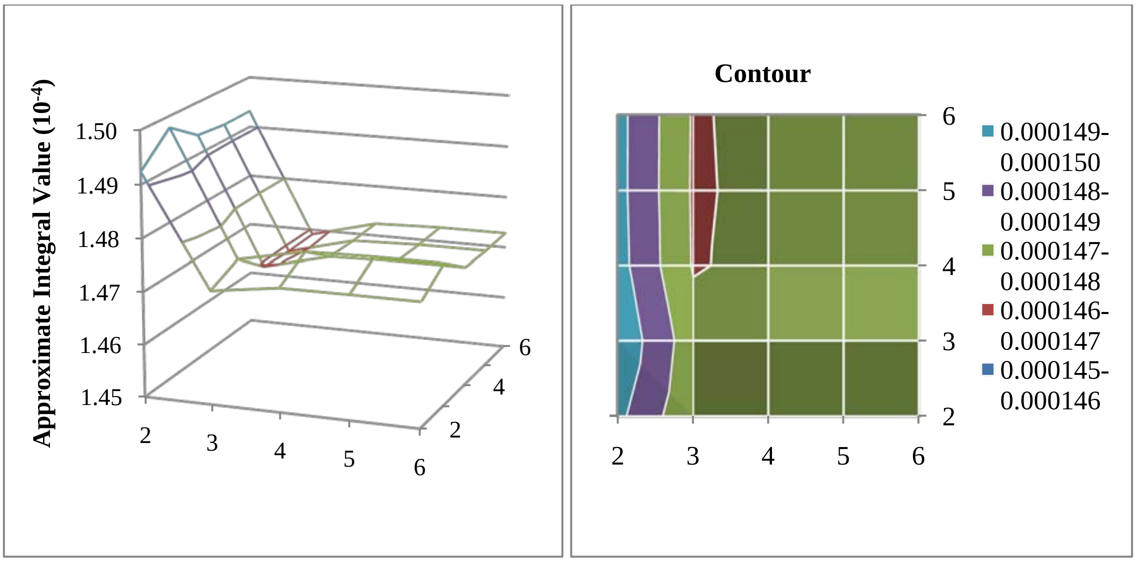
Computation Time for Simulation with N Monte Carlo Samples

Number of Monte Carlo Samples, N	100	200	500	800	1000
Computation Time (Minutes)	6.4684	12.4945	31.2583	48.5399	62.8610

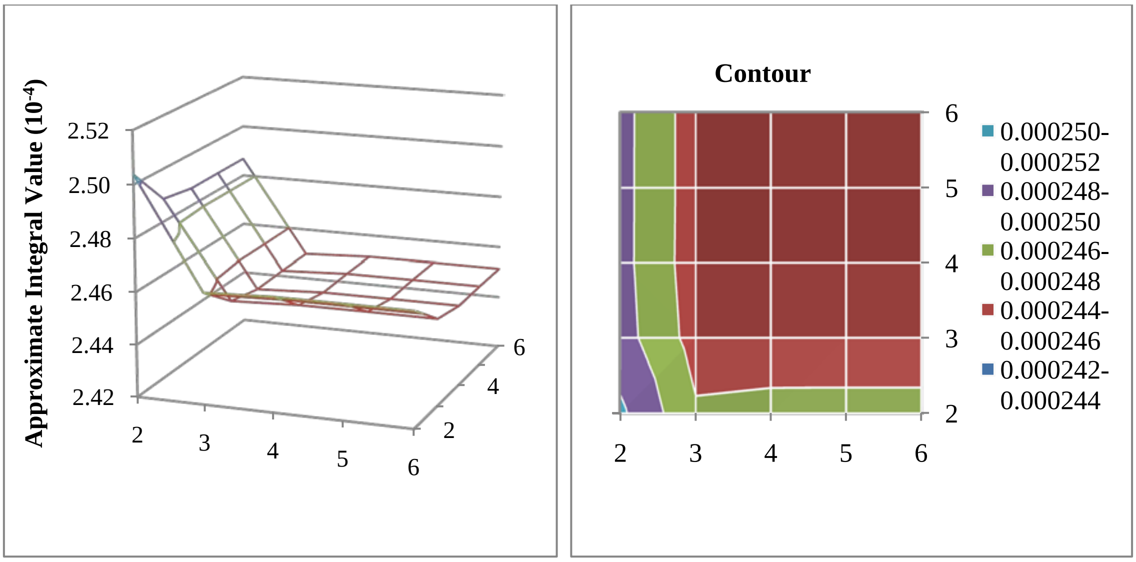
In the simulation study, the corresponding likelihood, distance or divergence measure is maximized or minimized by using simulated annealing technique. This is done over closed and bounded intervals for the parameters and the parameter space Ω may be assumed as compact. For bivariate distributions, the MD_α estimates are given by

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}\in\Omega}\left(\int_0^1\int_0^1\left|\hat{G}_n^\alpha(\mathbf{z}) - G_{\boldsymbol{\theta}}^\alpha(\mathbf{z})\right|^2 dz_1 dz_2\right). \quad (4.3)$$

The integral involved is numerically approximated by the Gauss quadrature method, which is known to work well in rectangular regions. The IMSL FORTRAN routine GQRUL or DGQRUL produces the quadrature points and corresponding weights required for the Gauss quadrature method. For bivariate distributions, the choice on the number of quadrature points used for each variable of integration in (4.3) is made based on the following empirical observations for $\alpha = 1$ and $\alpha = 1/2$ for the same set of parameter values, $p_1 = 0.4$, $p_2 = 0.5$, $\theta_3 = 0.3$ and $\nu = 4.0$, and sample size, $n = 500$ for *BNB* distribution.



a) MD_1



b) $MD_{1/2}$

Figure 4.2 (Data in Table A2)

Approximation with Corresponding Contour of the Distance Measure Integral Values

Figure 4.2 shows that the Gauss quadrature approximation method converges after 3 quadrature points are used for each variable of integration. Basically, the more quadrature points are used, the more accurate is the integral approximation and the longer is the computation duration. In order to make the most parsimonious choice in terms of computation time, the least number of points that leads to convergence is chosen. Here, that chosen number of quadrature points is 3. Thus, a 3×3 quadrature points are used in the approximation of the double integrals in the distance measure (4.3) for bivariate distribution. Furthermore, since the range of integration is over a narrow range from 0 to 1, three quadrature points are sufficient for a good approximation of a single integral.

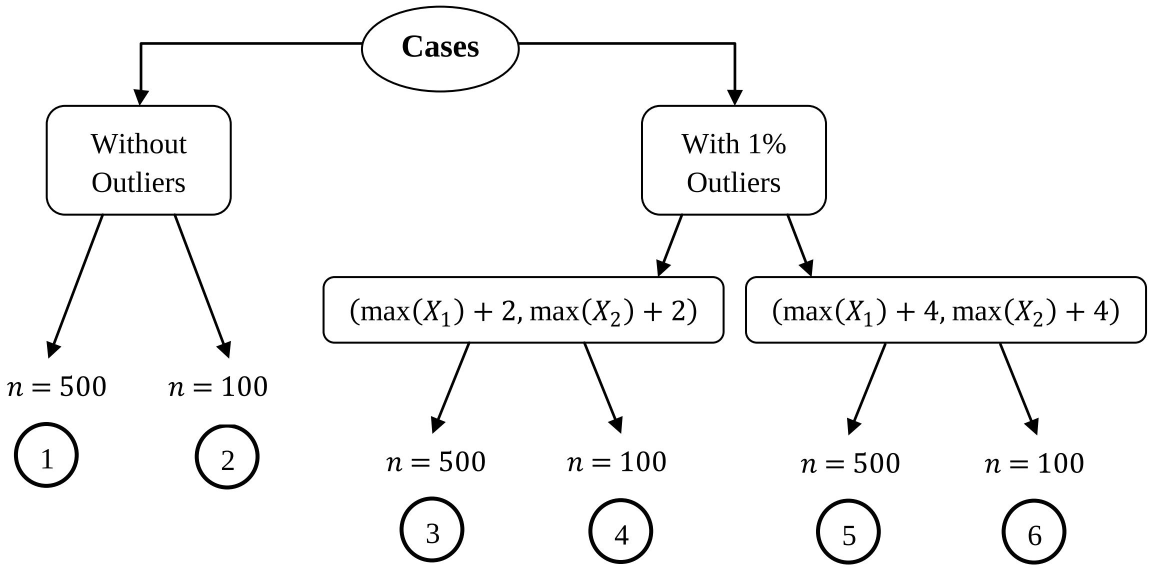


Figure 4.3

Cases Considered in the Simulation Study of MD_α Method

With the setup for simulation study as explained above, the behaviour of the MD_α method in the presence or absence of outliers are considered for six cases as mapped out in Figure 4.3. Simulated samples of size n = 100 or n = 500 are generated without outliers (Cases 1 and 2) or with an added 1% of outliers with respect to sample size n (Cases 3 to 6). The outliers are created by placing them at positions away from the rest of the data: a) 2 counts (Cases 3 and 4) and b) 4 counts (Cases 5 and 6) away from the maximum of X_1 and X_2, that is at the cell positions $(\max(X_1)+2, \max(X_2)+$

2) and $(\max(X_1)+4, \max(X_2)+4)$. On a computer, FORTRAN source codes for multiple cases of the study are executed simultaneously.

For Cases 1 and 2, the MD_1 and $MD_{1/2}$ parameter estimates are compared against the *ML* estimates. It is well known that *MLE* method is not robust against outliers. Therefore, for Cases 3 to 6 where outliers are present, the estimates from *MPGHD* instead of *MLE* are given for comparison with the estimates of MD_1 and $MD_{1/2}$ methods. *MGHD* method is affected by the empty cells in a data set whereas this effect has been corrected in the *MPGHD* method (Basu *et al*., 1997). Due to this and the fact that there are usually a number of empty cells in bivariate cases, the *MPGHD* estimates are chosen as the baseline for comparison in the simulation study when outliers are present in the data sets.

4.4 Simulation Results and Discussions

The simulation results along with the measures of bias and mean squared errors (MSE) for 4 different sets of selected parameters are given in the subsequent sections. The computation time for each case is given in minutes.

4.4.1 Without Outliers

Results shown in Tables 4.2 and 4.3 correspond to Cases 1 and 2. Figures 4.4 and 4.5 enable the graphical comparison of the parameter biases for the three estimation methods.

Table 4.2

Simulation Results when n = 500 with No Outliers (Computation Duration in Minutes)

n = 500	MD_1		$MD_{1/2}$		MLE	
Parameter	Estimate	Bias (MSE)	Estimate	Bias (MSE)	Estimate	Bias (MSE)
Set 1						
$p_1 = 0.4$	0.408227	0.008227 (0.002022)	0.412247	0.012247 (0.003345)	0.402726	0.002726 (0.001045)
$p_2 = 0.5$	0.509213	0.009213 (0.002145)	0.513122	0.013122 (0.003363)	0.502974	0.002974 (0.001117)
$\theta_3 = 0.3$	0.303968	0.003968 (0.002153)	0.302972	0.002972 (0.003299)	0.300814	0.000814 (0.000421)
$\nu = 4.0$	3.926721	-0.073279 (0.433761)	3.905012	-0.094988 (0.583802)	3.983285	-0.016715 (0.267652)
Duration:	7.249717		7.183417		30.444517	
Set 2						
$p_1 = 0.3$	0.310406	0.010406 (0.002319)	0.312532	0.012532 (0.002703)	0.304457	0.004457 (0.001527)
$p_2 = 0.2$	0.208510	0.008510 (0.001485)	0.209995	0.009995 (0.001769)	0.203796	0.003796 (0.000946)
$\theta_3 = 0.1$	0.099981	-0.000019 (0.000332)	0.101047	0.001047 (0.000997)	0.100031	0.000031 (0.000186)
$\nu = 4.0$	3.930723	-0.069277 (0.629331)	3.914868	-0.085132 (0.676913)	3.992861	-0.007139 (0.488250)
Duration:	2.547067		2.611883		6.592700	
Set 3						
$p_1 = 0.8$	0.807946	0.007946 (0.000864)	0.806411	0.006411 (0.001305)	0.801965	0.001965 (0.000186)
$p_2 = 0.6$	0.615890	0.015890 (0.002931)	0.611201	0.011201 (0.004944)	0.603264	0.003264 (0.000433)
$\theta_3 = 0.2$	0.210276	0.010276 (0.017321)	0.202912	0.002912 (0.017309)	0.195467	-0.004533 (0.000452)
$\nu = 4.0$	3.838572	-0.161428 (0.368277)	3.885545	-0.114455 (0.531359)	3.949643	-0.050357 (0.102458)
Duration:	48.013967		43.380467		283.006617	
Set 4						
$p_1 = 0.4$	0.401457	0.001457 (0.003775)	0.400839	0.000839 (0.003664)	0.398351	-0.001649 (0.002347)
$p_2 = 0.5$	0.500336	0.000336 (0.004037)	0.500052	0.000052 (0.004098)	0.497767	-0.002233 (0.002515)
$\theta_3 = 0.3$	0.302111	0.002111 (0.002537)	0.298424	-0.001576 (0.002276)	0.298654	-0.001346 (0.001043)
$\nu = 0.5$	0.514846	0.014846 (0.014462)	0.515526	0.015526 (0.014808)	0.513978	0.013978 (0.010288)
Duration:	2.019550		2.151950		4.760933	

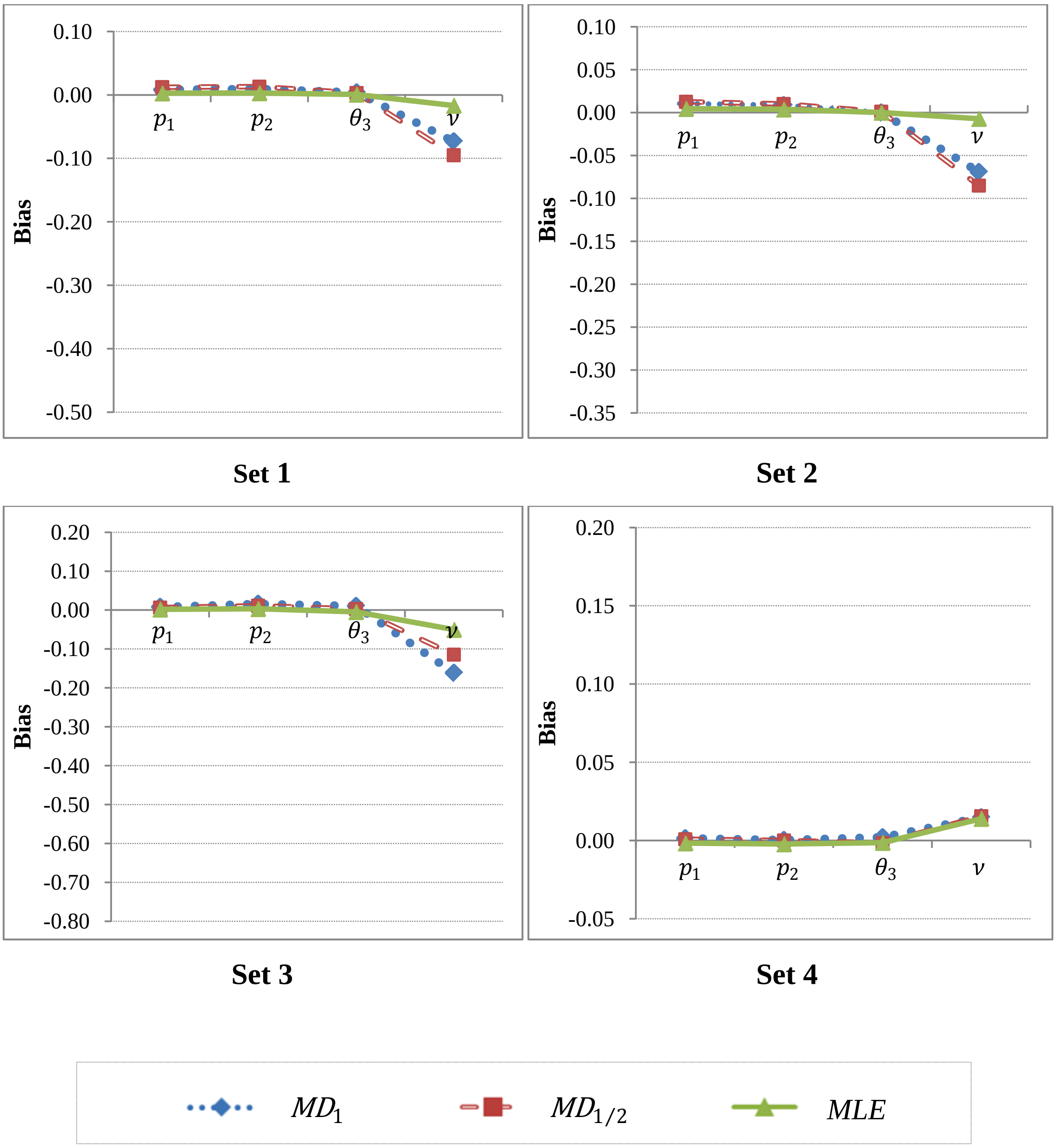


Figure 4.4

Bias Measures in Parameter for MD_1, $MD_{1/2}$ and MLE Methods when n = 500 with No Outliers

Table 4.3

Simulation Results when n = 100 with No Outliers (Computation Duration in Minutes)

n = 100	MD_1		$MD_{1/2}$		MLE	
Parameter	Estimate	Bias (MSE)	Estimate	Bias (MSE)	Estimate	Bias (MSE)
Set 1						
$p_1 = 0.4$	0.439091	0.039091 (0.006782)	0.447210	0.047210 (0.008752)	0.411839	0.011839 (0.003485)
$p_2 = 0.5$	0.537254	0.037254 (0.006893)	0.546282	0.046282 (0.008308)	0.510670	0.010670 (0.003500)
$\theta_3 = 0.3$	0.319004	0.019004 (0.010057)	0.322407	0.022407 (0.008896)	0.306324	0.006324 (0.001483)
$\nu = 4.0$	3.628951	-0.371049 (1.085702)	3.569747	-0.430253 (1.230148)	3.901468	-0.098532 (0.720146)
Duration:	5.222683		4.747600		20.848167	
Set 2						
$p_1 = 0.3$	0.338202	0.038202 (0.007443)	0.341767	0.041767 (0.008168)	0.319231	0.019231 (0.004686)
$p_2 = 0.2$	0.231046	0.031046 (0.005471)	0.233779	0.033779 (0.006203)	0.216005	0.016005 (0.003097)
$\theta_3 = 0.1$	0.103050	0.003050 (0.001872)	0.103538	0.003538 (0.001911)	0.103604	0.003604 (0.000770)
$\nu = 4.0$	3.693848	-0.306152 (1.454646)	3.673803	-0.326197 (1.527762)	3.850516	-0.149484 (1.091255)
Duration:	1.727950		1.815200		4.981767	
Set 3						
$p_1 = 0.8$	0.833766	0.033766 (0.002869)	0.827178	0.027178 (0.003376)	0.806697	0.006697 (0.000774)
$p_2 = 0.6$	0.670426	0.070426 (0.009263)	0.659241	0.059241 (0.013945)	0.611140	0.011140 (0.001846)
$\theta_3 = 0.2$	0.333721	0.133721 (0.086209)	0.317735	0.117735 (0.073302)	0.209286	0.009286 (0.002445)
$\nu = 4.0$	3.302221	-0.697779 (0.906485)	3.444127	-0.555873 (1.232588)	3.828822	-0.171178 (0.405671)
Duration:	31.226883		28.775733		179.798283	
Set 4						
$p_1 = 0.4$	0.378067	-0.021933 (0.015575)	0.376454	-0.023546 (0.015579)	0.369597	-0.030403 (0.012272)
$p_2 = 0.5$	0.471649	-0.028351 (0.01817)	0.469752	-0.030248 (0.018248)	0.463906	-0.036094 (0.014162)
$\theta_3 = 0.3$	0.280870	-0.019130 (0.009124)	0.277538	-0.022462 (0.009029)	0.275199	-0.024801 (0.005609)
$\nu = 0.5$	0.666446	0.166446 (0.286226)	0.671162	0.171162 (0.290995)	0.662190	0.162190 (0.240780)
Duration:	1.353683		1.407683		2.667183	

