Models and Statistical Inference for Multivariate Count Data

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Pankaj Bhagwat

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2019

Supervisor: Prof. Eric Marchand

Certificate

This is to certify that this dissertation entitled **Models and Statistical Inference for Multivariate Count Data** towards the partial fulfilment of the **BS-MS** dual degree programme at the **Indian Institute of Science Education and Research, Pune** represents study/work carried out by **Pankaj Bhagwat** at the Université de Sherbrooke, Canada and Indian Institute of Science Education and Research, Pune under the supervision of **Prof. Eric Marchand**, Department of Mathematics and Statistics , Université de Sherbrooke, Quebec, Canada , during the academic year **2018-2019**.

Éric Marchand

Prof. Eric Marchand

Committee:

Prof. Eric Marchand

Prof. Uttara Naik-Nimbalkar

This thesis is dedicated to my parents and my brother Rohit

Declaration

I hereby declare that the matter embodied in the report entitled **Models and Statistical Inference for Multivariate Count Data** are the results of the work carried out by me at the Université de Sherbrooke, Quebec, Canada and Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Eric Marchand and the same has not been submitted elsewhere for any other degree.

Bhagwat

Pankaj Bhagwat

Acknowledgments

I would like to express my sincere gratitude towards Prof. Eric Marchand for his constant support, insightful discussions and guidance. I am grateful to him for encouraging me to think independently. I am also thankful for providing opportunities to interact with other researchers which were certainly helpful for the project. I thank Prof. Uttara Naik Nimbalkar for monitoring the project, providing constructive feedback and discussions. I would like to thank the Université de Sherbrooke, Quebec, Canada for hosting me. I am also thankful to the Mathematics Department at IISER Pune for allowing me to carry out this project at the Université de Sherbrooke. I would also like to thank Mitacs and DST-INSPIRE for the financial support which made my stay in Canada possible. Finally, I would like to thank my parents and friends for their consistent support.

Abstract

We investigate different multivariate discrete distributions. In particular, we study the multivariate sums and shares model for multivariate count data proposed by Jones and Marchand. One such model consists of Negative binomial sums and Polya shares. We address the parameter estimation problem for this model using the method of moments, maximum likelihood, and a Bayesian approach. We also propose a general Bayesian setup for the estimation of parameters of a Negative binomial distribution and a Polya distribution. Simulation studies are conducted to compare the performances of different estimators. The methods developed are implemented on real datasets. We also present an example of a proper Bayes point estimator which is inadmissible. Other intriguing features are exhibited by the Bayes estimator, one such feature is the constancy with respect to the large class of priors.

Contents

List of Tables

List of Figures

Notations and Abbreviations

$\Gamma(a)$	$\int_0^\infty z^{a-1}e^{-z}dz$, Gamma function
$\mathbb{R}_{\geq 0}$	$\{r \in \mathbb{R} : r \geq 0\}$, set of non-negative real numbers
$\mathbf{I}$	$[0,1]$
$(a)_m$	$\frac{\Gamma(a+m)}{\Gamma(a)}$, Pochhammer symbol
$\mathbb{E}[X]$	Expectation of a random variable X
$\mathbb{I}_{(a,b)}(x)$	Indicator function
$\mathbb{N}$	$\{0, 1, 2, 3, \cdots\}$, set of non-negative integers
$\mathbb{V}[X]$	Variance of a random variable X
$\rho(X,Y)$	correlation between random variables X and Y
$B(a,b)$	Beta function $\left(\int\limits_0^1 \theta^{a-1}(1-\theta)^{b-1}d\theta\right)$
$Cov(X,Y)$	covariance between random variables X and Y
$N_d(\mu, \Sigma)$	d-variate Normal distribution
$X \sim F(X)$	random variable X has a distribution $F(X)$
c.d.f.	cumulative distribution function
log (x)	natural logarithm of x
MCMC	Monte Carlo Morkov Chain