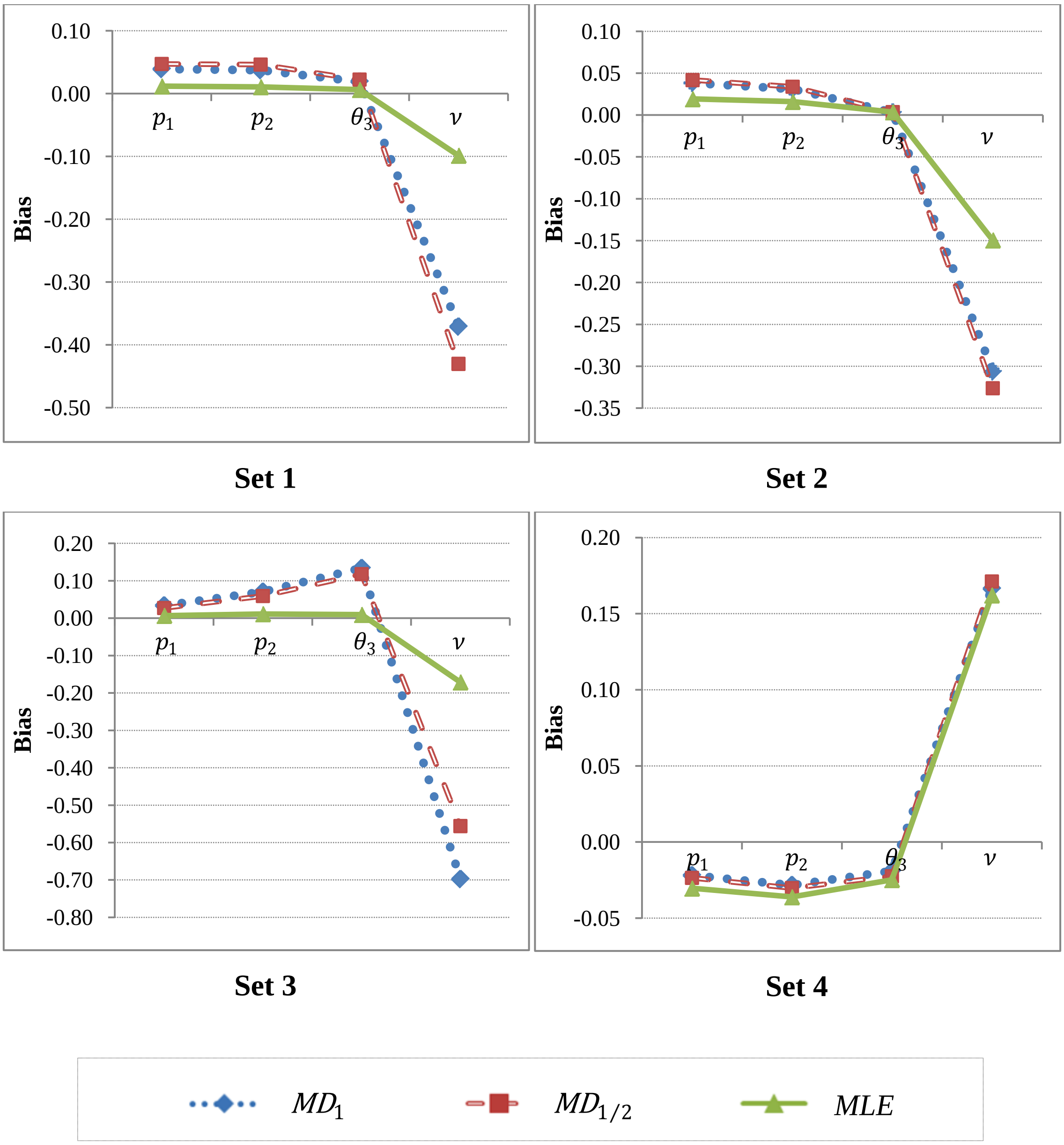


Figure 4.5

Bias Measures in Parameter for MD_1, $MD_{1/2}$ and MLE Methods when n = 100 with No Outliers

4.4.2 With Outliers

a) Outliers Positioned at Cell $(\max(X_1) + 2, \max(X_2) + 2)$

Results shown in Tables 4.4 and 4.5 correspond to Cases 3 and 4. Figures 4.6 and 4.7 illustrate the parameter biases for the three estimation methods.

Table 4.4

Simulation Results when n = 500 with 1% Outliers at Cell $(\max(X_1) + 2, \max(X_2) + 2)$ (Computation Duration in Minutes)

n = 500	MD_1		$MD_{1/2}$		$MPGHD$	
Parameter	Estimate	Bias (MSE)	Estimate	Bias (MSE)	Estimate	Bias (MSE)
Set 1						
$p_1 = 0.4$	0.422211	0.022211 (0.002281)	0.419452	0.019452 (0.002917)	0.389542	-0.010458 (0.002907)
$p_2 = 0.5$	0.524060	0.024060 (0.002374)	0.521029	0.021029 (0.002966)	0.488007	-0.011993 (0.002994)
$\theta_3 = 0.3$	0.307559	0.007559 (0.002493)	0.308892	0.008892 (0.002151)	0.276072	-0.023928 (0.004714)
$\nu = 4.0$	3.772157	-0.227843 (0.438702)	3.825498	-0.174502 (0.546128)	4.180034	0.180034 (0.677924)
Duration:	8.282667		7.951117		53.413233	
Set 2						
$p_1 = 0.3$	0.332773	0.032773 (0.003479)	0.327607	0.027607 (0.004079)	0.281029	-0.018971 (0.003383)
$p_2 = 0.2$	0.226240	0.026240 (0.002199)	0.222509	0.022509 (0.002567)	0.185757	-0.014243 (0.001965)
$\theta_3 = 0.1$	0.105798	0.005798 (0.000387)	0.104459	0.004459 (0.000460)	0.090491	-0.009509 (0.000545)
$\nu = 4.0$	3.706760	-0.293240 (1.183116)	3.824134	-0.175866 (1.540243)	4.547886	0.547886 (1.765489)
Duration:	2.241467		2.245483		11.107750	
Set 3						
$p_1 = 0.8$	0.809059	0.009059 (0.002201)	0.808141	0.008141 (0.001253)	0.771647	-0.028353 (0.000936)
$p_2 = 0.6$	0.618380	0.018380 (0.003421)	0.615145	0.015145 (0.004612)	0.559410	-0.040590 (0.001408)
$\theta_3 = 0.2$	0.204354	0.004354 (0.016551)	0.201095	0.001095 (0.016569)	0.202333	0.002333 (0.005205)
$\nu = 4.0$	3.802067	-0.197933 (0.384546)	3.865058	-0.134942 (0.519880)	4.483183	0.483183 (0.397416)
Duration:	38.787217		34.715767		508.795900	
Set 4						
$p_1 = 0.4$	0.463709	0.063709 (0.003238)	0.460404	0.060404 (0.003905)	0.392709	-0.007291 (0.007641)
$p_2 = 0.5$	0.565306	0.065306 (0.003335)	0.561331	0.061331 (0.003969)	0.489423	-0.010577 (0.008330)
$\theta_3 = 0.3$	0.329790	0.029790 (0.002217)	0.329308	0.029308 (0.002178)	0.276729	-0.023271 (0.005601)
$\nu = 0.5$	0.436097	-0.063903 (0.009533)	0.438617	-0.061383 (0.008975)	0.540314	0.040314 (0.129224)
Duration:	2.422017		2.435850		8.128383	

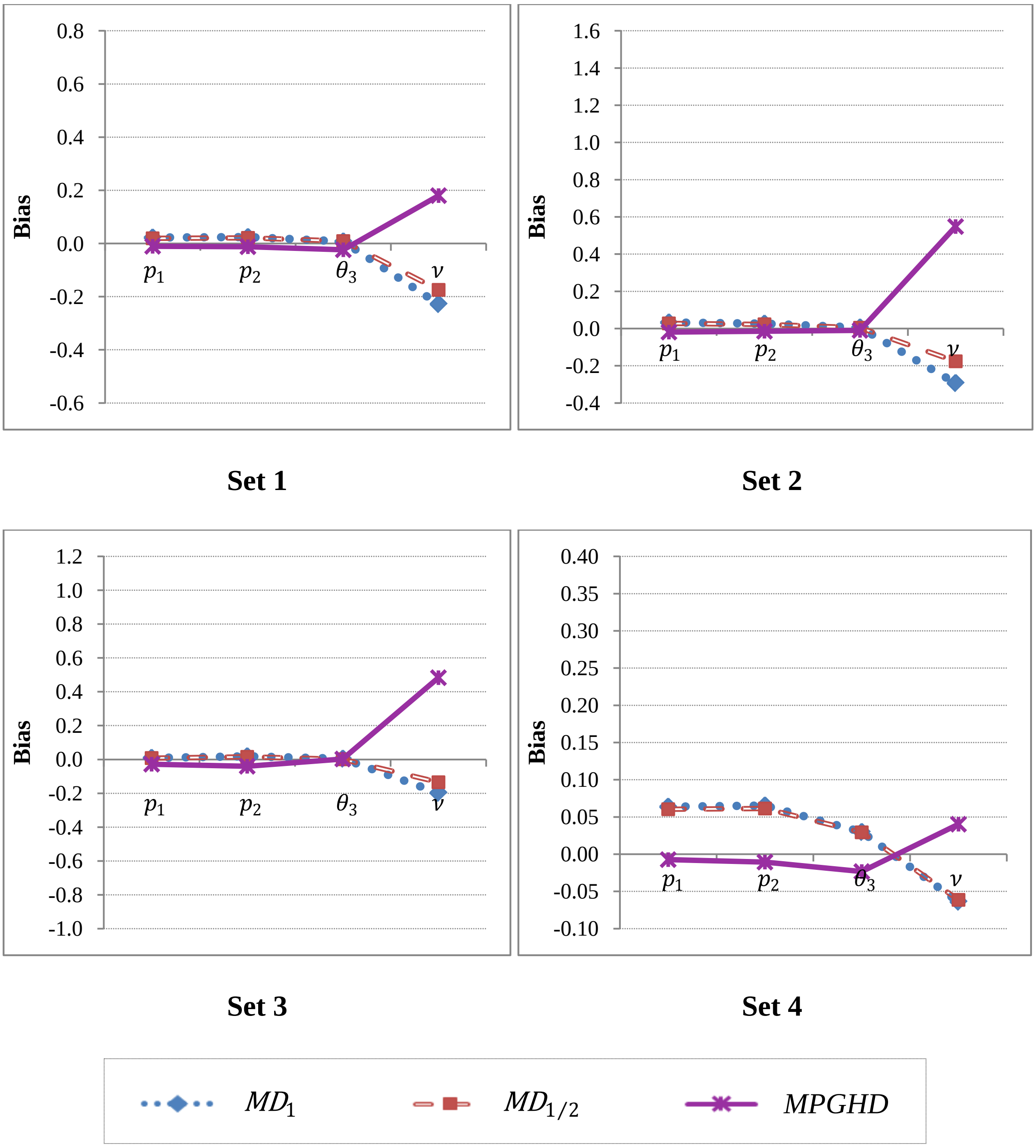


Figure 4.6

Bias Measures in Parameter for MD_1, $MD_{1/2}$ and $MPGHD$ Methods when n = 500 with 1% Outliers at Cell $(\max(X_1) + 2, \max(X_2) + 2)$

Table 4.5

Simulation Results when n = 100 with 1% Outliers at Cell $(\max(X_1) + 2, \max(X_2) + 2)$ (Computation Duration in Minutes)

n = 100	MD_1		$MD_{1/2}$		$MPGHD$	
Parameter	Estimate	Bias (MSE)	Estimate	Bias (MSE)	Estimate	Bias (MSE)
Set 1						
$p_1 = 0.4$	0.449382	0.049382 (0.007017)	0.454757	0.054757 (0.008915)	0.378850	-0.021150 (0.010482)
$p_2 = 0.5$	0.548291	0.048291 (0.006935)	0.553573	0.053573 (0.008357)	0.472756	-0.027244 (0.011497)
$\theta_3 = 0.3$	0.326028	0.026028 (0.009406)	0.326966	0.026966 (0.008700)	0.264465	-0.035535 (0.008494)
$\nu = 4.0$	3.536089	-0.463911 (1.083552)	3.511678	-0.488322 (1.228945)	4.642401	0.642401 (4.186887)
Duration:	5.637483		5.488150		27.894550	
Set 2						
$p_1 = 0.3$	0.333663	0.033663 (0.014513)	0.331141	0.031141 (0.016069)	0.266263	-0.033737 (0.010789)
$p_2 = 0.2$	0.231136	0.031136 (0.009356)	0.229661	0.029661 (0.010693)	0.176754	-0.023246 (0.00628)
$\theta_3 = 0.1$	0.105329	0.005329 (0.002165)	0.103766	0.003766 (0.002271)	0.084025	-0.015975 (0.001588)
$\nu = 4.0$	4.419362	0.419362 (7.396000)	4.549245	0.549245 (7.979569)	5.479294	1.479294 (7.888147)
Duration:	1.599667		1.773217		10.815117	
Set 3						
$p_1 = 0.8$	0.836548	0.036548 (0.002398)	0.818694	0.018694 (0.006432)	0.739273	-0.060727 (0.005797)
$p_2 = 0.6$	0.673529	0.073529 (0.009041)	0.651875	0.051875 (0.016592)	0.522854	-0.077146 (0.007572)
$\theta_3 = 0.2$	0.331425	0.131425 (0.086525)	0.330248	0.130248 (0.070930)	0.214878	0.014878 (0.008776)
$\nu = 4.0$	3.280508	-0.719492 (0.897631)	3.617985	-0.382015 (2.204145)	5.040627	1.040627 (2.557858)
Duration:	34.226117		23.243200		290.589000	
Set 4						
$p_1 = 0.4$	0.433973	0.033973 (0.013782)	0.433117	0.033117 (0.013804)	0.349438	-0.050562 (0.024963)
$p_2 = 0.5$	0.531051	0.031051 (0.014653)	0.530921	0.030921 (0.014270)	0.436059	-0.063941 (0.02811)
$\theta_3 = 0.3$	0.309398	0.009398 (0.007973)	0.312606	0.012606 (0.008696)	0.237155	-0.062845 (0.012549)
$\nu = 0.5$	0.535754	0.035754 (0.098187)	0.532747	0.032747 (0.089141)	0.847953	0.347953 (1.318471)
Duration:	1.580283		1.698317		4.430650	

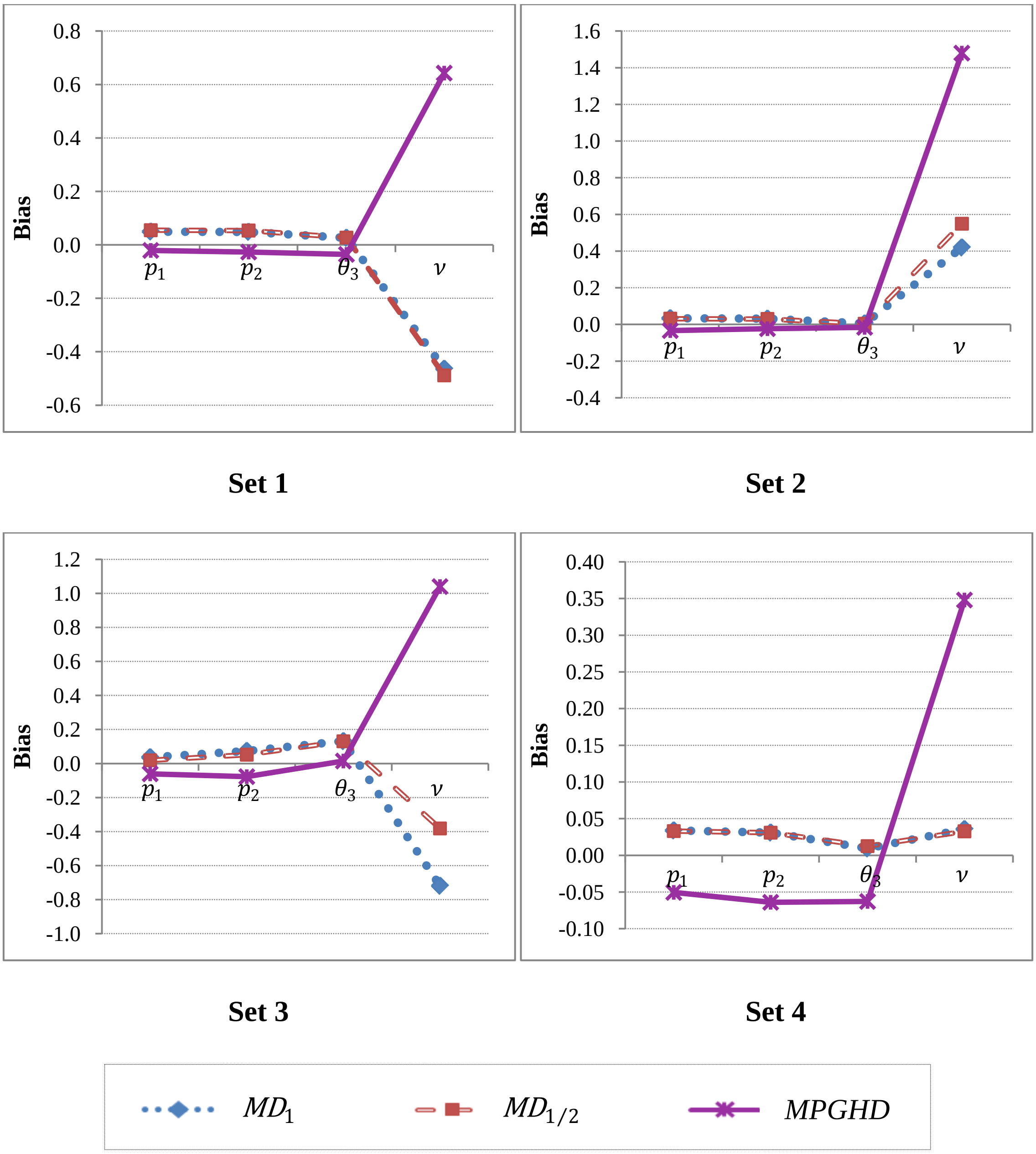


Figure 4.7

Bias Measures in Parameter for MD_1, $MD_{1/2}$ and $MPGHD$ Methods when n = 100 with 1% Outliers at Cell $(\max(X_1) + 2, \max(X_2) + 2)$

b) Outliers Positioned at Point $(\max(X_1) + 4, \max(X_2) + 4)$

Results shown in Tables 4.6 and 4.7 correspond to Cases 5 and 6. Figures 4.8 and 4.9 illustrate the parameter biases for the three estimation methods.

Table 4.6

Simulation Results when n = 500 with 1% Outliers at Cell $(\max(X_1) + 4, \max(X_2) + 4)$ (Computation Duration in Minutes)

n = 500	MD_1		$MD_{1/2}$		$MPGHD$	
Parameter	Estimate	Bias (MSE)	Estimate	Bias (MSE)	Estimate	Bias (MSE)
Set 1						
$p_1 = 0.4$	0.422070	0.022070 (0.002138)	0.421465	0.021465 (0.003167)	0.380728	-0.019272 (0.002106)
$p_2 = 0.5$	0.523961	0.023961 (0.002230)	0.523095	0.023095 (0.003217)	0.479284	-0.020716 (0.002378)
$\theta_3 = 0.3$	0.309070	0.009070 (0.002175)	0.307041	0.007041 (0.002831)	0.275191	-0.024809 (0.003976)
$\nu = 4.0$	3.771453	-0.228547 (0.424643)	3.804836	-0.195164 (0.564412)	4.296306	0.296306 (0.669366)
Duration:	10.456533		10.035150		62.898683	
Set 2						
$p_1 = 0.3$	0.334854	0.034854 (0.003460)	0.329040	0.029040 (0.004259)	0.276199	-0.023801 (0.003828)
$p_2 = 0.2$	0.227762	0.027762 (0.002196)	0.223014	0.023014 (0.002594)	0.182251	-0.017749 (0.002191)
$\theta_3 = 0.1$	0.106003	0.006003 (0.000383)	0.105697	0.005697 (0.000572)	0.086231	-0.013769 (0.000662)
$\nu = 4.0$	3.673041	-0.326959 (1.138937)	3.810375	-0.189625 (1.529734)	4.671142	0.671142 (2.009540)
Duration:	3.114867		3.125150		18.089000	
Set 3						
$p_1 = 0.8$	0.809613	0.009613 (0.000946)	0.808140	0.008140 (0.001260)	0.772264	-0.027736 (0.000552)
$p_2 = 0.6$	0.617973	0.017973 (0.003124)	0.615230	0.015230 (0.004634)	0.558206	-0.041794 (0.001226)
$\theta_3 = 0.2$	0.211857	0.011857 (0.016471)	0.202424	0.002424 (0.016837)	0.203340	0.003340 (0.004951)
$\nu = 4.0$	3.830413	-0.169587 (0.420453)	3.866399	-0.133601 (0.524173)	4.491988	0.491988 (0.290153)
Duration:	43.720733		38.977900		599.812517	
Set 4						
$p_1 = 0.4$	0.465812	0.065812 (0.003565)	0.464836	0.064836 (0.003079)	0.377906	-0.022094 (0.007218)
$p_2 = 0.5$	0.567438	0.067438 (0.003769)	0.566174	0.066174 (0.003229)	0.475530	-0.024470 (0.008587)
$\theta_3 = 0.3$	0.329559	0.029559 (0.002395)	0.329565	0.029565 (0.002538)	0.260173	-0.039827 (0.006465)
$\nu = 0.5$	0.440250	-0.059750 (0.042225)	0.434281	-0.065719 (0.008292)	0.564330	0.064330 (0.091400)
Duration:	3.576550		3.597150		14.076400	

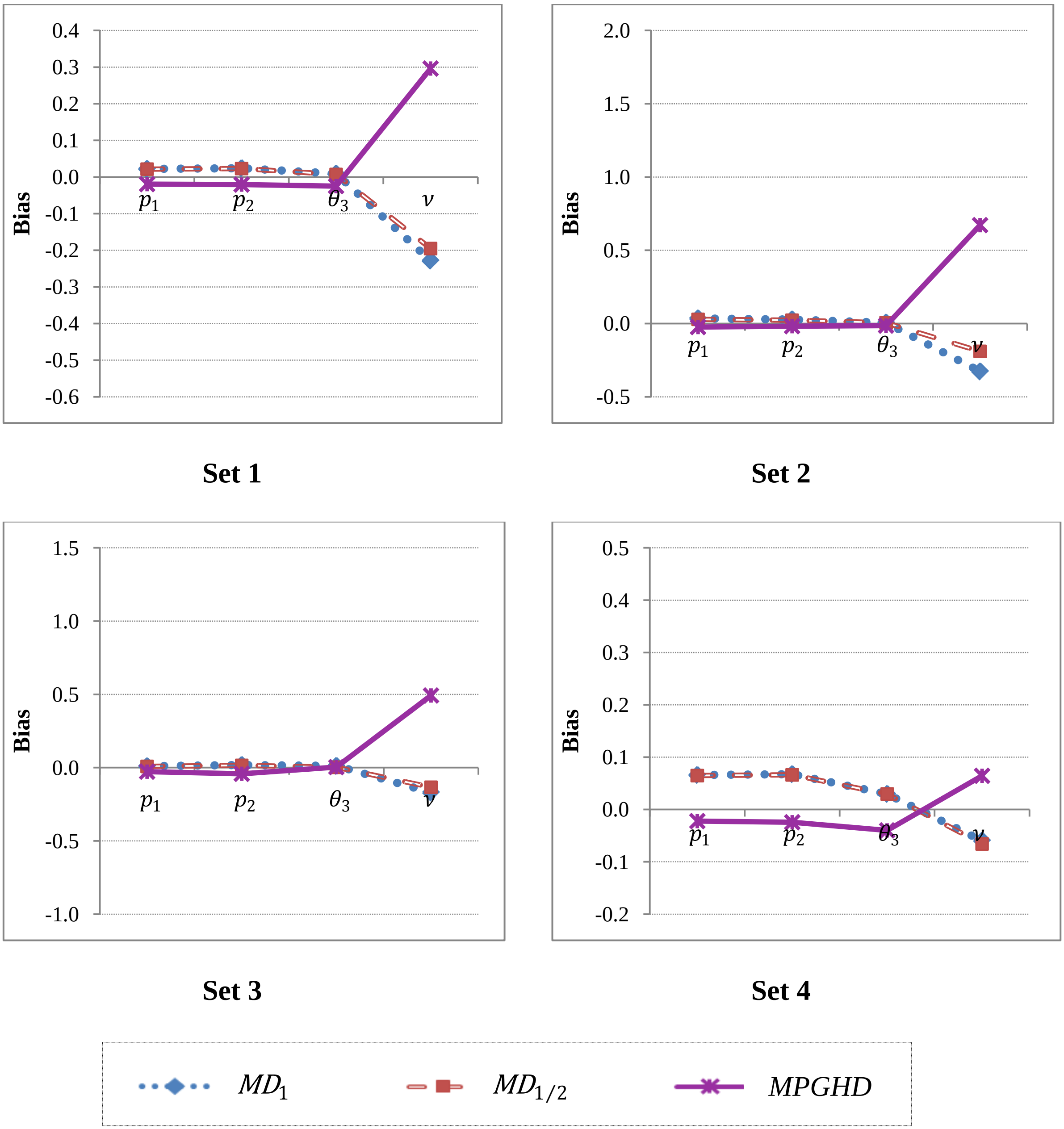


Figure 4.8

Bias Measures in Parameter for MD_1, $MD_{1/2}$ and $MPGHD$ Methods when n = 500 with 1% Outliers at Cell $(\max(X_1) + 4, \max(X_2) + 4)$

Table 4.7

Simulation Results when n = 100 with 1% Outliers at Cell $(\max(X_1) + 4, \max(X_2) + 4)$ (Computation Duration in Minutes)

n = 100	MD_1		$MD_{1/2}$		$MPGHD$	
Parameter	Estimate	Bias (MSE)	Estimate	Bias (MSE)	Estimate	Bias (MSE)
Set 1						
$p_1 = 0.4$	0.450658	0.050658 (0.007201)	0.454963	0.054963 (0.008926)	0.379540	-0.020460 (0.002566)
$p_2 = 0.5$	0.549609	0.049609 (0.007065)	0.554039	0.054039 (0.008402)	0.477970	-0.022030 (0.002689)
$\theta_3 = 0.3$	0.324154	0.024154 (0.009883)	0.326548	0.026548 (0.008742)	0.268189	-0.031811 (0.005052)
$\nu = 4.0$	3.523628	-0.476372 (1.096381)	3.508175	-0.491825 (1.230844)	4.323036	0.323036 (0.733868)
Duration:	7.348733		7.198533		63.313900	
Set 2						
$p_1 = 0.3$	0.335584	0.035584 (0.014431)	0.334721	0.034721 (0.016083)	0.251251	-0.048749 (0.009323)
$p_2 = 0.2$	0.232344	0.032344 (0.009371)	0.232483	0.032483 (0.010644)	0.165418	-0.034582 (0.005165)
$\theta_3 = 0.1$	0.106021	0.006021 (0.002109)	0.105684	0.005684 (0.002687)	0.078745	-0.021255 (0.001399)
$\nu = 4.0$	4.379788	0.379788 (7.252881)	4.480790	0.480790 (7.874835)	5.805823	1.805823 (7.949555)
Duration:	2.622700		2.403033		12.331117	
Set 3						
$p_1 = 0.8$	0.836411	0.036411 (0.002385)	0.825257	0.025257 (0.004918)	0.736315	-0.063685 (0.00342)
$p_2 = 0.6$	0.673686	0.073686 (0.008920)	0.659811	0.059811 (0.013613)	0.512885	-0.087115 (0.005762)
$\theta_3 = 0.2$	0.335692	0.135692 (0.085331)	0.320634	0.120634 (0.074232)	0.216332	0.016332 (0.008286)
$\nu = 4.0$	3.281912	-0.718088 (0.892124)	3.444059	-0.555941 (1.220883)	5.162477	1.162477 (2.447958)
Duration:	29.896083		42.869033		353.604550	
Set 4						
$p_1 = 0.4$	0.440168	0.040168 (0.013785)	0.438470	0.038470 (0.013970)	0.310459	-0.089541 (0.017217)
$p_2 = 0.5$	0.538511	0.038511 (0.014518)	0.535883	0.035883 (0.014645)	0.395997	-0.104003 (0.02164)
$\theta_3 = 0.3$	0.312904	0.012904 (0.008893)	0.311652	0.011652 (0.007624)	0.217228	-0.082772 (0.00978)
$\nu = 0.5$	0.521576	0.021576 (0.088694)	0.529733	0.029733 (0.105775)	0.891716	0.391716 (0.946063)
Duration:	2.524750		2.580833		8.349650	

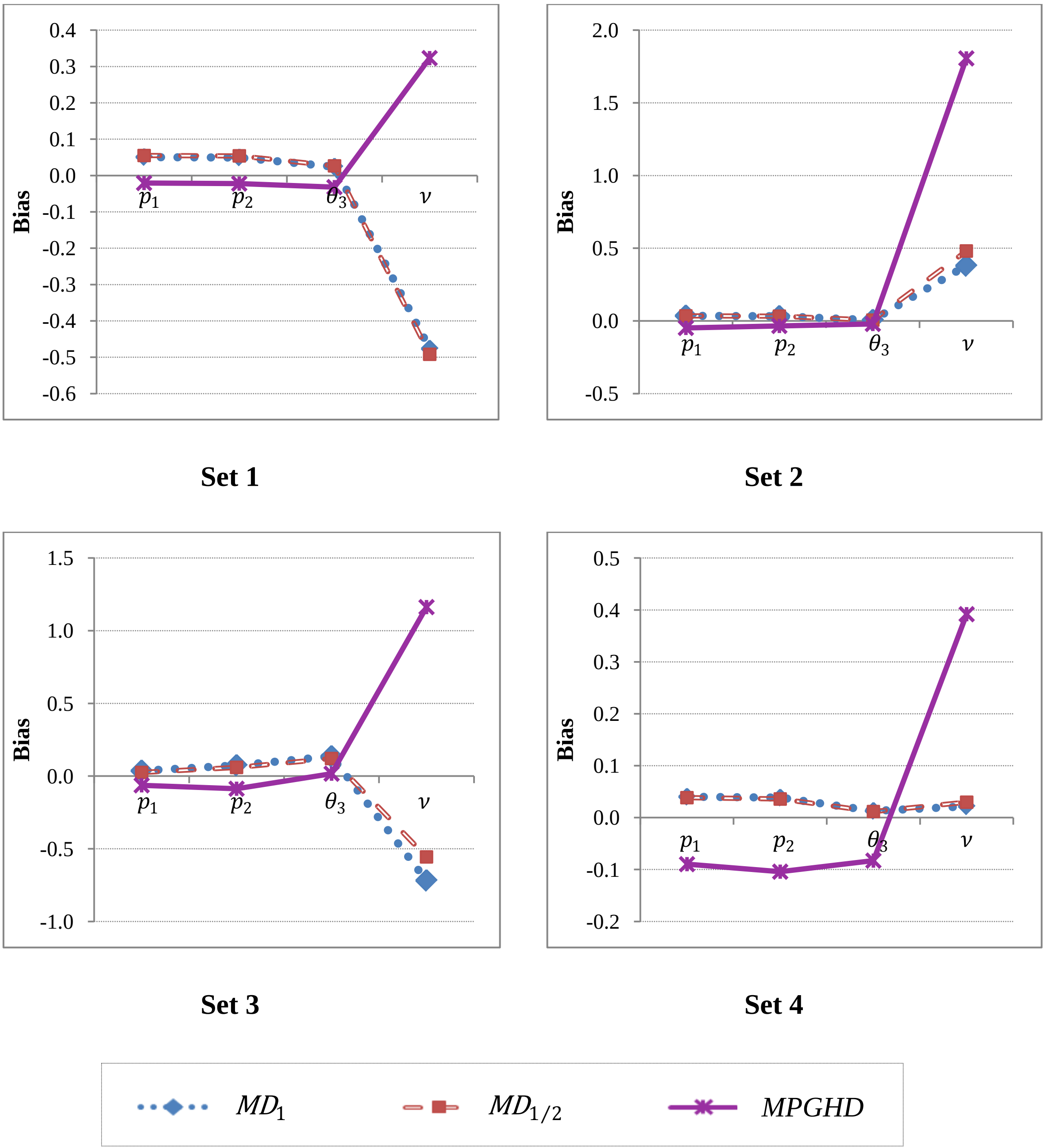


Figure 4.9

Bias Measures in Parameter for MD_1, $MD_{1/2}$ and $MPGHD$ Methods when n = 100 with 1% Outliers at Cell $(\max(X_1) + 4, \max(X_2) + 4)$

4.4.3 Discussion

For well-behaved data sets (no outliers), estimates for MD_1 and $MD_{1/2}$ methods are comparably close to the true values for n = 500. However, as with most minimum divergence methods, the larger parameter biases indicate that estimations for the smaller sample size, n = 100 deviate somewhat more from the true parameters. For all three estimation methods, the estimates for the index parameter ν deviate most from the true value as clearly shown in Figures 4.4 and 4.5. The estimates from the *MLE* method remain rather close to the true parameters as expected.

On the other hand, MD_1 and $MD_{1/2}$ methods are not greatly affected by the presence of outliers as shown by the good parameter estimations of all Sets 1 to 4 for both sample sizes in Tables 4.4 to 4.7. The results suggest that the estimates for MD_1 and $MD_{1/2}$ methods are comparable to that of *MPGHD*. In spite of the similar performances, MD_1 and $MD_{1/2}$ tend to have smaller biases for the estimates of ν compared to *MPGHD*. On top of that, due to empty cells, *MPGHD* requires a correction or penalty in its procedure which has not been supported theoretically (Basu *et al*., 1997) whereas MD_1 and $MD_{1/2}$ methods do not need such correction.

The simulation study also reveals that both MD_1 and $MD_{1/2}$ methods produced very similar results with no particular advantage to either method in terms of computation time. In comparison, the computation times for *MLE* and *MPGHD* methods are notably (at least twice) longer than that for MD_1 or $MD_{1/2}$ method. For most of the data sets, the computation time is 4 to 6 times longer than is needed for MD_1 or $MD_{1/2}$ method. Hence, it can be deduced that MD_1, $MD_{1/2}$ or, in general, MD_α method can be recommended for fast, robust parameter estimation.

4.5 Examples of Goodness-of-Fit

4.5.1 Univariate Case

To illustrate goodness-of-fit, a data set on the distribution of the number of infestation spots by the southern pine beetle, *Dendroctonus frontalis* Zimmermann, (Coleoptera: Scolytidae), in 5’ × 5’ geographic areas in Southeast Texas obtained from

Lin (1985) is used. This data set is fitted with the univariate distribution proposed by Sugita *et al*. (to appear), which has a relatively simple pgf given by

$$G(z)=\left(\frac{1-p_3}{1-p_3 z}\right)^{k_1}\left(\frac{p_1+p_2 z}{p_1+p_2}\right)^{k_2}\frac{{}_1F_1\ (k_1;k_2+1;\lambda\,(p_1+p_2 z)/(1-p_3 z))}{{}_1F_1\ (k_1;k_2+1;\lambda)}$$

where $k_1\geq 0$, $p_1,p_2,p_3>0$, $p_1+p_2+p_3=1$, $\lambda>0$ and k_2 is a non-negative integer or $k_2+1<0$ (k_2 real), as compared to its complicated pmf as follows.

(A) When k_2 is a non-negative integer,

(I) If $x\leq k_2$,

$$\Pr(X=x)=\sum_{s=0}^{x}\frac{(k_1)_{x-s}}{(x-s)!}(1-p_3)^{k_1}p_3^{x-s}\binom{k_2}{s}$$

$$\times\left(\frac{p_1}{p_1+p_2}\right)^{k_2-s}\left(\frac{p_2}{p_1+p_2}\right)^{s}\frac{{}_1F_1\ (k_1+x-s;k_2+1-s;\lambda p_1)}{{}_1F_1\ (k_1;k_2+1;\lambda)}.$$

(II) If $x>k_2$,

$$\Pr(X=x)=\sum_{s=0}^{k_2}\frac{(k_1)_{x-s}}{(x-s)!}(1-p_3)^{k_1}p_3^{x-s}\binom{k_2}{s}$$

$$\times\left(\frac{p_1}{p_1+p_2}\right)^{k_2-s}\left(\frac{p_2}{p_1+p_2}\right)^{s}\frac{{}_1F_1\ (k_1+x-s;k_2+1-s;\lambda p_1)}{{}_1F_1\ (k_1;k_2+1;\lambda)}$$

$$+\sum_{s=k_2+1}^{x}\frac{(k_1)_{x-k_2}}{(x-s)!}(1-p_3)^{k_1}p_3^{x-s}$$

$$\times\frac{k_2!}{s!\,(s-k_2)!}\lambda^{s-k_2}\frac{p_2^s}{(p_1+p_2)^{k_2}}\frac{{}_1F_1\ (k_1+x-k_2;s-k_2+1;\lambda p_1)}{{}_1F_1\ (k_1;k_2+1;\lambda)}.$$

(B) When $k_2+1<0$ (k_2 real),

$$\Pr(X=x)=\sum_{s=0}^{x}\frac{(k_1-k_2)_x}{(x-s)!}(1-p_3)^{k_1-k_2}p_3^{x-s}$$

$$\times\frac{(\lambda p_2)^s}{s!\,(1-k_2)_s}\frac{{}_1F_1\ (k_1+x-k_2;s-k_2+1;\lambda p_1)}{{}_1F_1\ (k_1;k_2+1;\lambda)}.$$

MLE, MD_1 and $MD_{1/2}$ estimations have been done by numerical optimization using simulated annealing method. The expected frequencies and corresponding χ^2 values as well as degrees of freedom (d.f.) are shown in Table 4.8.

Table 4.8

Observed and Expected Frequencies of Geographic Areas with Corresponding Number of Infestation Spots by the Southern Pine Beetle (Lin, 1985)

Number of Spots, x	Number of Areas			
	Observed	Expected		
		MLE	MD_1	$MD_{1/2}$
0	1169	1168.7975	1168.5271	1168.5480
1	144	150.8730	149.1437	149.0216
2	92	81.0680	82.5711	82.6389
3	54	50.5578	51.9411	52.0093
4	29	32.9782	33.7774	33.8179
5	18	21.8774	22.1767	22.1923
6	10	14.6067	14.5800	14.5792
7	12	9.7701	9.5668	9.5572
8	6	6.5322	6.2564	6.2433
9	9	4.3606	4.0755	4.0620
10	3	2.9048	2.6440	2.6317
11	2	1.9304	1.7083	1.6980
12	0	1.2796	1.0994	1.0911
{ 13	0	0.8460	0.7048	0.6984
14	1	0.5579	0.4502	0.4454
{ 15	0	0.3670	0.2866	0.2831
16	0	0.2409	0.1818	0.1793
17	0	0.1577	0.1150	0.1132
18	0	0.1030	0.0725	0.0713
19	1	0.1911	0.1215	0.1188
χ^2 (p-value)		11.53 (0.24)	12.10 (0.21)	12.15 (0.21)
d.f.		9	9	9
Duration (Seconds)		6.01	1.75	1.86

ML estimates: $\hat{p}_2 = 2.27 \times 10^{-7}$, $\hat{p}_3 = 0.568428$, $\hat{k}_1 = 0.113725$, $\hat{k}_2 = 0$, $\hat{\lambda} = 1.668486$
MD_1 estimates: $\hat{p}_2 = 2.0 \times 10^{-10}$, $\hat{p}_3 = 0.535226$, $\hat{k}_1 = 0.099844$, $\hat{k}_2 = 0$, $\hat{\lambda} = 1.959887$
$MD_{1/2}$ estimates: $\hat{p}_2 = 4.0 \times 10^{-10}$, $\hat{p}_3 = 0.533580$, $\hat{k}_1 = 0.099078$, $\hat{k}_2 = 0$, $\hat{\lambda} = 1.975669$

All three methods are able to fit the data reasonably. Even though the χ^2 value for *MLE* is slightly lower than MD_1 and $MD_{1/2}$, the MD_1 and $MD_{1/2}$ methods have the clear advantage of faster computation time as shown in Table 4.8. Furthermore, the higher χ^2 value for MD_1 and $MD_{1/2}$ method is due to the biggest difference between the observed and expected frequencies for $x = 9$. Omitting this point would actually produce a lower χ^2 value for both MD_1 and $MD_{1/2}$ methods compared to *MLE*. This is in line with the characteristic of MD_1 and $MD_{1/2}$ method being a more robust estimation method and thus, mitigating the effect of possible sampling error.

4.5.2 Bivariate Case

Example 1. A comparison of the estimation methods MD_1, $MD_{1/2}$, *MPGHD* and *MLE* is made based on the chi-square, χ^2 goodness-of-fit statistic (2.13) obtained for a real data set. The data on the numbers of accidents sustained by 166 London omnibus drivers over two successive years from Edwards & Gurland (1961) is used. The *BNB* with pgf (4.1) and pmf (4.2) is fitted to the data set. The parameter estimates with corresponding χ^2 values as well as d.f. are given in the following Table 4.9.

Table 4.9

Parameter Estimates and χ^2 Statistics for MD_1, $MD_{1/2}$, *MPGHD* and *MLE* Methods

Parameter	Estimate			
	MD_1	$MD_{1/2}$	*MPGHD*	*MLE*
p_1	0.119763	0.109187	0.093077	0.232907
p_2	0.112254	0.102381	0.087041	0.224823
θ_3	0.034128	0.031740	0.023562	0.030518
ν	11.807725	13.037853	14.819167	5.317321
χ^2	14.72	14.75	15.35	17.40
d.f.	21	21	21	21
p-value	0.84	0.84	0.80	0.69
Duration (Seconds)	0.34	0.37	1.69	1.42

Based on the χ^2 values, MD_1, $MD_{1/2}$ and *MPGHD* methods give a better fit than *MLE*. Between the three, MD_1 and $MD_{1/2}$ show a slightly better fit than *MPGHD*. It is also noted that there is a marked difference between the estimates of MD_1, $MD_{1/2}$ and *MPGHD* as compared to the *ML* estimates, suggesting the possibility of outliers in the data set. These *ML* estimates are quite similar to those obtained by Subrahmaniam & Subrahmaniam (1973), where the index parameter, ν has been fixed in that paper. Subrahmaniam and Subrahmaniam obtained the estimates of $\hat{p}_1 = 0.244142$, $\hat{p}_2 = 0.235474$ and $\hat{\theta}_3 = 0.032278$ with $\nu = 5$. The computation times given in seconds clearly show the superiority of MD_1 and $MD_{1/2}$ methods for rapid estimation. The observed and expected frequencies for all four methods are given in Table 4.10.