MLE	maximum likelihood estimator
MoM	method of moments
p.d.f.	probability density function
p.m.f.	probability mass function
s.d.	standard deviation

Introduction

Problem and Motivation

Multivariate count data arises very often, for example, when one wants to analyze the number of insurance claims falling in to different time periods, or the number of dengue cases falling in to different locations. To model general dependencies among the counts in such scenarios, multivariate discrete distributions are needed. There is a vast literature on multivariate continuous distributions. It is because of the availability of natural generalizations of univariate continuous distributions to their multivariate counterparts which covers full ranges of correlations. This natural generalization is not always possible for the discrete distributions. There are some methods to construct multivariate discrete models. Johnson, Kotz and Balkrishnan (2004) have provided a book-length treatment on discrete multivariate distributions in [16] with the focus on strategies for the construction of multivariate discrete distributions. Kocherlakota , S. and Kocherlakota, K. (1992) also provides a survey of generating methods for bivariate discrete distributions in [18]. But most of them fail to cater to all correlation structures in the data, even for bivariate cases. Unlike construction of multivariate discrete distributions, the problem of parameter estimation for these models has not been addressed thoroughly. The main reason is the complexity of the likelihood functions which forbid the development of the estimation methods for such models. As the computational facilities became available in the recent years, parameter estimation methods can be addressed more effectively.

A recent manuscript of Jones and Marchand [17] introduces a simple and appealing two-step strategy for decomposing or generating multivariate count data. The strategy is highly appealing since for the bivariate cases, it covers the whole range of correlations. A rich ensemble of distributions arise using this strategy. Several known distributions can be recovered, including bivariate cases which are prominent in the literature, and several extensions or novel distributions can also be obtained. One such model consists of Negative Binomial (a, θ) sums and Polya shares with parameters $t, \alpha_1, \ldots, \alpha_d$, with the corresponding probability mass function (p.m.f.) for $M_1, \ldots, M_d$ reducing to :

$$p(m_1, \cdots, m_d) = \frac{(a)_{m_1+\cdots+m_d} \prod_{i=1}^{d} (\alpha)_{m_i}}{m_1! m_2! \cdots m_d! \big(\sum_{i=1}^{d} \alpha_i \big)_{m_1+m_2+\cdots+m_d}} \theta^{\alpha} (1-\theta)^{m_1+m_2+\cdots+m_d} \qquad (1)$$

$m_1, \ldots, m_d \in \mathbb{N}$. Several challenges related to the estimation of the parameters $a, \theta, \alpha_1, \ldots, \alpha_d$ arise, as well as fitting the above p.m.f. to actual datasets, namely for the purposes of prediction. The main goal of the thesis is to study the problem of parameter estimation for Negative binomial sums and Polya shares model. We explore the implementation of estimation based on the methods of moments, maximum likelihood, as well as Bayesian methods. In particular, Bayesian methods are developed with a focus on interpretation and prior-posterior analysis. This problem of parameter estimation is decomposed into two separate sub-problems:

- Parameter estimation for a Negative Binomial distribution
- Parameter estimation for a Polya distribution

The two parameter Negative binomial distribution model has been studied in terms of method of moments and maximum likelihood (for instance, Fisher (1941, [11]) , Dropkin (1959, [9]) , Savani, *et al.* (2006, [31])). It is well known that both methods may lead to infeasible estimators. A Bayesian approach can be used to avoid such problems. But, not much work is done from the Bayesian perspective. Due to the complexity of the likelihood function, the posterior distributions becomes intractable. Bradlow, *et al.* (2002), in [7], suggest closed form approximations for the posterior moments of the parameters using polynomial expansions. We provide a family of distributions for the parameters which is semiconjugate for Negative binomial distribution. This allows us to use Gibbs sampler for sampling from the posterior distributions. The comparison of the risk performances of the available estimators is also of high interest. Besides, the methods are implemented on data sets and compared.