Example 2. To illustrate further the advantage of the MD_α estimation method, the classic data set on 122 shunters from Arbous & Kerrich (1951), given in Table 4.11, has been fitted with the new *EBNB-II* distribution formed using the *BNB* of Mitchell & Paulson (1981). This new distribution has a complicated pmf (as given in Section 3.4.1) but, comparatively, a simple pgf (3.4). Parameter estimation using the classical methods, such as *MLE* and methods of moments, which made use of the pmf will be difficult and tedious. Estimation has been done using the proposed MD_1 and $MD_{1/2}$ methods, and the estimates for the parameters, following the notations in Section 3.1.1 e), are given in Table 4.12. The result for each of MD_1 and $MD_{1/2}$ methods is obtained using 5×5 quadrature points with a computation time of less than 1.5 seconds.

Table 4.10

Observed and Expected Frequencies for MD_1, $MD_{1/2}$, *MPGHD* and ***MLE*** Methods

	X_2									
X_1	0	1	2	3	4	5	6	7	8	9
	15	15	4	2	0	0	1	0	0	0
	15.73	12.90	5.74	1.83	0.47	0.10	0.02	0.00	0.00	0.00
0	15.65	12.94	5.76	1.83	0.47	0.10	0.02	0.00	0.00	0.00
	16.12	13.84	6.34	2.06	0.53	0.12	0.02	0.00	0.00	0.00
	16.34	**13.65**	**6.77**	**2.59**	**0.85**	**0.25**	**0.07**	**0.02**	**0.00**	**0.00**
	17	18	9	3	3	0	1	1	0	0
	14.37	19.12	11.77	4.80	1.50	0.39	0.09	0.02	0.00	0.00
1	14.39	19.28	11.87	4.82	1.49	0.38	0.08	0.02	0.00	0.00
	15.32	19.66	11.99	4.86	1.50	0.38	0.08	0.02	0.00	0.00
	14.41	**16.95**	**10.84**	**5.09**	**1.97**	**0.66**	**0.20**	**0.06**	**0.02**	**0.00**
	4	16	12	6	2	3	0	0	0	0
	7.11	13.10	10.91	5.71	2.19	0.67	0.17	0.04	0.01	0.00
2	7.12	13.20	11.01	5.73	2.18	0.66	0.17	0.04	0.01	0.00
	7.77	13.28	10.52	5.31	1.98	0.59	0.15	0.03	0.01	0.00
	7.55	**11.45**	**9.11**	**5.14**	**2.33**	**0.90**	**0.31**	**0.10**	**0.03**	**0.01**
	2	6	5	2	4	0	0	0	0	0
	2.53	5.95	6.35	4.19	1.97	0.72	0.22	0.06	0.01	0.00
3	2.52	5.96	6.37	4.18	1.95	0.70	0.21	0.05	0.01	0.00
	2.79	5.96	5.88	3.64	1.62	0.57	0.16	0.04	0.01	0.00
	3.05	**5.67**	**5.43**	**3.61**	**1.89**	**0.84**	**0.32**	**0.11**	**0.04**	**0.01**
	1	4	4	0	1	0	0	0	0	0
	0.72	2.07	2.72	2.19	1.24	0.54	0.19	0.06	0.01	0.00
4	0.71	2.05	2.69	2.16	1.22	0.52	0.18	0.05	0.01	0.00
	0.80	2.04	2.42	1.80	0.95	0.39	0.13	0.04	0.01	0.00
	1.05	**2.31**	**2.59**	**2.00**	**1.19**	**0.60**	**0.26**	**0.10**	**0.04**	**0.01**
	1	0	0	0	1	0	0	0	0	0
	0.18	0.60	0.93	0.89	0.60	0.31	0.13	0.04	0.01	0.00
5	0.17	0.58	0.91	0.87	0.58	0.29	0.12	0.04	0.01	0.00
	0.19	0.58	0.80	0.69	0.43	0.20	0.08	0.02	0.01	0.00
	0.33	**0.83**	**1.06**	**0.93**	**0.63**	**0.35**	**0.17**	**0.07**	**0.03**	**0.01**
	0	0	1	0	0	0	0	0	0	0
	0.04	0.15	0.27	0.30	0.24	0.14	0.07	0.03	0.01	0.00
6	0.04	0.14	0.26	0.29	0.22	0.13	0.06	0.02	0.01	0.00
	0.04	0.14	0.22	0.22	0.16	0.09	0.04	0.01	0.00	0.00
	0.09	**0.27**	**0.39**	**0.38**	**0.29**	**0.18**	**0.09**	**0.04**	**0.02**	**0.01**
	0	0	0	0	0	0	0	0	0	0
	0.01	0.04	0.09	0.12	0.11	0.08	0.04	0.02	0.00	0.00
7+	0.01	0.04	0.07	0.10	0.10	0.07	0.04	0.01	0.00	0.00
	0.01	0.04	0.06	0.07	0.06	0.04	0.02	0.01	0.00	0.00
	0.03	**0.11**	**0.18**	**0.21**	**0.18**	**0.12**	**0.08**	**0.04**	**0.02**	**0.00**

Note: The dotted lines indicate grouping of the data for the χ^2 goodness-of-fit test to yield a minimum expected frequency of 2.

Table 4.11

Observed Frequencies of the Number of Accidents Sustained by 122 Experienced Shunters over Two Successive Periods of Time

Number of Accidents from (1943-1947), X_1	Number of Accidents from (1937-1942), X_2						
	0	1	2	3	4	5	6
0	21	18	8	2	1	0	0
1	13	14	10	1	4	1	0
2	4	5	4	2	1	0	1
3	2	1	3	2	0	1	0
4	0	0	1	1	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0

Table 4.12

Results from MD_1 and $MD_{1/2}$ Estimations in Comparison with Results from Arbous & Kerrich (1951) and Kocherlakota & Kocherlakota (1992, p. 153)

Estimation	MD_1	$MD_{1/2}$	Arbous & Kerrich (1951)	Kocherlakota & Kocherlakota (1992)
Parameter Estimates	$\hat{p}_1 = 0.211768$ $\hat{p}_2 = 0.261421$ $\hat{v} = 3.612849$ $\hat{\alpha}_1 = 5.00 \times 10^{-8}$ $\hat{\alpha}_2 = 5.07 \times 10^{-7}$ $\hat{a} = 1.04 \times 10^{-6}$ $\hat{b} = 0.00$ $\hat{c} = 0.00$	$\hat{p}_1 = 0.212035$ $\hat{p}_2 = 0.262266$ $\hat{v} = 3.602972$ $\hat{\alpha}_1 = 1.12 \times 10^{-5}$ $\hat{\alpha}_2 = 4.29 \times 10^{-5}$ $\hat{a} = 2.22 \times 10^{-7}$ $\hat{b} = 0.00$ $\hat{c} = 0.00$	$\hat{p}_1 = 0.217$ $\hat{p}_2 = 0.265$ $\hat{v} = 3.524$	$\hat{p}_1 = 0.222$ $\hat{p}_2 = 0.271$ $\hat{v} = 3.420$
Correlation	0.235288	0.235815	0.239744	0.325489

The estimates obtained for MD_1 and $MD_{1/2}$ methods are comparable to the estimates obtained by Arbous & Kerrich (1951) and Kocherlakota & Kocherlakota (1992). Also note that the MD_1 and $MD_{1/2}$ parameter estimates for α_1, α_2, b and c are very close to zero, indicating that the *BNB* of Edward & Gurland (1981), a special case of the *EBNB-II*, may be sufficient to describe this data set.

Example 3. In Mitchell & Paulson (1981), the pmf of the proposed *BNB* is restricted to the index parameter, ν being an integer despite the distribution admitting a full range of correlation $[-1, 1]$. They argued that there is no representation for which ν can take on any arbitrary real values since the pmf derived is not infinitely divisible. Thus, having to fix as an integer, $\nu = 1$ in the *MLE* although it is more possible for $\nu > 1$, they are not able to produce a better fit to the data on the number of flight aborts for 109 aircrafts given in Table 4.13. On the other hand, the MD_α method involves only the pgf, where ν can take any arbitrary real values. Now, the data on the number of flight aborts will be fitted with *EBNB-II* distribution to allow for flexibility in the marginals, and parameter estimation is done with MD_1 and $MD_{1/2}$ methods for comparison with the result from Mitchell & Paulson (1981).

Table 4.13

Observed Frequencies of the Number of Flight Aborts for 109 Aircrafts over Two Successive Periods of Time

Number of Aborts for First Six Months, X_1	Number of Aborts for Second Six Months, X_2				
	0	1	2	3	4
0	34	20	4	6	4
1	17	7	0	0	0
2	6	4	1	0	0
3	0	4	0	0	0
4	0	0	0	0	0
5	2	0	0	0	0

Univariates $X_1 \sim NB(0.95, 0.40)$ and $X_2 \sim NB(1.51, 0.32)$ (Mitchell & Paulson, 1981).

Table 4.14 gives the results for the estimations. The estimates for p_2 are approximately the same for all estimation methods but the estimates for p_1 are very different for the MD_1 and $MD_{1/2}$ estimates compared to the *ML* estimate from Mitchell & Paulson (1981). Unfortunately, χ^2 goodness-of-fit test could not be performed since there is no explicit representation for the pmf of *EBNB-II* when ν is not an integer.

However, comparing the estimated marginals with the result for the univariates of X_1 and X_2 given in Table 4.13, the estimated marginals from both MD_1 and $MD_{1/2}$ methods are much closer to the given univariates than those from Mitchell and Paulson (1981). For MD_1 method, the marginals are $X_1 \sim NB(1.1138, 0.3525)$ and $X_2 \sim NB(1.4718, 0.3332)$ while for $MD_{1/2}$ method, the marginals are $X_1 \sim NB(1.0644, 0.3573)$ and $X_2 \sim NB(1.4857, 0.3311)$. The marginals from Mitchell & Paulson (1981) are $X_1 \sim NB(1, 0.1655)$ and $X_2 \sim NB(1, 0.3299)$. The estimated correlation coefficients for all three methods are not much different from the sample correlation of -0.16.

Table 4.14

Results from MD_1 and $MD_{1/2}$ Estimations in Comparison with Result from Mitchell & Paulson (1981)

Estimation	MD_1	$MD_{1/2}$	Mitchell & Paulson (1981)
Parameter Estimates	$\hat{p}_1 = 0.352531$ $\hat{p}_2 = 0.333210$ $\hat{a} = 0.709167$ $\hat{b} = 1.928684$ $\hat{c} = 5.62 \times 10^{-7}$ $\hat{v} = 1.113808$ $\hat{\alpha}_1 = 2.00 \times 10^{-10}$ $\hat{\alpha}_2 = 0.358033$	$\hat{p}_1 = 0.357294$ $\hat{p}_2 = 0.331096$ $\hat{a} = 0.499473$ $\hat{b} = 8.735757$ $\hat{c} = 0.020573$ $\hat{v} = 0.952946$ $\hat{\alpha}_1 = 0.141434$ $\hat{\alpha}_2 = 0.532748$	$\hat{p}_1 = 0.1655$ $\hat{p}_2 = 0.3299$ $\hat{a} = 0.00$ $\hat{b} = 0.6820$ $\hat{c} = 0.3179$ $v = 1$
Correlation	-0.131282	-0.119507	-0.13

4.6 Extension to Multivariate Distributions

The proposed method of parameter estimation is readily extended to the case of three or more variables because the Gauss quadrature method is easily extended for the approximation of multiple integrations (see Haber, 1970, p. 488; Burden & Faires, 2005, p. 226). However, the accuracy of the higher-dimension integration is of concern and this would be affected by the numbers of quadrature points selected in the computation of the integral. The following Table 4.15 shows the approximate integral values for several selected number of quadrature points for each variable of integration when finding the integral of MD_1 and $MD_{1/2}$ for a 5-variate *NB* distribution with the pgf

$$G_{(X_1,X_2,X_3,X_4,X_5)}(z_1,z_2,z_3,z_4,z_5)$$

$$= \left[\frac{1-\theta}{1-\theta\{(1-\theta_1)\cdots(1-\theta_5)+\theta_1(1-\theta_2)\cdots(1-\theta_5)z_1+\cdots+\theta_1\cdots\theta_5 z_1\cdots z_5\}}\right]^{\nu}$$

and marginals $X_i \sim NB(p_i, \nu)$ where $p_i = \theta\theta_i/(1-\theta+\theta\theta_i)$, $i = 1,2,\ldots,5$. Two sets of parameters are selected for computation.

(i) Set 1: $\theta_1 = 0.2, \theta_2 = 0.5, \theta_3 = 0.3, \theta_4 = 0.6, \theta_5 = 0.5, \theta = 0.3, \nu = 4.0$

(ii) Set 2 : $\theta_1 = 0.2, \theta_2 = 0.5, \theta_3 = 0.6, \theta_4 = 0.6, \theta_5 = 0.1, \theta = 0.3, \nu = 0.8$

Table 4.15

Approximate Integration Values for Different Number of Quadrature Points

No. of Quadrature Points	Set 1		Set 2	
	MD_1	$MD_{1/2}$	MD_1	$MD_{1/2}$
2	0.001178838	0.000918151	6.891733290	3.144593212
3	0.001173603	0.000918121	6.892291443	3.144656158
4	0.001173891	0.000918174	6.892298716	3.144658084
5	0.001173892	0.000918175	6.892298756	3.144658086
10	0.001173892	0.000918175	6.892298756	3.144658086

From the results above, a minimum of 2 quadrature points for each variable of integration is enough to obtain an accuracy of four significant figures. To have a higher confidence in the accuracy of the integral approximation, 3 quadrature points can be used with a slightly longer computation time. The integral approximation converges with more quadrature points.

Next, the use of the MD_1 and $MD_{1/2}$ methods in estimating the parameters of a 5-variate *NB* distribution has been investigated with a simulated sample of size n = 500. The parameters for Set 1 above are used for generating the random sample. The result is given in Table 4.16.

Table 4.16

Estimates for A Simulated 5-Variate *NB* Sample of Size 500

No. of Quadrature Points	Parameters	Estimates	
		MD_1	$MD_{1/2}$
2	$\theta_1 = 0.2$	0.173043	0.170282
	$\theta_2 = 0.5$	0.493745	0.4929726
	$\theta_3 = 0.3$	0.272964	0.271419
	$\theta_4 = 0.6$	0.579709	0.579059
	$\theta_5 = 0.5$	0.498450	0.498641
	$\theta = 0.3$	0.247742	0.255043
	$\nu = 4.0$	5.221724	5.049222
	Duration (Minutes):	0.525783	1.077600
3	$\theta_1 = 0.2$	0.178044	0.172070
	$\theta_2 = 0.5$	0.501627	0.495670
	$\theta_3 = 0.3$	0.279266	0.273963
	$\theta_4 = 0.6$	0.588056	0.581884
	$\theta_5 = 0.5$	0.506287	0.500621
	$\theta = 0.3$	0.220468	0.246223
	$\nu = 4.0$	5.939513	5.250783
	Duration (Minutes):	8.820317	6.347917

The estimates are close to the true parameter values with the computation time under 10 minutes. Although computation time is processor and computer dependent, this indicates that the MD_α method can be used for quick parameter estimation in multivariate setting.

CHAPTER 5 : A CLASS OF DISTRIBUTIONS DEFINED BY DIFFERENCE OF TWO DISCRETE RANDOM VARIABLES

5.0 Introduction

In this chapter, we consider the difference of two discrete random variables when they are (a) independent and (b) jointly distributed. The difference of two discrete random variables (rv's) has been discussed by various researchers (Irwin, 1937; Skellam, 1946; Johnson, 1959; Consul, 1988; Karlis & Ntzoufras, 2003; Ong & Shimizu, 2003); most of these researchers considered independent rv's. Furthermore, application to the analysis of paired data involving counts has not been given due attention in statistical literature. One particular area where such an analysis occurs naturally is in the study of fluctuating asymmetry of organisms involving meristic (count) traits, where models based upon the difference of two correlated random variables are required. A zero inflated count model is proposed to test for fluctuating asymmetry and a simulation study on the power of the test is considered. Examples using real data sets are then given.

5.1 Two Independent Random Variables

Let two independent rv's, X_1 and X_2, be from the Panjer's family of distributions with probability generating function (pgf) and probability mass function (pmf), respectively,

$$G_{X_i}(z) = \left(\frac{q_i}{1 - p_i z}\right)^{\alpha_i}, \text{ and}$$

$$\Pr(X_i = r) = \frac{(\alpha_i)_r}{r!} p_i^r (1 - p_i)^{\alpha_i}, \qquad 0 < p_i < 1, \alpha_i > 0, r = 0,1,2, \dots, i = 1,2.$$

Let $p_i = \frac{\lambda_i \theta_i}{1+\lambda_i \theta_i}$ and $\alpha_i = \frac{1}{\theta_i}$ for $i = 1,2$. Then, the Poisson pgf is obtained when $\theta_i \to 0$. When $\theta_i < 0$ ($\theta_i > 0$), the binomial (negative binomial) pgf is obtained. Further, define the difference of X_1 and X_2 as $X = X_1 - X_2$. Then, the rv X has the difference distribution with the pmf

$$\Pr(X = k) = \begin{cases} p_1^k (1-p_1)^{\alpha_1} (1-p_2)^{\alpha_2} \dfrac{(\alpha_1)_k}{k!} \, {}_2F_1\ (\alpha_1 + k, \alpha_2; k+1; p_1 p_2), & k \geq 0 \\ p_2^{-k} (1-p_1)^{\alpha_1} (1-p_2)^{\alpha_2} \dfrac{(\alpha_2)_{-k}}{(-k)!} \, {}_2F_1\ (\alpha_2 - k, \alpha_1; -k+1; p_1 p_2), & k < 0 \end{cases} \tag{5.1}$$

where k is an integer and the Gauss hypergeometric function is defined by

$${}_2F_1\ (a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!}, \qquad |x| < 1 .$$

This pmf (5.1) is arrived at by making use of the fact that

$$\Pr(X_1 - X_2 = k) = \sum_{r=0}^{\infty} \Pr(X_1 = k + r) \Pr(X_2 = r), \qquad -\infty < k < \infty \tag{5.2}$$

where $\Pr(X_1 = r)$ and $\Pr(X_2 = r)$ are the pmf for the rv's X_1 and X_2 for $r = 0,1,2,\ldots$. This relation (5.2) combined with the probability recurrence formulae for the two rv's belonging to the Panjer's family, given by

$$\Pr(X_i = r + 1) = \left(\frac{(\alpha_i + r) p_i}{r + 1} \right) \Pr(X_i = r), \quad r = 0,1,2, \ldots, i = 1,2, \tag{5.3}$$

is useful in computing the pmf (5.1) due to its relative simplicity without involving the Gauss hypergeometric function.

The pgf of $X = X_1 - X_2$ is given by

$$G_X(z) = \left(\frac{q_1}{1 - p_1 z}\right)^{\alpha_1} \left(\frac{q_2}{1 - p_2 \frac{1}{z}}\right)^{\alpha_2}, \qquad q_i = 1 - p_i, i = 1,2. \tag{5.4}$$

Through suitable transformation and restrictions, several models of differences nested in this class of distribution can be obtained as shown in Table 5.1.

Table 5.1

Nested Models

Parameter	Distribution		
	X_1	X_2	$X = X_1 - X_2$
$\theta_1 > 0$ $\theta_2 > 0$	$NB\left(\frac{1}{\theta_1}, \frac{\lambda_1\theta_1}{1+\lambda_1\theta_1}\right)$	$NB\left(\frac{1}{\theta_2}, \frac{\lambda_2\theta_2}{1+\lambda_2\theta_2}\right)$	$NB\left(\frac{1}{\theta_1}, \frac{\lambda_1\theta_1}{1+\lambda_1\theta_1}\right) - NB\left(\frac{1}{\theta_2}, \frac{\lambda_2\theta_2}{1+\lambda_2\theta_2}\right)$
$\theta_1 \to 0$	$Poisson(\lambda_1)$	$NB\left(\frac{1}{\theta_2}, \frac{\lambda_2\theta_2}{1+\lambda_2\theta_2}\right)$	$Poisson(\lambda_1) - NB\left(\frac{1}{\theta_2}, \frac{\lambda_2\theta_2}{1+\lambda_2\theta_2}\right)$
$\theta_2 \to 0$	$NB\left(\frac{1}{\theta_1}, \frac{\lambda_1\theta_1}{1+\lambda_1\theta_1}\right)$	$Poisson(\lambda_2)$	$NB\left(\frac{1}{\theta_1}, \frac{\lambda_1\theta_1}{1+\lambda_1\theta_1}\right) - Poisson(\lambda_1)$
$\theta_1 \to 0$ $\theta_2 \to 0$	$Poisson(\lambda_1)$	$Poisson(\lambda_2)$	$Poisson(\lambda_1) - Poisson(\lambda_2)$
$P_i = -\frac{p_i}{q_i}$ $\alpha_i = -N$ $i = 1,2$	$Binomial(N, P_1)$	$Binomial(N, P_2)$	$Binomial(N, P_1) - Binomial(N, P_2)$

5.1.1 Distributional Properties

A recurrence formula for pmf (5.1) is

$$(k+1)(k+1-\alpha_2)\gamma \Pr(k+1)$$
$$= [(2k + \alpha_1 - \alpha_2)\gamma - k]\Pr(k) + (1-\gamma)p_1(\alpha_1 + k)(\alpha_1 + k - 1)\Pr(k-1)$$

where $\gamma = \frac{p_1 p_2}{p_1 p_2 - 1}$.

The r-th factorial cumulants, denoted by $\kappa_{(r)}$ and derived by differentiating the factorial cumulant generating function $\ln G(z+1)$, is given by

$$\kappa_{(r)} = (r-1)!\left\{\alpha_1\left(\frac{p_1}{q_1}\right)^r + (-1)^r\alpha_2\left[\left(\frac{1}{1-q_2}\right)^r - 1\right]\right\}.$$

The mean and variance are then, respectively,

$$\mu_1 = \kappa_{(1)} = \alpha_1\left(\frac{p_1}{q_1}\right) - \alpha_2\left(\frac{p_2}{q_2}\right) \text{ and } \mu_2 = \kappa_{(1)} + \kappa_{(2)} = \alpha_1\left(\frac{p_1}{q_1^2}\right) + \alpha_2\left(\frac{p_2}{q_2^2}\right).$$

Since $\kappa_{(2)} > 0$, the variance is always larger than the mean. It follows that this class of distribution exhibits over dispersion.

The indices of skewness and kurtosis are given by $\sqrt{\beta_1} = \sqrt{\mu_3^2/\mu_2^3}$ and $\beta_2 = \mu_4/\mu_2^2$, respectively. Thus,

$$\beta_1 = \frac{[\alpha_1 a(1+a)(1+2a) - \alpha_2 p_2(1+p_2)/(1-p_2)^3]^2}{[\alpha_1 a(1+a) + \alpha_2 p_2/(1-p_2)^2]^3}, \qquad a = \frac{p_1}{q_1},$$

$$\frac{\beta_2 - 3}{\beta_1}$$

$$= \frac{\left\{\alpha_1 a(1+a)(1+6a+6a^2) + \frac{\alpha_2 p_2[(1+p_2)^2 + 2p_2]}{(1-p_2)^4}\right\}\left[\alpha_1 a(1+a) + \frac{\alpha_2 p_2}{(1-p_2)^2}\right]}{\left[\alpha_1 a(1+a)(1+2a) - \frac{\alpha_2 p_2(1+p_2)}{(1-p_2)^3}\right]^2}$$

$$> 0.$$

Ong *et al*. (2008) have discussed results of skewness and kurtosis for several specific cases. For $\theta_2 \to 0$ (*NB-Poisson*),

$$\beta_1 = \frac{[\alpha_1 a(1+a)(1+2a) - \lambda_2]^2}{[\alpha_1 a(1+a) + \lambda_2]^3}, \qquad a = \frac{p_1}{q_1},$$

$$\beta_2 = 3 + \frac{\alpha_1 a(1+a)(1+6a+6a^2) + \lambda_2}{[\alpha_1 a(1+a) + \lambda_2]^2}, \text{ and}$$

$$\frac{\beta_2 - 3}{\beta_1} = \frac{[\alpha_1 a(1+a)(1+6a+6a^2) + \lambda_2][\alpha_1 a(1+a) + \lambda_2]}{[\alpha_1 a(1+a)(1+2a) - \lambda_2]^2}.$$

For the Poisson distribution, $\beta_2 - \beta_1 - 3 = 0$ while for the *NB-Poisson* distribution, $(\beta_2 - 3)/\beta_1 > 1$. In comparison, the Neyman Type A distribution, a well known contagious distribution, has this ratio $(\beta_2 - 3)/\beta_1$ close to 1 (Johnson *et al*., 2005, p. 406), which limits its utility.

5.1.2 Difference of Two Independent Random Variables $X = X_1 - X_2$ when $X_2 \sim Binomial\ (n, p)$

The following result allows the application of known results for convolutions to derive various quantities for the difference $X = X_1 - X_2$ involving the binomial rv.

Result 5.1 (Ong *et al*., 2008): The difference $X = X_1 - X_2$, where X_1 is any rv belonging to the family defined by (5.3) and $X_2 \sim Binomial\ (n, p)$, is the convolution of X_1 and $T \sim Binomial\ (n, 1-p)$ shifted n steps to the left. (Note: If $X_1 \sim Binomial\ (n, p)$, then consider $X = -(X_2 - X_1)$.)

To illustrate, the pmf for the differences *Binomial-Binomial* is derived based on the above result. Now, let $X_1 \sim Binomial(m, \theta)$. Then, $X = X_1 - X_2$ is the shifted convolution of $X_1 \sim Binomial(m, \theta)$ and $T \sim Binomial\ (n, 1-p)$, that is $X \sim shifted_{-n}(X_1 + T)$. The pmf for $X_1 + T$, the convolution of two binomials (see Ong, 1995) is given by

$$\Pr(k) = p^n \binom{m+n}{k} \theta^k (1-\theta)^{m-k}\ {}_2F_1\left(-n, -k; -m-n; 1 + \frac{(1-\theta)(1-p)}{\theta p}\right),$$

$$k = 0,1,2, \dots, m+n.$$

Then, X has pmf

$$\Pr(x) = p^n \binom{m+n}{x+n} \theta^{x+n} (1-\theta)^{m-n-x} \ {}_2F_1 \left(-n, -x-n; -m-n; 1 + \frac{(1-\theta)(1-p)}{\theta p}\right),$$

$$x = -n, -n+1, \dots, 0, 1, 2, \dots, m.$$

Another example is the derivation of the *Poisson-Binomial* pmf. Let $X_1 \sim Poisson(\lambda)$. Then, $X = X_1 - X_2$ has pgf $G(z) = z^{-n} e^{\lambda(z-1)} (p + (1-p)z)^n$, that is, $X \sim shifted_{-n}(X_1 + T)$. $X_1 + T$ has pmf given by

$$\Pr(k) = e^{-\lambda} (1-p)^k p^{n-k} L_k^{(n-k)} \left(\frac{-\lambda p}{1-p}\right), \qquad k = 0,1,2, \dots$$

with $L_k^{(\alpha)}(x) = \frac{(\alpha+1)_n}{n!} \ {}_1F_1 \ (-n; \alpha+1; x)$ being the Laguerre function. If we let $\lambda = \theta(1-p)$, then this becomes the pmf of the Charlier series distribution (Ong, 1988). Thus, $X = X_1 - X_2$ has pmf

$$\Pr(x) = e^{-\lambda} (1-p)^{x+n} p^{-x} L_{x+n}^{(-x)} \left(\frac{-\lambda p}{1-p}\right), \qquad x = -n, -n+1, \dots, 0,1,2, \dots \ .$$

5.1.3 Computation of Probabilities

In order to use recurrence formulae to facilitate the computation of pmf's which are in terms of special functions, initial values like $\Pr(0)$ and $\Pr(1)$ are required. However, $\Pr(0)$ and $\Pr(1)$ are infinite series (except for the *Binomial-Binomial*); for example, from (5.2), $\Pr(0) = \sum_{r=0}^{\infty} \Pr(X_1 = r) \Pr(X_2 = r)$ where $\Pr(X_i = r), i = 1,2$ have recurrences as in (5.3). To aid the computations, error bounds for these infinite series are derived. Here, the error bound for calculating $\Pr(0)$ is presented. The error bounds for $\Pr(1)$ and other probabilities are obtained similarly. Let

$$\Pr(0) = \sum_{r=0}^{\infty} \Pr(X_1 = r) \Pr(X_2 = r) = \sum_{r=0}^{t} \Pr_{X_1}(r) \Pr_{X_2}(r) + e_t$$

where $e_t = \sum_{r=t+1}^{\infty} \Pr_{X_1}(r)\Pr_{X_2}(r)$ is the error bound. $\Pr(0)$ is evaluated by summing until e_t is below a prescribed value (accuracy). The following result gives a bound for e_t.

Result 5.2: Suppose that $\Pr_{X_1}(r)$ and $\Pr_{X_2}(r)$ are unimodal distributions. The error bound e_t is given by

$$e_t = \sum_{r=t+1}^{\infty} \Pr_{X_1}(r)\Pr_{X_2}(r) < \frac{\Pr_{X_1}(r+1)\Pr_{X_2}(r+1)}{1 - T_1T_2}$$

where $t \geq \max(M_1, M_2)$, $M_i =$ mode of $\Pr_{X_i}(r)$ and $T_i = (\alpha_i + t + 1)p_i/(t+2)$, $i = 1,2$.

Proof:

Consider $\Pr_{X_i}(t+2) = [(\alpha_i + t + 1)p_i/(t+2)]\Pr_{X_i}(t+1) = T_i\Pr_{X_i}(t+1)$, $i = 1,2$. Suppose $t \geq \max(M_1, M_2)$ where unimodality of the distributions is assumed. Then, $\Pr_{X_i}(t+2)/\Pr_{X_i}(t+1) = T_i < 1$ and $T_1T_2 < 1$. It follows that

$$e_t < \Pr_{X_1}(r+1)\Pr_{X_2}(r+1)[1 + T_1T_2 + (T_1T_2)^2 + \cdots] = \frac{\Pr_{X_1}(r+1)\Pr_{X_2}(r+1)}{1 - T_1T_2}.$$

■

To illustrate the utility of the error bound, comparison is done for the computation of $\Pr(0) = \sum_{r=0}^{\infty} \Pr_{X_1}(r)\Pr_{X_2}(r)$ by summing this infinite series until (a) the term $\Pr_{X_1}(r)\Pr_{X_2}(r) \leq \varepsilon$, and (b) error bound is realized, that is, $e_t \leq \varepsilon$ where ε is a designated small number, say $\varepsilon = 10^{-10}$. Some computational results are shown in Table 5.2. The $\Pr(0)$ in the first column have been computed to an accuracy of 10^{-10}. For small values of α_i and p_i, there is no appreciable difference between stopping criteria (a) and (b). It is noted that when the parameters α_i and p_i are large, computation of the infinite series by stopping criterion (a) is not advisable due to premature termination because all the terms $\Pr_{X_1}(r)\Pr_{X_2}(r)$ are very small. Thus, in general, it is more advantageous to use the error bound in the computation of $\Pr(0)$ and other probabilities for a more accurate result.

Table 5.2

Comparison between Stopping Criteria (a) and (b)

$\varepsilon = 1.0 \times 10^{-10}$	Estimated $\Pr(0)$/ Number of terms	
Parameters/ $\Pr(0)$	Criterion (a) Term is very small	Criterion (b) Error bound achieved
$\alpha_1 = 5, p_1 = 0.1$ $\alpha_2 = 5, p_2 = 0.2$ 0.3097102247	0.3097102247 10	0.3097102246 9
$\alpha_1 = 5, p_1 = 0.1$ $\alpha_2 = 10, p_2 = 0.2$ 0.152199031879	0.1521990319 11	0.1521990318 10
$\alpha_1 = 5, p_1 = 0.1$ $\alpha_2 = 5, p_2 = 0.9$ 0.000037854543	0.0000378545 12	0.0000378545 37
$\alpha_1 = 5, p_1 = 0.1$ $\alpha_2 = 10, p_2 = 0.9$ 0.000000001380	5.90490×10^{-11} 1	0.0000000014 82
$\alpha_1 = 5, p_1 = 0.4$ $\alpha_2 = 5, p_2 = 0.5$ 0.104458093643	0.1044580936 24	0.1044580936 23
$\alpha_1 = 5, p_1 = 0.4$ $\alpha_2 = 10, p_2 = 0.5$ 0.034065499221	0.0340654992 27	0.0340654992 27
$\alpha_1 = 5, p_1 = 0.85$ $\alpha_2 = 5, p_2 = 0.95$ 0.003023311781	2.37305×10^{-11} 1	0.0030233117 164
$\alpha_1 = 5, p_1 = 0.85$ $\alpha_2 = 10, p_2 = 0.95$ 0.000031024588	7.45177×10^{-18} 1	0.0000310245 173

5.2 Two Correlated Random Variables

Considering the case when the two rv's are not independent, the difference of two correlated X_1 and X_2 rv's having extended bivariate negative binomial (*EBNB*) distribution arising from two different formulations is examined as follows.

a) *EBNB* Distribution Formed by Extended Trivariate Reduction

If two rv's X_1 and X_2 involve a third random element in common as in the usual trivariate reduction formulation (2.1), the difference of these two rv's can be treated as the difference between two independent rv's. To illustrate, let Y_1, Y_2 and W be three independent rv's. Defined as (2.1), the difference between X_1 and X_2 denoted by X is $X = Y_1 - Y_2$, which does not involve W. Therefore, X can be considered as the difference of two independent rv's.

For formulation (3.1), where (W_1, W_2) is a pair of randomly correlated elements independent of Y_1 and Y_2, it is found that $X = (Y_1 - Y_2) + (W_1 - W_2)$. Thus, X is a convolution of two independent differences. The first difference involves two independent variables whereas the second difference may be considered as a difference of two dependent variables depending on the formulation of the joint (W_1, W_2) distribution. If (W_1, W_2) is formulated by the usual trivariate reduction (2.1) or *EBNB* described in the next section, we obtain the difference of two independent rv's.

b) *EBNB* Distribution as A Bivariate Mixed Poisson Distribution

The *EBNB* distribution has pgf as given in (3.7). Thus, pgf of the difference $X = X_1 - X_2$ will have the form of

$$G_X(z) = G_{(X_1,X_2)}\left(z, \frac{1}{z}\right)$$

$$= \left(\frac{\Theta}{1 - \theta_2 - (\theta_1 + \theta_3)z}\right)^{\alpha_1 - \nu} \left(\frac{\Theta}{1 - \theta_1 - (\theta_2 + \theta_3)\frac{1}{z}}\right)^{\alpha_2 - \nu} \left(\frac{\Theta}{1 - \theta_1 z - \theta_2 \frac{1}{z} - \theta_3}\right)^{\nu}. \tag{5.5}$$

Let $\theta_1 = \pi$, $\theta_2 = \eta$ and $\theta_3 = -\pi\eta$ where $\theta_3 < 0$. This still represents a legitimate *EBNB* (see Section 3.1.2). Equation (5.5) is then

$$G_X(z) = \left(\frac{1-\pi}{1-\pi z}\right)^{\alpha_1} \left(\frac{1-\eta}{1-\eta\frac{1}{z}}\right)^{\alpha_2}$$

which is of the same form as (5.4), the pgf of the difference of two independent $NB(\alpha_1, \pi)$ and $NB(\alpha_2, \eta)$ rv's. Thus, the differences of correlated data from *EBNB* distribution can be fitted using the *NB-NB* distribution.

As a special case, when $\alpha_1 = \alpha_2 = \nu$, (X_1, X_2) has the *BNB* (compound correlated bivariate Poisson) distribution of Edwards & Gurland (1961). The difference $X = X_1 - X_2$ is then the difference of two independent $NB(\nu, \pi)$ and $NB(\nu, \eta)$ rv's.

5.3 Applications of the Distribution

5.3.1 Model Selection

The *NB-NB* distribution has several other distributions nested within it through certain conditions imposed on the distribution parameters as shown in Table 5.1. By testing suitable hypotheses involving only these parameters, the *NB-NB* distribution can be used to find the best model for fitting a set of data. For example, let the null and alternative hypotheses of interest be, respectively,

H_0 : $\theta_1 = \theta_2 = \theta_0$; λ_1, λ_2 unspecified

H_1 : $\lambda_1, \lambda_2, \theta_1, \theta_2$ unspecified

where $\lambda_1, \lambda_2, \theta_1, \theta_2$ are as defined in Section 5.1 and θ_0 is known. Non-rejection of the null hypothesis for a very small value of θ_0 indicates that the *Poisson-Poisson* distribution will fit adequately. Conversely, the rejection of the null hypothesis will indicate that the more general *NB-NB* distribution could be more suitable.

5.3.2 Test for Equality of Means and Index Parameters

The null hypothesis for testing the equality of means and index parameters of the random variables X_1 and X_2 is

$$H_0 : \quad \mu = 0, \delta = 0; \;\; \lambda_2, \theta_2 \text{ unspecified}$$

against the alternative hypothesis of

$$H_1 : \quad \mu, \lambda_2, \delta, \theta_2 \text{ unspecified}$$

where $\mu = \lambda_1 - \lambda_2$ and $\delta = \theta_1 - \theta_2$. If the null hypothesis is not rejected, the distributions for X_1 and X_2 are identical. The converse is true if the null hypothesis is rejected.

5.3.3 Test for Bilateral Asymmetry in Organisms

There are three forms of bilateral asymmetry, namely fluctuating asymmetry (FA), directional asymmetry (DA) and antisymmetry (Van Valen, 1962). FA, a pattern of bilateral variation in a sample of individuals where the mean of the difference between the bilateral sides is zero with deviations normally distributed about the mean (Palmer, 1994), is a commonly used measure for developmental stability. Developmental stability refers to the ability of defending against small, random developmental perturbations originating from the environment on a particular morphological structure or trait.

On the other hand, DA shows a deviation from perfect bilateral symmetry by favouring development of a trait on one side of the body than the other, causing the mean of that side to be larger. Antisymmetry is a pattern of bilateral variation in a sample of individuals which shows a platykurtic (broad-peaked) or bimodal distribution of the differences between the bilateral sides about a mean of zero. DA and antisymmetry may render some traits unusable for studies of developmental stability (Palmer & Strobeck, 1986; Palmer, 1994).

Homogenous Population

The morphological trait considered may be meristic (counts) or metrical (continuous variable). Here, it is of interest to consider an additive error model for meristic trait. This is the case if inert structures like bristles or fins are considered (Mosimann & Campbell, 1988). Let W_1 and W_2 be rv's that represent the development of the left (L) and right (R) side of a meristic (count) trait for a homogenous group of individuals (Palmer, 1994) such that $W_1 = W_2 = W$. If Y_1 and Y_2 are the corresponding *independent* random errors in development for the left and right side of the trait, then by trivariate reduction (2.1)

$$L = W + Y_1 \text{ and } R = W + Y_2 \,.$$

Thus, (L, R) has a joint distribution with marginals given by L and R. The difference $X = L - R = Y_1 - Y_2$ is the difference of two independent discrete rv's as described in Section 5.1.

For example, let $Poisson(\mu)$ represent a Poisson random variable with mean μ. Suppose W, Y_1 and Y_2 are independent Poisson random variables $Poisson(\lambda)$, $Poisson(\lambda_1)$ and $Poisson(\lambda_2)$ respectively. Then, (L, R) has a bivariate Poisson distribution with joint probability mass function (Mardia, 1970) given by

$$\Pr(L = l, R = r) = e^{-(\lambda+\lambda_1+\lambda_2)} \frac{{\lambda_1}^l}{l!} \frac{{\lambda_2}^r}{r!} \sum_{i=0}^{\min(l,r)} \binom{l}{i} \binom{r}{i} i! \left(\frac{\lambda}{\lambda_1 \lambda_2}\right)^i .$$

Then, $X = Y_1 - Y_2$ has pmf $\Pr(X = k) = f(k)$ (see, for example, Johnson *et al.*, 2005, p. 198) given by

$$f(k) = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{k}{2}} I_k\left(2\sqrt{\lambda_1\lambda_2}\right), \qquad -\infty < k < \infty \tag{5.6}$$

where k is an integer and $I_k(x)$ is the modified Bessel function of the first kind defined by

$$I_k(x) = \left(\frac{x}{2}\right)^k \sum_{n=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^n}{n!\,\Gamma(k+n+1)}$$

with $\Gamma(y)$ being the gamma function.

For simplicity suppose that the random errors in development, Y_1 and Y_2 are also identically distributed (Graham *et al*., 2003, p. 58), that is $\lambda_1 = \lambda_2 = \mu$. Then, $X = Y_1 - Y_2$ is the difference of two identical Poisson rv's (Irwin, 1937; see also Johnson *et al*., 2005, p. 198 for references therein). In this case,

$E[X] = E[Y_1] - E[Y_2] = 0$ and

$Var[X] = Var[Y_1] + Var[Y_2] = 2\mu$.

Also, $f(-k) = f(k)$ since the distribution X is symmetrical.