The Polya distribution also leads to convoluted likelihood function which makes the parameter estimation for such model challenging. In this case, maximum likelihood estimators (MLE) may not exist (Levin and Reeds (1977, [23])). Whenever, they exist, one needs to employ numerical methods to find the estimators. We propose the use of data cloning method for the estimation of MLE. Lele, *et al.* (2010) proposed a method of data cloning for MLE estimators in a random effect models in [22]. This method also provides estimates for the standard errors related to MLE. We also provide a semiconjugate family of priors for the Polya distribution and the use of the Gibbs sampler for sampling from the posterior distributions which enables to approximate posterior densities and posterior moments of the parameters.

While studying Bayesian inference for Negative binomial model, we found an interesting example of proper Bayes estimator which is also inadmissible. Proper Bayes estimators are generally admissible. There are very few constructive examples in the literature of inadmissible proper Bayes estimators. We provide an occurrence of such instance in a very natural setting. Other intriguing features are exhibited by this estimator, one such is the constancy of the Bayes estimator with respect to the large class of priors.

Outline

In Chapter 1, we provide lists of notions used throughout this work. In Chapter 2, we review different methods of generating or constructing multivariate discrete distributions. Chapter 3 introduces the multivariate sums and shares model proposed by Jones and Marchand [17]. Statistical properties such as moments and correlation structure of the model are discussed in this chapter. This chapter also introduces the main aim of the thesis which is the parameter estimation problem for the negative binomial sums and Polya shares model. In Chapter 4, we provide the details of the different estimators for the two parameter unknown negative binomial distribution. This involves the novel Bayesian setup for such model which is an important outcome of the thesis. In Chapter 5, we study the parameter estimation problem for the Polya distribution. This involves maximum likelihood estimators and Bayesian estimators for the parameters of a Polya distribution. Later in Chapter 6, we combine methods of parameter estimations obtained in previous chapters to address the estimation problem for the Negative binomial sums and Polya shares model. In Chapter 7, we present an example of proper Bayes but an inadmissible estimator. An article based on this chapter is accepted for publication in *The American Statistician.*

Contributions

This thesis builds on the article of Jones and Marchand [17] which introduces sums and shares model, but extends the material in several ways. The main contributions of the thesis are as follows:

1. We provide a more general Bayesian setup for the Negative binomial distribution which allows the use of Gibbs sampler for the sampling from the posterior distribution.

2. We propose a new class of priors which is semiconjugate for the negative binomial distribution. This is a generalization of Gamma and Lindley distributions, which are used frequently.

3. We propose the use of a data cloning method for the estimation of parameters of a Polya distribution. This method of estimating MLE was proposed by Lele, *et al.* (2010) for a random effect models.

4. We extend the work of Jones and Marchand on multivariate discrete distributions via sums and shares in terms of the parameter estimation for the Negative binomial sums and Polya shares model proposed by them.

5. We also provide an interesting example of proper Bayes estimator which is also inadmissible. This estimator also has other interesting features as well. (*Accepted for the publication in The American Statistician.*)

Chapter 1

Preliminaries

The following list summarizes basic definitions and notations used throughout this work. The reader may skip this chapter and revisit when needed.

Definition 1.0.1. *A* ***multivariate discrete probability mass function*** *on* $\mathbb{N}^d$ *is any function* $f : \mathbb{N}^d \to \mathbb{R}_{\geq 0}$ *for some positive integer* d*, such that* $\sum\limits_{x \in \mathbb{N}^d} f(x) = 1$*. In case of* $d = 1$*,* f *is the usual univariate probability mass function.*

Definition 1.0.2. *Any function* $H : \mathbb{R} \to [0, 1]$ *is a* ***distribution function*** *if following holds:*

1) $H(x)$ *is non-decreasing in* x*.*
2) $\lim\limits_{x \to -\infty} H(x) = 0$*.*
3) $\lim\limits_{x \to +\infty} H(x) = 1$*.*
4) H *is right continuous i.e.* $\lim\limits_{x \downarrow x_0} H(x) = H(x_0)$