Heterogeneous Population

Relaxing the assumption that the group of individuals under study forms a homogenous group, it is assumed that Y_1 and Y_2 above are now independent Poisson random variables $Poisson(\lambda_{1i})$ and $Poisson(\lambda_{2i})$ for the *i*th individual. That is, the means λ_{1i} and λ_{2i} may differ from individual to individual due to size, length and so on, representing the non homogeneity in the population. However, this constitutes a great number of unknown parameters λ_{1i} and λ_{2i}, $i = 1,2, \ldots$ in the model. To overcome this, compounding technique is applied by assuming that λ_{1i} and λ_{2i} are values of rv's with probability density functions (pdf's) $f(\lambda_1; \psi)$ and $g(\lambda_2; \phi)$ respectively, where ψ and ϕ

are vectors of unknown parameters. Both $f(\lambda_1;\psi)$ and $g(\lambda_2;\phi)$ are known as mixing distributions. This reduces a large number of parameters to be considered to only ψ and ϕ in the two distributions.

Mathematically, let $Y_1|\lambda_1 \sim Poisson(\lambda_1)$ and $Y_2|\lambda_2 \sim Poisson(\lambda_2)$ having pmf's $f_{Y_1}(k|\lambda_1)$ and $f_{Y_2}(k|\lambda_2)$ respectively. Then, the unconditional pmf's are given by

$$f_{Y_1}(k) = \int_0^\infty e^{-\lambda_1} \frac{{\lambda_1}^k}{k!} f(\lambda_1;\psi)d\lambda_1 \text{ and } f_{Y_2}(k) = \int_0^\infty e^{-\lambda_2} \frac{{\lambda_2}^k}{k!} g(\lambda_2;\phi)d\lambda_2 \, .$$

If $f(\lambda_1;\psi)$ and $g(\lambda_2;\phi)$ are taken to be gamma pdf's, the negative binomial (*NB*) distributions $Y_1 \sim NB(\alpha_1,\pi)$ and $Y_2 \sim NB(\alpha_2,\eta)$ are obtained (Greenwood & Yule, 1920). Then, the pmf of $X = Y_1 - Y_2$ is given by (5.1).

Correlated Random Errors

Extending to Y_1 and Y_2 being the corresponding *correlated* random errors in development for the left and right side of the morphological trait studied and letting the joint (Y_1, Y_2) has the *EBNB* distribution in Section 3.4, the result in Section 5.2 is applicable for the difference $X = Y_1 - Y_2$ here.

a) Directional Asymmetry

If $\lambda_1 \neq \lambda_2$, then

$E[X] = E[Y_1] - E[Y_2] = \lambda_1 - \lambda_2 \neq 0$ and

$Var[X] = Var[Y_1] + Var[Y_2] = \lambda_1 + \lambda_2$.

This case may be used as a model for DA. It follows that the set of hypotheses of interest is

$$H_0 : \lambda_1 = \lambda_2 \quad \text{against} \quad H_1 : \lambda_1 \neq \lambda_2 \, . \tag{5.7}$$

Non rejection of H_0 indicates that the means of the left and right sides may be taken to be the same and consideration of FA is more appropriate. There are a variety of statistical tests for the hypotheses (5.7) which are similar to those for Section 5.3.2. The popular ones are the likelihood ratio, score and Wald tests. Rao's score test is usually preferred due to its computation simplicity.

b) Fluctuating Asymmetry

The presence of excess zeros (inflation) in count data indicates a zero-inflated model having the pmf

$$P(0) = \omega + (1 - \omega)f(0), \qquad 0 < \omega < 1$$

$$P(k) = (1 - \omega)f(k), \qquad k = 1,2,3,\ldots \tag{5.8}$$

where $f(k)$ is the pmf of a rv X. It is possible for $\omega < 0$ provided that $\omega \geq \frac{-f(0)}{1-f(0)}$. The literature abounds with the applications of the well-researched zero-inflated models. See, for instance, the recent articles of Gupta *et al*. (1996, 2004) and papers in the reference.

If the difference in the development of the left and right side of a meristic trait is merely due to random variation, then a high degree of symmetry will be reflected in a great number of $X = 0$. It follows that an appropriate model will be given by the zero-inflated model (5.8) where $f(k)$ is as in (5.1) or (5.6) depending on the model. A threshold value ω_0 is predetermined and if the value ω exceeds ω_0 it will mean that there is a significant degree of symmetry.

With zero-inflated model (5.8), a statistical test of FA corresponds to testing the one-sided hypotheses

$$H_0: \ \omega = \omega_0 \quad \text{against} \quad H_1: \ \omega \geq \omega_0.$$

Therefore, rejection of H_0: $\omega = \omega_0$ implies perfect bilateral symmetry for the trait apart from insignificant chance variation.

5.4 Numerical Illustrations

The required formulae for the partial derivatives of *NB-NB* and *Poisson-Poisson* distributions and related quantities in the information matrix are given in Appendices B and C, respectively.

5.4.1 Model Selection and Test for Equality of Means

Thirty two randomly selected participants from a prospective study of male sexual contacts of men with AIDS or an AIDS-related condition (ARC) were assessed for the presence or absence of generalized lymphadenopathy by two different physicians with mean number of assessed palpable lymph nodes being 7.91 and 5.16 (Rosner, 2000, p. 319). Since the means of the number of assessed lymph nodes are clearly different, the $NB(\theta, \lambda) - NB(\theta, \lambda)$ distribution is not considered and $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$ distribution has been fitted to the data. The parameterization using p_1, p_2, α_1 and α_2 has not been adopted because of the numerical instability in the computations for the score test. Due to the complicated expression of the pmf, maximum likelihood estimation (*MLE*) has been done by numerical optimization.

The expected frequencies and their graphical representation are shown in Table 5.3 and Figure 5.1, respectively. The hypothesis that the data follows the $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$ distribution is not rejected at the significance level, $\alpha = 0.05$ since $\chi_7^2 = 14.067 > 11.73$. Note that the χ^2 goodness of fit value is inflated due to the relatively high count at $k = 3$.

Table 5.3

Observed and Expected Frequencies of Differences in Assessment of Number of Palpable Lymph Nodes among Sexual Contacts of AIDS or ARC Patients by Two Physicians

Difference, k	Observed Frequency	Expected Frequency, $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$
-4 ≤	0	0.35
-3	1	0.50
-2	2	1.03
-1	1	1.85
0	3	2.89
1	3	3.90
2	1	4.55
3	10	4.60
4	4	4.07
5	3	3.18
6	2	2.20
7≥	2	2.86
χ^2		11.73
d.f.		7

ML estimates: $\hat{\lambda}_1 = 5.1876$, $\hat{\lambda}_2 = 2.4376$, $\hat{\theta}_1 = 1.44 \times 10^{-4}$, $\hat{\theta}_2 = 0.02591$

Since the *ML* estimates of θ_i's are small, which indicates the *Poisson-Poisson* distribution, the data have been fitted with the $Poisson(\lambda_1) - Poisson(\lambda_2)$ distribution. The parameter estimates obtained are $(\hat{\lambda}_1, \hat{\lambda}_2) = (5.2707, 2.5207)$ and $\chi^2 = 11.80$. The expected frequencies are very close to those given in Table 5.3 and will not be displayed.

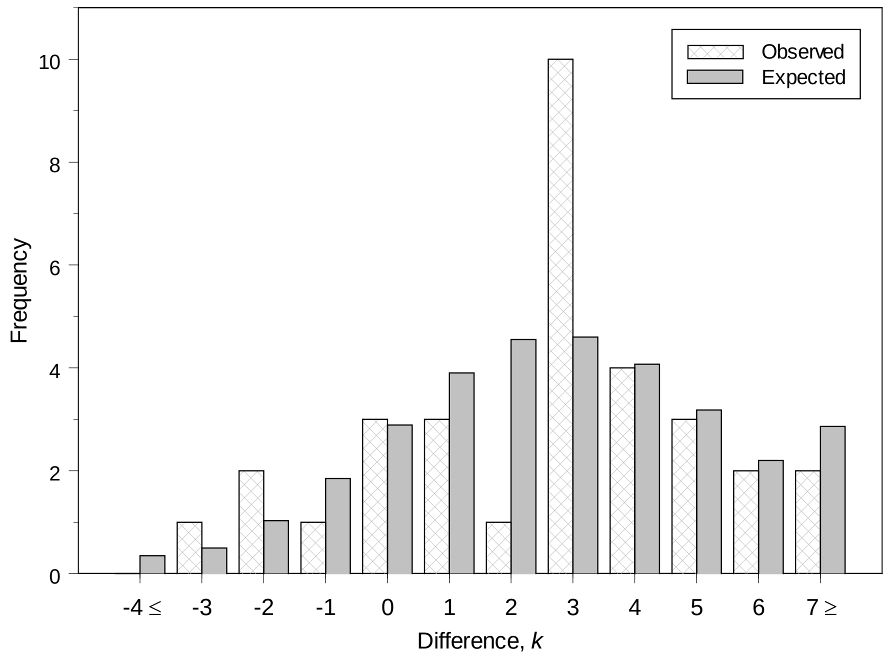


Figure 5.1

Observed and Expected Frequencies of Difference in Assessment of Number of Palpable Lymph Nodes by Two Physicians

Since $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$ $(Poisson(\lambda_1) - Poisson(\lambda_2))$ distribution is not rejected, we now consider the test of parameters based upon the likelihood ratio and score tests. The $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$ will be used to illustrate the tests.

The degree of clinical agreement between the two physicians may be examined by testing for equality of means in the distributions of the number of palpable lymph nodes assessed by them when a $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$ distribution is fitted to the data. It is assumed that the assessments made by the physicians are independent of each other. The hypotheses of interest are as in Section 5.3.2.

Remark: The sets of parameters under the null and alternative hypotheses correspond to the $NB(\theta, \lambda) - NB(\theta, \lambda)$ and $NB(\theta_1, \lambda_1) - NB(\theta_2, \lambda_2)$ distributions respectively with parameter estimates

$\mu = 0, \delta = 0; \hat{\lambda}_2^0 = 7.6968, \hat{\theta}_2^0 = 1.4041 \times 10^{-6}$ (null),

$\hat{\mu} = 2.7500, \hat{\lambda}_2 = 2.4376, \hat{\delta} = -0.02577, \hat{\theta}_2 = 0.02591$ (alternative).

The likelihood ratio (*LR*) and score test statistic values for the hypothesis test are 10.969 and 18.373, respectively. These test statistics as given in Section 2.5.2 have been obtained by numerical computation of the various quantities involved. For the two-sided score test, following the sequence of μ, λ_2, δ and θ_2, the efficient scores and information matrix for a single observation are computed as

$$\mathbf{U}^{*T} = [1.010563 \quad -4.103907 \times 10^{-8} \quad -1.926742 \quad 0.229783] \text{ and}$$

$$\mathbf{\Gamma}^* = \begin{bmatrix} 0.067010 & 0.004098 & 0.120730 & 0.241461 \\ & 0.008196 & 0.241461 & 0.482921 \\ & & 8.513046 & 14.386421 \\ & & & 28.772841 \end{bmatrix}.$$

At the significance level $\alpha = 0.05$, both test statistics above exceeded the critical value $\chi_2^2 = 5.991$. Therefore, both tests reject the null hypothesis at $\alpha = 0.05$. Thus, there is little evidence to suggest that the mean of number of palpable lymph nodes as assessed by each of the two physicians are equal.

5.4.2 Test for Fluctuating Asymmetry

For illustration purposes, the simpler *Poisson-Poisson* distribution, which is nested in *NB-NB* distribution as shown in Section 5.1, is considered. The pmf of this distribution is as given in (5.6). For the zero-inflated model (5.8), it is of interest to apply the above theory to test the following null and alternative hypotheses.

$H_0: \quad \omega = \omega_0; \quad \lambda_1, \lambda_2$ unspecified

$H_1: \quad \omega \geq \omega_0; \quad \lambda_1, \lambda_2$ unspecified. (5.9)

Since the alternative hypothesis is one-sided, the one-sided score test (Silvapulle & Silvapulle, 1995) mentioned in Section 2.5 is employed here. Note that the usual score test, which is a two-sided test, is inappropriate or not meaningful since it is of interest only to show ω greater than a threshold value of ω_0 as no evidence of FA. As a comparison, the one-sided likelihood ratio test is also considered. Asymptotic null distribution of these one-sided test statistics is $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ distribution with χ_0^2 denoting the distribution with point mass at zero (Self & Liang, 1987; Silvapulle & Silvapulle, 1995). The *p*-value of both test statistics is as given in (2.16). First, the computer programs used in implementing the tests were validated using 5 sets of generated data from several combinations of λ_1, λ_2 and ω for $\omega_0 = 0.5$ and 0.7 as shown in Table 5.4. Assuming a significance level of $\alpha = 0.05$ for the hypothesis testing, the results of the tests of hypotheses for these well-behaved data sets are consistent with the parameters chosen.

Next, a simulation study is conducted to compare the power of both one-sided score and likelihood ratio tests. The values $\omega_0 = 0.3, 0.5$ and 0.7, and various combinations of λ_1, λ_2 and ω are selected for two sample sizes, $n = 50$ and 100. We simulated 10000 Monte Carlo samples for each combination of parameter values. For each simulated sample, the *p*-value is calculated. The proportion of samples with *p*-value less than α is computed as an estimate for the power of the test with the results as shown in Table 5.5 ($\alpha = 0.05$) and Table 5.6 ($\alpha = 0.10$).

Table 5.4

Hypothesis Testing for 5 Sets of Generated Data

Parameters	*ML* Estimate		
	$\omega_0 = 0.5$	$\omega_0 = 0.7$	Unconstrained
Set 1			
$\lambda_1 = 0.5$	0.114391	0.205082	0.215989
$\lambda_2 = 0.5$	0.114391	0.205082	0.215989
$\omega = 0.8$			0.713131
Score, S_{c_1} (p-value)	0.031846 (0.43)	0.000309 (0.49)	
LR_T (p-value)	0.727826 (0.20)	0.006273 (0.47)	
	Do not reject H_0	Do not reject H_0	
Set 2			
$\lambda_1 = 2.0$	1.888086	2.069489	1.764248
$\lambda_2 = 0.8$	0.616240	0.665064	0.575384
$\omega = 0.4$			0.407281
Score, S_{c_1} (p-value)	0.00 (1.00)	0.00 (1.00)	
LR_T (p-value)	0.00 (1.00)	0.00 (1.00)	
	Do not reject H_0	Do not reject H_0	
Set 3			
$\lambda_1 = 4.2$	1.478640	2.517537	2.991034
$\lambda_2 = 4.2$	1.478639	2.517536	2.991033
$\omega = 0.8$			0.815328
Score, S_{c_1} (p-value)	29.785568 (2.4×10^{-8})	3.410228 (0.03)	
LR_T (p-value)	56.184720 (3.3×10^{-14})	10.199098 (7.0×10^{-4})	
	Reject H_0	Reject H_0	
Set 4			
$\lambda_1 = 4.2$	3.078538	3.586511	3.404922
$\lambda_2 = 4.2$	3.078531	3.586517	3.404937
$\omega = 0.6$			0.619395
Score, S_{c_1} (p-value)	5.767871 (8.2×10^{-3})	0.00 (1.00)	
LR_T (p-value)	7.944254 (2.4×10^{-3})	0.00 (1.00)	
	Reject H_0	Do not reject H_0	
Set 5			
$\lambda_1 = 4.2$	4.139961	4.228670	3.615909
$\lambda_2 = 4.2$	4.139961	4.228670	3.615909
$\omega = 0.1$			0.099489
Score, S_{c_1} (p-value)	0.00 (1.00)	0.00 (1.00)	
LR_T (p-value)	0.00 (1.00)	0.00 (1.00)	
	Do not reject H_0	Do not reject H_0	

Table 5.5

a) Power of One-Sided Score Test ($\alpha = 0.05$)

ω_0	ω	$n = 50$ (λ_1, λ_2)			$n = 100$ (λ_1, λ_2)		
		(2, 2)	(4, 2)	(4, 4)	(2, 2)	(4, 2)	(4, 4)
0.3	0.3	0.0412	0.0510	0.0573	0.0417	0.0580	0.0596
	0.4	0.2142	0.3359	0.3310	0.3925	0.5569	0.5470
	0.5	0.5530	0.7768	0.7691	0.8631	0.9693	0.9561
	0.6	0.7920	0.9746	0.9699	0.9854	0.9999	0.9997
0.5	0.5	0.0129	0.0177	0.0188	0.0158	0.0220	0.0213
	0.6	0.0981	0.1918	0.1741	0.2484	0.4117	0.3709
	0.7	0.3307	0.6565	0.6026	0.7553	0.9480	0.9303
	0.8	0.4495	0.8708	0.8492	0.8136	0.9962	0.9930
0.7	0.7	0.0003	0.0009	0.0004	0.0006	0.0019	0.0015
	0.8	0.0055	0.0334	0.0163	0.0460	0.1794	0.1240
	0.9	0.0490	0.2293	0.1870	0.2536	0.7452	0.7041

b) Power of One-Sided Likelihood Ratio Test ($\alpha = 0.05$)

ω_0	ω	$n = 50$ (λ_1, λ_2)			$n = 100$ (λ_1, λ_2)		
		(2, 2)	(4, 2)	(4, 4)	(2, 2)	(4, 2)	(4, 4)
0.3	0.3	0.0394	0.0332	0.0409	0.0398	0.0377	0.0412
	0.4	0.2238	0.2734	0.2775	0.4019	0.4813	0.4842
	0.5	0.6030	0.7285	0.7255	0.8848	0.9539	0.9427
	0.6	0.8917	0.9681	0.9636	0.9957	0.9999	0.9995
0.5	0.5	0.0327	0.0329	0.0344	0.0379	0.0381	0.0401
	0.6	0.2114	0.2805	0.2580	0.4036	0.5083	0.4746
	0.7	0.6049	0.7778	0.7322	0.9144	0.9718	0.9623
	0.8	0.8528	0.9800	0.9727	0.9932	1.0000	0.9998
0.7	0.7	0.0234	0.0385	0.0275	0.0311	0.0432	0.0327
	0.8	0.1865	0.3684	0.2701	0.4569	0.6424	0.5593
	0.9	0.5843	0.8472	0.8050	0.9213	0.9935	0.9895

Table 5.6

a) Power of One-Sided Score Test ($\alpha = 0.10$)

ω_0	ω	$n = 50\ (\lambda_1, \lambda_2)$			$n = 100\ (\lambda_1, \lambda_2)$		
		(2, 2)	(4, 2)	(4, 4)	(2, 2)	(4, 2)	(4, 4)
0.3	0.3	0.0812	0.0935	0.1031	0.0855	0.1071	0.1084
	0.4	0.3404	0.4607	0.4619	0.5381	0.6758	0.6759
	0.5	0.7016	0.8633	0.8586	0.9292	0.9858	0.9784
	0.6	0.8953	0.9906	0.9868	0.9960	1.0000	0.9999
0.5	0.5	0.0368	0.0480	0.0507	0.0463	0.0562	0.0590
	0.6	0.2122	0.4190	0.3113	0.4175	0.5809	0.5431
	0.7	0.5423	0.8119	0.7677	0.8920	0.9827	0.9726
	0.8	0.6264	0.9389	0.9332	0.9166	0.9987	0.9970
0.7	0.7	0.0033	0.0090	0.0056	0.0060	0.0127	0.0085
	0.8	0.0378	0.1519	0.0873	0.1822	0.4200	0.3267
	0.9	0.1655	0.4844	0.4259	0.4538	0.8799	0.8531

b) Power of One-Sided Likelihood Ratio Test ($\alpha = 0.10$)

ω_0	ω	$n = 50\ (\lambda_1, \lambda_2)$			$n = 100\ (\lambda_1, \lambda_2)$		
		(2, 2)	(4, 2)	(4, 4)	(2, 2)	(4, 2)	(4, 4)
0.3	0.3	0.0813	0.0708	0.0821	0.0838	0.0794	0.0865
	0.4	0.3497	0.4020	0.4117	0.5430	0.6224	0.6286
	0.5	0.7384	0.8353	0.8330	0.9383	0.9796	0.9725
	0.6	0.9497	0.9876	0.9853	0.9986	1.0000	0.9999
0.5	0.5	0.0710	0.0703	0.0759	0.0812	0.0792	0.0838
	0.6	0.3338	0.3416	0.3941	0.5499	0.6460	0.6249
	0.7	0.7446	0.8756	0.8427	0.9612	0.9898	0.9856
	0.8	0.9203	0.9913	0.9878	0.9969	1.0000	0.9999
0.7	0.7	0.0532	0.0794	0.0639	0.0667	0.0899	0.0699
	0.8	0.3113	0.5205	0.4209	0.6148	0.7754	0.7093
	0.9	0.7218	0.9144	0.8909	0.9558	0.9960	0.9951

The study reveals the following:

(i) In most cases, the empirical power for score test is smaller than the power of likelihood ratio test.

(ii) The probability of Type-I error for score test is much lower than the nominal level when the value of ω_0 is larger. Thus, the likelihood ratio test would be preferred for such cases.

(iii) The power for both tests increases when the difference between ω_0 and ω increases.

(iv) The empirical powers for both tests also increase as the sample size increases.

A numerical example based on a real data set is given next.

Example 1. A data set from Mardia (1970) on the distribution of number of Mullerian glands on the left and right forelegs of 2000 male pigs will be used to show application of the above theory. The frequencies for the difference in the number of glands on the bilateral sides $(R - L)$ of the pigs are given in Table 5.7 below.

Table 5.7

Observed Frequencies for the Difference of Number of Mullerian Glands on the Left and Right Sides $(R - L)$ of 2000 Male Pigs

Difference, k	Frequency
-6	0
-5	0
-4	4
-3	28
-2	116
-1	444
0	809
1	450
2	111
3	34
4	4
5	0
6	0

The zero-inflated *Poisson-Poisson* difference distribution is fitted to the data to yield the estimates of $\hat{\lambda}_1 = 0.627606$, $\hat{\lambda}_2 = 0.620612$ and $\widehat{\omega} = 1.39 \times 10^{-8}$. Since the value of $\widehat{\omega}$ is close to 0, the hypotheses (5.9) with $\omega_0 = 0$ will be tested using the one-sided likelihood ratio and score tests. The estimates obtained under the null hypothesis, when $\omega = 0$, are $\hat{\lambda}_1 = 0.627594$ and $\hat{\lambda}_2 = 0.620610$. Following the sequence of parameters in the alternative hypothesis, the efficient scores and information matrix for a single observation are computed as

$$\mathbf{U}^{*T} = [-0.619260 \quad 5.6698 \times 10^{-4} \quad -5.5038 \times 10^{-4}] \text{ and}$$

$$\mathbf{\Gamma}^* = \begin{bmatrix} 1.437955 & -0.475490 & -0.469588 \\ & 1.064594 & -0.534744 \\ & & 1.070557 \end{bmatrix}.$$

Both the *LR* and score test statistics are found to be 0.00 with a *p*-value of 1.00 due to the unconstrained *ML* estimates being very close to the *ML* estimates under H_0. Thus, the null hypothesis is not rejected. There is no evidence of perfect bilateral symmetry for the number of Mullerian glands on the left and right sides of male pigs.

Example 2. A data set on the difference of number of pored lateral line scales on the left and right sides $(R - L)$ of 40 pure banded sunfish (*Enneacanthus gloriosus* Girard) from Collier's Mill Pond, New Jersey is used to investigate developmental stability for the population. The data is from Graham & Felley (1985) (cited in Graham *et al*., 1993). The frequencies for the differences in the number of pored lateral line scales are given in Table 5.8.

Graham & Felley (1985) examined the patterns of asymmetry in 11 populations of fishes including this population of *E. gloriosus*. They conducted *t*-tests on the means and variances of $(L - R)$ of 7 bilateral traits in the 11 populations to check for directional asymmetry. Graham & Felley (1985) concluded that there is no evidence of directional asymmetry and asserted that the asymmetry shown was FA.

Table 5.8

Observed Frequencies for the Difference of Number of Pored Lateral Line Scales on Left and Right Sides $(R - L)$ of 40 *Enneacanthus gloriosus* Individuals

Difference, k	Frequency
-4	0
-3	2
-2	4
-1	8
0	12
1	4
2	5
3	3
4	2
5	0

Now, the zero-inflated *Poisson-Poisson* difference distribution is considered as a model for the data and tested for FA. A fit of the model yields the *ML* estimates of $\hat{\lambda}_1 = 1.901434$, $\hat{\lambda}_2 = 1.650508$ and $\hat{\omega} = 0.103329$. The value of $\hat{\omega}$ is not much larger than 0, showing a slight zero inflation. Thus, the hypotheses (5.9) with $\omega_0 = 0$ will be tested. The *ML* estimates obtained under the null hypothesis, when $\omega = 0$, are $\hat{\lambda}_1 = 1.666655$ and $\hat{\lambda}_2 = 1.441650$. Following the sequence of parameters in the alternative hypothesis, the efficient scores and information matrix for a single observation are computed as

$$\mathbf{U}^{*T} = [1.695795 \quad -4.8589 \times 10^{-6} \quad 1.8962 \times 10^{-5}] \text{ and}$$

$$\mathbf{\Gamma}^{*} = \begin{bmatrix} 3.227096 & -0.240032 & -0.121420 \\ & 0.361808 & -0.275372 \\ & & 0.375299 \end{bmatrix}.$$

The one-sided *LR* and score test statistics are 1.095466 and 0.891118, respectively. The corresponding *p*-values are 0.15 and 0.17, respectively. Thus, the null hypothesis is not rejected, implying that there is evidence of fluctuating asymmetry among the sunfish based on the asymmetry shown in the meristic character of pored lateral line scales. This inference corresponds with the findings of Graham & Felley (1985).

CHAPTER 6 : CONCLUSION AND FURTHER WORK

The extension of trivariate reduction method to construct a class of bivariate and multivariate distribution belonging to the Meixner family of distributions has been considered in this research. These constructed distributions are of a more general form as their marginal distributions are allowed to have different parameters. Such flexibility is required not only in Monte Carlo simulation experiments but also, more importantly, in empirical modelling for better understanding of, and solutions to, real life problems. Although the extended bivariate Meixner hypergeometric distribution has not been considered in this thesis due to the complicated nature of the distribution, this distribution is of interest for future work.

The existence of canonical expansion for this extended class of distributions has also been established to help in the study of the distribution structure. Using the given algorithms, bivariate samples of these distributions can be easily generated on a computer. It may be interesting to apply the extended trivariate reduction method to cover the exponential or related family of distribution, forming flexible marginal distributions and possibly, a wider range of correlations in some cases.

One of the distributions highlighted in the research is a new, extended bivariate negative binomial (*EBNB-I*) distribution constructed by the extended trivariate reduction method and also as a bivariate mixed Poisson model. Aside from the derived basic distributional properties such as the joint probability mass function, factorial moments, correlations and regressions, this distribution is shown to be positive quadrant dependent and hence making it useful, for example, in the field of reliability analysis. Among the other potential applications for this distribution are in the analysis of accidents, absenteeism and ecology. The relevance of this distribution in practice has been demonstrated clearly in the analysis of rain-forest data in Chapter 3. A fit with simulated data shows that when the negative binomial marginals are very different, the more flexible, extended bivariate negative binomial distribution is to be preferred.

Unfortunately, the joint probability functions for bivariate and multivariate distributions usually have complicated or worse, intractable mathematical expressions. Most of them also involve special functions such as the orthonormal Meixner polynomials in the case of *EBNB-I*. Due to this, the use of classical parameter estimation methods such as maximum likelihood estimation has proved to be very tedious and taxing. The situation becomes even more complex when there is a need to account for outliers in the data. Therefore, the MD_α estimation based on the distribution generating function proposed here will be a very appealing method of parameter estimation. This method is fast and robust against outliers. The estimators are also consistent.

The simpler Edwards and Gurland's bivariate negative binomial distribution has been used throughout the simulation study on the competency of the MD_α estimation method. Without jeopardising the robustness and accuracy of the estimators, the MD_α method has been shown to be far superior in the computation time taken as compared to the maximum likelihood and minimum penalized generalized Hellinger distance estimation methods. This method is usually 4 to 6 times faster in obtaining the estimates for a set of data. The clear advantage of MD_α method is due to the simpler generating function used in the computations. The simulation study also shows that the MD_α method works as well as the minimum penalized generalized Hellinger distance method of Basu *et al*. (1997) in the presence of empty cells in the data. In addition, this method is easily extensible to multivariate cases. Thus, MD_α method will indeed be a suitable parameter estimation method for the extended class of bivariate and multivariate distributions in Chapter 3 as illustrated by Example 2 in Chapter 4.

Application of MD_α method to the omnibus driver data in Chapter 4 yields a result consistent with the minimum penalized generalized Hellinger distance method, pointing out the possibility of a presence of outliers in the data set.

In spite of the many attractive characteristics, as with many estimation methods, MD_α method does not perform very well when the sample size of data is small. Further research along this line can be pursued to improve this method. A possible modification is to add an appropriate penalty to the distance measure of the method, as similarly done to produce the minimum penalized generalized Hellinger distance method. Another potential research consideration is to derive the explicit asymptotic efficiency of the MD_α estimators, which has not been investigated here. The asymptotic distribution of the MD_α distance measure will also be of interest for statistical inference. On a different tangent, it may be feasible to use the MD_α distance measure as a test statistic in the hypothesis testing for goodness-of-fit.

Aside from the joint distribution, the distribution of the difference between two random variables is also an important area of study, especially for paired count data which does not seem to be well studied. Various results have been derived and computational issues considered for the case of two independent random variables. For the dependent case, the distribution of the difference between two correlated negative binomial random variables has been proposed to model fluctuating asymmetry, where this distribution is fitted to a sample of differences between an organism's bilateral sides for a meristic trait. Based on a zero-inflated count model, a test for fluctuating asymmetry has also been proposed. Clearly high incidence of zero counts indicates perfect bilateral symmetry for the trait apart from chance variations. In Chapter 5, the one-sided score and likelihood ratio hypothesis tests for fluctuating asymmetry performed on several generated data sets and two real data sets, where computations are based on the simpler Poisson difference distribution, demonstrated the feasibility of implementing this model for fluctuating asymmetry. Further work may be considered in either a regression or Bayesian context.

The contributions of this thesis clearly have useful and interesting applications in many areas. For instance, statistical inference based on generating functions in multivariate situations will reduce the complexity of the computational problems involved. Further and more comprehensive work in the field of multivariate distributions is required, and this field is still attracting the attention of many researchers with a focus towards applications.

REFERENCES

Aitchison, J., and Ho, C. H. (1989). The Multivariate Poisson-Log Normal Distribution. *Biometrika*, **76**, 643-653.

Aitken, A. C., and Gonin, H. T. (1935). On Fourfold Sampling With and Without Replacement. *Proceedings of the Royal Society of Edinburgh*, **55**, 114-125.

Arbous, A. G., and Kerrich, J. E. (1951). Accident Statistics and the Concept of Accident Proneness. *Biometrics*, **7**, 340-432.

Arellano-Valle, R. B., and Azzalini, A. (2008). The Centred Parametrization for the Multivariate Skew-Normal Distribution. *Journal of Multivariate Analysis*, **99**, 1362-1382. Correction: Vol. **100** (2009), 816.

Arnold, B. C. (1967). A Note on Multivariate Distributions with Specified Marginals. *Journal of the American Statistical Association*, **62**, 1460-1461.

Athreya, K. B., and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. New York: Springer.

Azzalini, A. (1985). A Class of Distributions which Includes the Normal Ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

Azzalini, A. (2005). The Skew-Normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, **32**, 159-188.

Balakrishnan, N., and Lai, C. D. (2009). *Continuous Bivariate Distributions*. New York: Springer.

Barnett, V. D. (1966). Evaluation of the Maximum Likelihood Estimator where the Likelihood Equation Has Multiple Roots. *Biometrika*, **53**, 151-166.

Barrett, J. F., and Lampard, D. G. (1955). An Expansion for Some Second-Order Distributions and Its Application to Noise Problems. *IRE Transaction on Information Theory*, *IT-1*, 10-15.

Basu, A. (2002). Outlier Resistant Minimum Divergence Methods in Discrete Parametric Models. *Sankhya: The Indian Journal of Statistics*, **64**, 128-140.

Basu, A., Basu, S., and Chaudhuri, G. (1997). Robust Minimum Divergence Procedures for Count Data Models. *Sankhya: The Indian Journal of Statistics*, **59**, 11-27.

Belzunce, F., Castaño, A., Olvera-Cervantes, A., and Suárez-Llorens, A. (2007). Quantile Curves and Dependence Structure for Bivariate Distributions. *Computational Statistics & Data Analysis*, **51**, 5112-5129.

Bennett, W. R., and Rice, S. O. (1934). Note on Methods of Computing Modulation Products. *Philosophical Magazine*, **18**, 422-424.

Beran, R. (1977). Minimum Hellinger Distance Estimates for Parametric Models. *The Annals of Statistics*, **5**, 445-463.

Biswas, A., and Hwang, J. S. (2002). A New Bivariate Binomial Distribution. *Statistics & Probability Letters*, **60**, 231-240.

Brooks, S. P. (1995). A Hybrid Optimization Algorithm. *Applied Statistics*, **44**, 530-533.

Brooks, S. P., and Morgan, B. J. (1995). Optimization Using Simulated Annealing. *The Statistician*, **44**, 241 -257.

Brown, J. J. (1958). A Criterion for the Diagonal Expansion of a Second-Order Probability Distribution in Orthogonal Polynomials. *IRE Transactions on Information Theory*, *IT-4*, 172.

Burden, R. L., and Faires, J. D. (2005). *Numerical Analysis* (8th ed.). Belmont: Thomson Brooks/Cole.

Campbell, J. T. (1934). The Poisson Correlation Function. *Proceedings of the Edinburgh Mathematical Society: Series 2*, **4**, 18-26.

Cherian, K. C. (1941). A Bivariate Correlated Gamma-Type Distribution Function. *Journal of the Indian Mathematical Society*, **5**, 133-144.

Chib, S., and Winkelmann, R. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business and Economic Statistics*, **19**, 428-435.

Colangelo, A., Hu, T., and Shaked, M. (2008). Conditional Orderings and Positive Dependence. *Journal of Multivariate Analysis*, **99**, 358-371.

Colangelo, A., Scarsini, M., and Shaked, M. (2006). Some Positive Dependence Stochastic Orders. *Journal of Multivariate Analysis*, **97**, 46-78.

Consul, P. C. (1988). On the Difference of Two Generalized Poisson Variates. *Communications in Statistics – Theory and Methods*, **15**, 761-767.

Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Cuadras, C. M. (2002). Correspondence Analysis and Diagonal Expansions in Terms of Distribution Functions. *Journal of Statistical Planning and Inference,* **103***,* 137-150.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer.

Dowling, M., and Nakamura, M. (1997). Estimating Parameters for Discrete Distributions Via the Empirical Probability Generating Function. *Communications in Statistics – Simulation and Computation*, **26**, 301-313.

Downton, F. (1970). Bivariate Exponential Distributions in Reliability Theory. *Journal of the Royal Statistical Society: Series B*, **32**, 408-417.

Eagleson, G. K. (1964). Polynomial Expansions of Bivariate Distributions. *The Annals of Mathematical Statistics,* **35**, 1208-1215.

Eagleson, G. K., and Lancaster, H. O. (1967). The Regression System of Sums with Random Elements In Common. *Australian Journal of Statistics*, **9**, 119-125.

Edwards, C. B., and Gurland, J. (1961). A Class of Distributions Applicable to Accidents. *Journal of the American Statistical Association*, **56***,* 503-517.

Ferrari, A., Letac, G., and Tourneret, J.-Y. (2004). Multivariate Mixed Poisson Distributions. *EUSIPCO* (p. 1067-1070). Vienna: Elsevier.

Fouskakis, D., and Draper, D. (2002). Stochastic Optimization: A Review. *International Statistical Review*, **70**, 315-349.

Graham, J. H., and Felley, J. D. (1985). Genomic Coadaptation and Developmental Stability within Introgressed Populations of *Enneacanthus gloriosus* and *E. obesus* (Pisces, Centrarchidae). *Evolution*, **39**, 104-114.

Graham, J. H., Freeman, D. C., and Emlen, J. M. (1993). Developmental Stability: A Sensitive Indicator of Populations under Stress. In W. G. Landis, J. S. Hughes, and M.

A. Lewis (eds.), *Environmental Toxicology and Risk Assessment* (Vol. 1, p. 136-158). Philadelphia: ASTM Special Technical Publication.

Graham, J. H., Shimizu, K., Emlen, J., Freeman, D. C., and Merkel, J. (2003). Growth Models and the Expected Distribution of Fluctuating Asymmetry. *Biological Journal of the Linnean Society,* **80**, 57-65.

Greenwood, M., and Yule, G. U. (1920). An Inquiry Into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society: Series A*, **83**, 255-279.

Gumbel, E. J. (1960). Bivariate Exponential Distributions. *Journal of the American Statistical Association*, **55**, 698-707.

Gupta, A. K. (1979). On the Expansion of Bivariate Gamma Distribution. *Journal of the Indian Statistical Association*, **17**, 41-50.

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1996). Analysis of Zero Adjusted Count Data. *Computational Statistics & Data Analysis*, **23**, 207-218.

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (2004). Score Test for Zero Inflated Generalized Poisson Regression Model. *Communications in Statistics – Theory and Mehtods,* **33**, 47-64.

Haber, S. (1970). Numerical Evaluation of Multiple Integrals. *SIAM Review*, **12**, 481-526.

Hamdan, M. A. (1972). Canonical Expansion of the Bivariate Binomial Distribution with Unequal Marginal Indices. *International Statistical Review*, **40**, 277-280.

Hamdan, M. A., and Al-Bayyati, H. A. (1971). Canonical Expansion of the Compound Correlated Bivariate Poisson Distribution. *Journal of the American Statistical Association*, **66**, 390-393.

Hamdan, M. A., and Jensen, D. R. (1976). A Bivariate Binomial Distribution and Some Applications. *Australian Journal of Statistics,* **18**, 163 -169.

Hewa-Kapuge, S., and Hoffmann, A. A. (2001). Composite Asymmetry as an Indicator of Quality in the Benificial Wasp *Trichogramma* nr. *brassicae* (Hymenoptera: Trichogrammatidae). *Journal of Economic Entomology*, **94**, 826-830.

Hogg, R. V., and Craig, A. T. (1995). *Introduction to Mathematical Statistics* (5th Edition ed.). New Jersey: Prentice-Hall.

Holgate, P. (1964). Estimation for the Bivariate Poisson Distribution. *Biometrika*, **51**, 241-245.

Holgate, P. (1966). Bivariate Generalizations of Neyman's Type A Distribution. *Biometrika*, **53**, 241-245.

Hutchinson, T. P., and Lai, C. D. (1990). *Continuous Bivariate Distributions: Emphasizing Applications.* Adelaide: Rumsby Scientific Publishing.

Ingber, L. (1992). Genetic Algorithms and Very Fast Simulated Reannealing: A Comparison. *Mathematical and Computer Modelling*, **16**, 87-100.

Irwin, W. (1937). The Frequency Distribution of Difference between Poisson Variates Following the Same Poisson Distribution. *Journal of the Royal Statistical Society: Series A*, **100**, 415-416.

Jensen, D. R. (1971). A Note on Positive Dependence and the Structure of Bivariate Distributions. *SIAM Journal on Applied Mathematics*, **20**, 749-753.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Boca Raton: Chapman and Hall/CRC.

Johnson, N. L. (1959). On An Extension of the Connexion between Poisson and χ^2 Distributions. *Biometrika*, **46**, 352-362.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions* (3rd ed.). Hoboken: Wiley.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions.* New York: Wiley.

Johnson, O., Neely, K., and Waples, R. (2004). Lopsided Fish in the Snake River Basin – Fluctuating Asymmetry as A Way of Assessing Impact of Hatchery Supplementation in Chinook Salmon, *Oncorhynchus tshawytscha*. *Environmental Biology of Fishes*, **69**, 379-393.

Jones, M. C. (2004). Families of Distributions Arising from Distributions of Order Statistics (with Discussion). *TEST*, **13**, 1-43.

Jones, M. C., and Larsen, P. V. (2004). Multivariate Distributions with Support Above the Diagonal. *Biometrika*, **91**, 975-986.

Karlis, D., and Ntzoufras, I. (2000). *Discrete Distributions with Applications in Sports, Technical Report **64***. Greece: Athens University of Economics and Business.

Katz, L. (1946). On the Class of Functions Defined by the Difference Equation $(x+1)f(x+1)=(a+bx)f(x)$. *Annals of Mathematical Statistics*, **17**, 501 (Abstract).

Kemp, A. W. (1979). Convolutions Involving Binomial Pseudo-Variables. *Sankhya: The Indian Journal of Statistics*, **41**, 232-243.

Kemp, C. D., and Kemp, A. W. (1988). Rapid Estimation for Discrete Distributions. *The Statistician*, **37**, 243-255.

Kibble, W. F. (1941). A Two-Variate Gamma Type Distribution. *Sankhya: The Indian Journal of Statistics*, **5**, 137-150.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, **220**, 671-680.

Knierim, U., Van Dongen, S., Forkman, B., Tuyttens, F. A., Spinka, M., Campo, J. L., et al. (2007). Fluctuating Asymmetry as An Animal Welfare Indicator — A Review of Methodology and Validity. *Physiology and Behavior*, **92**, 398-421.

Kocherlakota, S., and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. New York: Marcel Dekker.

Kotz, S. (1974). Multivariate Distributions at A Cross-Road. In *Statistical Distributions in Scientific Work* (Vol. 1, p. 247-270). Dordrecht: Reidel.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions, Volume 1* (2nd ed.). New York: Wiley.

Kundu, D., and Gupta, R. D. (2009). Bivariate Generalized Exponential Distribution. *Journal of Multivariate Analysis*, **100**, 581-593.

Kundu, D., Balakrishnan, N., and Jamalizadeh, A. (2010). Bivariate Birnbaum-Saunders Distribution and Associated Inference. *Journal of Multivariate Analysis*, **101**, 113-125.

Laha, R. G., and Lukacs, E. (1960). On A Problem Connected with Quadratic Regression. *Biometrika*, **47**, 335-343.

Lai, C. D. (1995). Construction of Bivariate Distributions by A Generalised Trivariate Reduction Technique. *Statistics & Probability Letters*, **25**, 265-270.

Lai, C. D., and Moore, T. (1984). Probability Integrals of A Bivariate Gamma Distribution. *Journal of Statistical Computation and Simulation*, **19**, 205-213.

Lai, C. D., and Xie, M. (2000). A New Family of Positive Quadrant Dependent Bivariate Distributions. *Statistics & Probability Letters*, **46**, 359-364.

Lancaster, H. O. (1957). Some Properties of the Bivariate Normal Distribution Considered in teh Form of A Contingency Table. *Biometrika*, **44**, 289-292.

Lancaster, H. O. (1958). The Structure of Bivariate Distributions. *The Annals of Mathematical Statistics*, **29**, 719-736.

Lancaster, H. O. (1963). Correlations and Canonical Forms of Bivariate Distributions. *The Annals of Mathematical Statistics*, **34**, 532-538.

Lancaster, H. O. (1975). Joint Probability Distributions in the Meixner Classes. *Journal of the Royal Statistical Society: Series B*, **37**, 434-443.

Lehmann, E. L. (1966). Some Concepts of Dependence. *The Annals of Mathematical Statistics*, **37**, 1137-1153.

Lin, S.-K. (1985). *Characterization of Lightning As A Disturbance to the Forest Ecosystem in East Texas.* College Station: Texas A & M University.

Lindsay, B. G. (1994). Efficiency versus Robustness: The Case for Minimum Hellinger Distance and Related Methods. *The Annals of Statistics*, **22**, 1081-1114.

Mardia, K. V. (1967). Some Contributions to Contingency-Type Bivariate Distributions. *Biometrika*, **54**, 235-249.

Mardia, K. V. (1970). *Families of Bivariate Distributions.* London: Charles Griffin and Company Limited.

Mather, K. (1953). Genetical Control of Stability in Development. *Heredity*, **7**, 297-336.

Meintanis, S., and Swanepoel, J. (2007). Bootstrap Goodness-of-Fit Tests with Estimated Parameters based on Empirical Transforms. *Statistics & Probability Letters*, **77**, 1004-1013.

Meixner, J. (1934). Orthogonale Polynomsysteme Mit Einer Besonderen Gestalt Der Erzeugenden Funktion. *Journal of the London Mathematical Society*, **9**, 6-13.

Mendonca, P. R., and Caloba, L. P. (1997). New Simulated Annealing Algorithms. *IEEE International Symposium on Circuits and Systems* (p. 1668-1671). Hong Kong: IEEE.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculation by Fast Computing Machines. *Journal of Chemical Physics*, **21**, 1087-1091.

Michael, J. R., and Schucany, W. R. (2002). The Mixture Approach for Simulating Bivariate Distributions with Specified Correlations. *The American Statistician*, **56**, 48-54.

Minhajuddin, A., Harris, I. R., and Schucany, W. R. (2004). Simulating Multivariate Distributions with Specific Correlations. *Journal of Statistical Computation and Simulation*, **74**, 599-607.

Mitchell, C. R., and Paulson, A. S. (1981). A New Bivariate Negative Binomial Distribution. *Naval Research Logistics Quarterly*, **28**, 359-374.

Morris, C. N. (1982). Natural Exponential Families with Quadratic Variance Function. *Annals of Statistics*, **10**, 65-80.

Mosimann, J. E., and Campbell, G. (1988). Applications in Biology: Simple Growth Models. In E. L. Crow, and K. Shimizu (eds.), *Lognormal Distributions: Theory and Applications* (p. 287-302). New York: Marcel Dekker.

Newey, W. K., and McFadden, D. (1994). Large Sample Estimation and Hypothesis Testing. In R. F. Engle, and D. L. McFadden (eds.), *Handbook of Econometrics* (Vol. IV, p. 2112-2245). Amsterdam: Elsevier Science Publishers.

Ong, S. H. (1988). A Discrete Charlier Series Distribution. *Biometrical Journal*, **30**, 1003-1009.

Ong, S. H. (1990). Mixture Formulations of A Bivariate Negative Binomial Distribution with Applications. *Communications in Statistics – Theory and Methods*, **19**, 1303-1322.

Ong, S. H. (1992). The Computer Generation of Bivariate Binomial Variables with Given Marginals and Correlation. *Communications in Statistics – Simulation and Computation*, **21**, 285-299.

Ong, S. H. (1993). Canonical Expansions, Correlation Structure and Conditional Distributions of Bivariate Distributions Generated by Mixtures. *Communications in Statistics – Theory and Methods*, **22**, 2527-2547.

Ong, S. H. (1995). Computation of Bivariate Gamma and Inverted Beta Distribution Functions. *Journal of Statistical Computation and Simulation*, **51**, 153-163.

Ong, S. H. (2008). Generation of Mixtures of Bivariate and Multivariate Distributions with Given Marginals and Correlation. In A. Sengupta (ed.), *Multivariate Statistical Methods with Recently Emerging Trends* (p. 13-23). New Delhi: Macmillan.

Ong, S. H., and Lee, P. A. (1986). Bivariate Non-Central Negative Binomial Distribution: Another Generalization. *Metrika*, **33**, 29-46.

Ong, S. H., and Shimizu, K. (2003). A Class of Discrete Distributions Expressed as Difference of Two Random Variables. *Bulletin of the ISI, 54th ISI Session, Berlin, 13-20 August 2003*, 194-195.

Ong, S. H., Shimizu, K., and Ng, C. M. (2008). A Class of Discrete Distributions Arising from Difference of Two Random Variables. *Computational Statistics & Data Analysis*, **52**, 1490-1499.

Ord, J. K. (1967). On A System of Discrete Distributions. *Biometrika*, **58**, 649-656.

Ord, J. K. (1972). *Families of Frequency Distributions.* London: Griffin.

Palmer, A. R. (1994). Fluctuating Asymmetry Analyses: A Primer. In T. A. Markow (ed.), *Developmental Instability: Its Origins and Evolutionary Implications* (p. 335-364). Dordrecht: Kluwer.

Palmer, A. R., and Strobeck, C. (1986). Fluctuating Asymmetry: Measurement, Analysis and Patterns. *Annual Review of Ecology and Systematics*, **17**, 391-421.

Parker, J. R. (2000). Simulated Annealing for Fitting Linear Combinations of Gaussians to Data. *Computing*, **65**, 291-312.

Patil, M. K., and Shirke, D. T. (2007). Testing Parameter of the Power Series Distribution of A Zero Inflated Power Series Model. *Statistical Methodology*, **4**, 393-406.

Pearson, K. (1897). Mathematical Contributions to the Theory of Evolution: On A Form of Spurious Correlation which May Arise when Indices are Used in the Measurement of Organs. *Proceedings of the Royal Society of London A*, **60**, 489-498.

Press, S. J. (1972). *Applied Multivariate Analysis.* New York: Holt, Reinhart and Winston.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (2nd ed., Fortran Numerical Recipes: Vol. 1). Cambridge: Cambridge University Press.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd ed.). New York: John Wiley and Sons.

Remillard, B., and Theodorescu, R. (2000). Inference Based on the Empirical Probability Generating Function for Mixtures of Poisson Distributions. *Statistics and Decisions*, **18**, 349-366.

Rosner, B. (2000). *Fundamentals of Biostatistics* (5th ed.). California: Duxbury.

Roulin, A., Ducrest, A. L., Balloux, F., Dijkstra, C., and Riols, C. (2003). A Female Melanin Ornament Signals Offspring Fluctuating Asymmetry in the Barn Owl. *Proceedings of the Royal Society of London B,* **270**, 167-171.

Rueda, R., and O'Reilly, F. (1999). Tests of Fit for Discrete Distributions Based on the Probability Generating Function. *Communications in Statistics – Simulation and Computation*, **28**, 259-274.

Self, S. G., and Liang, K. Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association,* **82**, 605-610.

Shapiro, A. (1988). Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis. *International Statistical Review*, **56**, 49-62.

Silvapulle, M. J., and Silvapulle, P. (1995). A Score Test Against One-Sided Alternatives. *Journal of the American Statistical Association*, **90**, 342-349.

Simpson, D. G. (1987). Minimum Hellinger Distance Estimation for the Analysis of Count Data. *Journal of the American Statistical Association*, **82**, 802-807.

Simpson, D. G. (1989). Hellinger Deviance Tests: Efficiency, Breakdown Points, and Examples. *Journal of the American Statistical Association*, **84**, 107-113.

Skellam, J. G. (1946). The Frequency Distribution of Difference between Two Poisson Variates Belonging to Different Populations. *Journal of the Royal Statistical Society: Series A*, **109**, 296.

Stein, G. Z., and Juritz, J. M. (1987). Bivariate Compound Poisson Distributions. *Communications in Statistics – Theory and Methods*, **16**, 3591-3607.

Subrahmaniam, K. (1966). A Test of 'Intrinsic Correlation' in the Theory of Accident Proneness. *Journal of the Royal Statistical Society: Series B*, **35**, 131-146.

Subrahmaniam, K., and Subrahmaniam, K. (1973). On the Estimation of the Parameters in the Bivariate Negative Binomial Distribution. *Journal of the Royal Statistical Society: Series B*, **35**, 131-146.

Sugita, T., Shimizu, K., Ong, S. H., and Ng, C. M. (2010). A Family which Integrates the Generalized Charlier Series and Extended Non-Central Negative Binomial Distributions. *Communications in Statistics – Theory and Methods*, (To appear).

Sundt, B., and Jewell, W. S. (1981). Further Results on Recursive Evaluation of Compound Distributions. *ASTIN Bulletin*, **18**, 27-39.

Szu, H., and Hartley, R. (1987). Fast Simulated Annealing. *Physics Letters A*, **122**, 157-162.

Tamura, R. N., and Boos, D. D. (1986). Minimum Hellinger Distance Estimation for Multivariate Location and Covariance. *Journal of the American Statistical Association*, **81**, 223-229.

Van Valen, L. (1962). A Study of Fluctuating Asymmetry. *Evolution*, **16**, 125-142.

Wald, A. (1949). Note on the Consistency of the Maximum Likelihood Estimate. *The Annals of Mathematical Statistics*, **20**, 595-601.

Woods, R. E., Hercus, M. J., and Hoffmann, A. A. (1998). Estimating the Heritability of Fluctuating Asymmetry in Field Drosophila. *Evolution*, **52**, 816-824.

Young, W. P., Frenyea, K., Wheeler, P. A., and Thorgaard, G. H. (2009). No Increase in Developmental Deformities or Fluctuating Asymmetry in Rainbow Trout (*Oncorhynchus mykiss*) Produced with Cryopreserved Sperm. *Aquaculture*, **289**, 13-18.

Zhao, Y., and Joe, H. (2005). Composite Likelihood Estimation in Multivariate Data Analysis. *The Canadian Journal of Statistics*, **33**, 335-356.

APPENDIX A : DATA FOR GRAPHS IN TEXT

Table A1 (Figure 4.1)

Parameter Bias Measures from Simulations with N Monte Carlo Samples with selected Parameter Values, $p_1 = 0.4$, $p_2 = 0.5$, $\theta_3 = 0.3$ and $\nu = 4.0$, for *BNB* Distribution

Parameter	Sample Size, N				
	100	200	500	800	1000
p_1	0.013027	0.008328	0.002726	0.001427	0.000885
p_2	0.014044	0.009034	0.002974	0.001412	0.000894
θ_3	0.006337	0.004114	0.000814	0.000036	-0.000230
ν	-0.196980	-0.114700	-0.016710	0.005439	0.013207
Log Likelihood	-1857.4495	-1856.0878	-1855.6510	-1855.8358	-1855.5725
Computation Time (Minutes)	6.4684	12.4945	31.2583	48.5399	62.8610

Table A2 (Figure 4.2)

Approximation of the Distance Measure Integral Values for Several Combinations of Number of Quadrature Points (z_1, z_2)

a) MD_1

z_1	z_2				
	2	3	4	5	6
2	1.492384	1.497824	1.493699	1.493091	1.493176
3	1.471525	1.474015	1.469258	1.468923	1.469100
4	1.473321	1.476654	1.472398	1.472225	1.472408
5	1.473380	1.476925	1.472774	1.472616	1.472798
6	1.473385	1.476939	1.472791	1.472634	1.472816

b) $MD_{1/2}$

z_1	z_2				
	2	3	4	5	6
2	2.503785	2.488645	2.486656	2.486768	2.486818
3	2.462397	2.452109	2.449587	2.449962	2.449996
4	2.463505	2.453181	2.450873	2.451249	2.451272
5	2.463542	2.453281	2.450971	2.451340	2.451364
6	2.463541	2.453275	2.450963	2.451333	2.451357

APPENDIX B : PARTIAL DERIVATES AND ELEMENTS OF THE INFORMATION MATRIX FOR *NB-NB* DISTRIBUTION

Let $X = X_1 - X_2$, $p_i = \frac{\lambda_i\theta_i}{1+\lambda_i\theta_i}$ and $\alpha_i = \frac{1}{\theta_i}$ for $i = 1,2$, where λ_1 and λ_2 are the means for $X_1 \sim NB(\alpha_1, p_1)$ and $X_2 \sim NB(\alpha_2, p_2)$ respectively. Also, $\Pr(X = k) = f(k)$ is as given in (5.1).

(B1) Partial Derivatives of *NB-NB* Probabilities

Let $u_i = \frac{1}{1+\lambda_i\theta_i}$ for $i = 1,2$.

$$\frac{\partial f(k)}{\partial \lambda_1} = \sum_{j=0}^{\infty} u_1(u_1\lambda_1\theta_1)^j\,[f(k-j-1) - f(k-j)]$$

$$\frac{\partial f(k)}{\partial \lambda_2} = \sum_{j=0}^{\infty} u_2(u_2\lambda_2\theta_2)^j\,[f(k+j+1) - f(k+j)]$$

$$\frac{\partial f(k)}{\partial \theta_1} = \sum_{j=1}^{\infty} (u_1\lambda_1)^{j+1}\theta_1^{j-1}\left[\frac{jf(k-j-1)}{j+1} - f(k-j) + \frac{f(k)}{j+1}\right]$$

$$\frac{\partial f(k)}{\partial \theta_2} = \sum_{j=1}^{\infty} (u_2\lambda_2)^{j+1}\theta_2^{j-1}\left[\frac{jf(k+j+1)}{j+1} - f(k+j) + \frac{f(k)}{j+1}\right]$$

(B2) Partial Derivatives of Log Likelihood Function

$n_k =$ Observed frequency for $(X = k)$

$$\frac{\partial \ln L}{\partial \lambda_i} = \sum_{k=-\infty}^{\infty} \frac{n_k}{f(k)}\frac{\partial f(k)}{\partial \lambda_i}, \qquad i = 1,2$$

$$\frac{\partial \ln L}{\partial \theta_i} = \sum_{k=-\infty}^{\infty} \frac{n_k}{f(k)}\frac{\partial f(k)}{\partial \theta_i}, \qquad i = 1,2$$

(B3) Elements of the Information Matrix

By using the relation $\mathbf{\Gamma}(\mathbf{\Phi}) = \frac{1}{n}\left\{E\left[-\frac{\partial^2 \ln L(\mathbf{\Phi};\mathbf{x})}{\partial\phi_i\partial\phi_j}\right]\right\} = \left\{E\left[\left(\frac{1}{f(k)}\right)^2 \frac{\partial f(k)}{\partial\phi_i}\frac{\partial f(k)}{\partial\phi_j}\right]\right\}$, expectations for the second partial derivatives of the log likelihood function, for $i, j = 1,2$, are given by

$$E\left[-\frac{\partial^2 \ln L}{\partial\lambda_i\partial\lambda_j}\right] = n\sum_{\mathrm{k}=-\infty}^{\infty}\frac{1}{f(k)}\frac{\partial f(k)}{\partial\lambda_i}\frac{\partial f(k)}{\partial\lambda_j}\ , \qquad i,j = 1,2$$

$$E\left[-\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_j}\right] = n\sum_{\mathrm{k}=-\infty}^{\infty}\frac{1}{f(k)}\frac{\partial f(k)}{\partial\theta_i}\frac{\partial f(k)}{\partial\theta_j}\ , \qquad i,j = 1,2$$

$$E\left[-\frac{\partial^2 \ln L}{\partial\lambda_i\partial\theta_j}\right] = E\left[-\frac{\partial^2 \ln L}{\partial\theta_j\partial\lambda_i}\right] = n\sum_{\mathrm{k}=-\infty}^{\infty}\frac{1}{f(k)}\frac{\partial f(k)}{\partial\lambda_i}\frac{\partial f(k)}{\partial\theta_j}\ , \qquad i,j = 1,2$$

where $n =$ sample size.

APPENDIX C : PARTIAL DERIVATES AND ELEMENTS OF THE INFORMATION MATRIX FOR *POISSON-POISSON* DISTRIBUTION

$f(k)$ and $P(k)$ are as given in (5.6) and (5.8), respectively.

(C1) Partial Derivatives of *Poisson-Poisson* Probabilities

$$\frac{\partial f(k)}{\partial \lambda_1} = f(k-1) - f(k)$$

$$\frac{\partial f(k)}{\partial \lambda_2} = f(k+1) - f(k)$$

$$\frac{\partial^2 f(k)}{\partial {\lambda_1}^2} = f(k-2) - 2f(k-1) + f(k)$$

$$\frac{\partial^2 f(k)}{\partial {\lambda_2}^2} = f(k+2) - 2f(k+1) + f(k)$$

$$\frac{\partial^2 f(k)}{\partial \lambda_1 \partial \lambda_2} = \frac{\partial^2 f(k)}{\partial \lambda_2 \partial \lambda_1} = 2f(k) - f(k-1) - f(k+1)$$

(C2) Partial Derivatives of Log Likelihood Function of Zero-Inflated *Poisson-Poisson* Distribution

$n_k =$ Observed frequency for $(X = k)$

$$\ln L = \sum_{k=-\infty}^{\infty} n_k \ln P(k)$$

$$\frac{\partial \ln L}{\partial \omega} = \frac{n_0\big(1 - f(0)\big)}{f(0) + \omega\big(1 - f(0)\big)} + \sum_{k \neq 0} \frac{-n_k}{(1-\omega)}$$

$$\frac{\partial \ln L}{\partial \lambda_i} = \frac{n_0(1-\omega)\dfrac{\partial f(0)}{\partial \lambda_i}}{f(0) + \omega\big(1 - f(0)\big)} + \sum_{k \neq 0} n_k \frac{1}{f(k)} \frac{\partial f(k)}{\partial \lambda_i}, \qquad i = 1,2$$

(C3) Elements of the Information Matrix

$n =$ sample size

$$E\left[-\frac{\partial^2 \ln L}{\partial \omega^2}\right] = \frac{nf(0)\big(1-f(0)\big)^2}{\Big(f(0)+\omega\big(1-f(0)\big)\Big)^2} + \sum_{k\neq 0}\frac{nf(k)}{(1-\omega)^2}$$

$$E\left[-\frac{\partial^2 \ln L}{\partial \omega \partial \lambda_i}\right] = E\left[-\frac{\partial^2 \ln L}{\partial \lambda_i \partial \omega}\right] = \frac{nf(0)\frac{\partial f(0)}{\partial \lambda_i}}{\Big(f(0)+\omega\big(1-f(0)\big)\Big)^2}, \qquad i = 1,2$$

$$E\left[-\frac{\partial^2 \ln L}{\partial \lambda_i \partial \lambda_j}\right] = n\left[\frac{f(0)(1-\omega)^2}{\Big(f(0)+\omega\big(1-f(0)\big)\Big)^2}\frac{\partial f(0)}{\partial \lambda_i}\frac{\partial f(0)}{\partial \lambda_j}\right.$$

$$\left. - \frac{f(0)(1-\omega)}{f(0)+\omega\big(1-f(0)\big)}\frac{\partial^2 f(0)}{\partial \lambda_i \partial \lambda_j}\right]$$

$$+ n\sum_{k\neq 0}\left(\frac{1}{f(k)}\frac{\partial f(k)}{\partial \lambda_i}\frac{\partial f(k)}{\partial \lambda_j} - \frac{\partial^2 f(k)}{\partial \lambda_i \partial \lambda_j}\right), \qquad i,j = 1,2$$