Definition 1.0.3. *A* ***joint distribution function*** *is a function* $F : \mathbb{R}^d \to [0, 1]$ *such that*

1) $\lim\limits_{x_1, \cdots, x_d \to +\infty} F(x_1, \cdots, x_d) = 1$*.*
2) $\lim_{x_i \to -\infty} F(x_1, \cdots, x_d) = 0$ *,* $\forall i \in \{1, 2, \cdots, d\}$*.*
3) *For any* $(x_1, \cdots, x_d)$ *and* $(y_1, \cdots, y_d) \in \mathbb{R}^d$ *such that* $x_i \leq y_i, \forall i \in \{1, 2, \cdots, d\}$*,*

$$\sum_{w_i \in \{x_i, y_i\}} (-1)^{\sum_{i=1}^{d} \mathbf{1}_{(y_i, w_i)}} F(w_1, \cdots, w_d) \geq 0, \tag{1.1}$$

where $\boldsymbol{1}_{(y_i,w_i)} = \begin{cases} 0 & if\ \ y_i = w_i \\ 1 & otherwise. \end{cases}$

Definition 1.0.4. *The **margins** of a joint distribution function F are given by*

$$F_i(x) = \lim_{x_j \to \infty; j \neq i} F(x_1, \cdots, x_i = x, \cdots, x_n).$$

Definition 1.0.5. *A discrete random variable X is said to have **Poisson distribution** with mean λ if p.m.f. is given as*

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad \lambda > 0, \tag{1.2}$$

where $x \in \mathbb{N}$.

Definition 1.0.6. *A continuous random variable X is said to have **Gamma distribution** with shape parameter $a > 0$ and scale parameter $b > 0$, if probability density function is given as*

$$f(x|a,b) = \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}, \quad x > 0. \tag{1.3}$$

Definition 1.0.7. *A continuous random variable X is said to have **Beta distribution** with parameters $a\,, b > 0$, if probability density function is given as*

$$f(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad x \in \boldsymbol{I}. \tag{1.4}$$

Definition 1.0.8. *A random variable $U = (U_1, \ldots, U_d)$, $d \geq 2$, is said to have **Dirichlet distribution** with parameters $\alpha_1, \ldots, \alpha_d > 0$ if probability density function is given as*

$$h(u_1, \ldots, u_{d-1}|\alpha_1, \ldots, \alpha_d) = \frac{\Gamma(\sum\limits_{i=1}^{d} \alpha_i) \prod_{i=1}^d u_i^{\alpha_i - 1}}{\prod_{i=1}^d \Gamma(\alpha_i)}, \tag{1.5}$$

where $u_i \in \boldsymbol{I}$ such that $\sum\limits_{i=1}^{d} u_i = 1$.

Definition 1.0.9. *A multivariate discrete random variable* $X = (X_1, \ldots, X_d)$, $d \geq 2$, *is said to have* ***Multinomial distribution*** *with parameters* $t \in \mathbb{N}$ *and* $t, u_1, \ldots, u_d > 0$, *if probability mass function is given as*

$$f(x_1, \ldots, x_{d-1}|t, u_1, \ldots, u_d) = \frac{t! \prod_{i=1}^{d} u_i^{x_i - 1}}{\prod_{i=1}^{d} x_i!}, \quad x_i \in \{0, \cdots, t\}, \quad \sum_{i=1}^{d} x_i = t, \tag{1.6}$$

where $u_i \in \boldsymbol{I}$ *such that* $\sum\limits_{i=1}^{d} u_i = 1$.

Definition 1.0.10. *The* ***moment generating function*** *for the random variable* X *with c.d.f* $\mathbb{F}_X$ *is defined as*

$$\mathbb{M}_X(s) = \mathbb{E}(e^{sX}). \tag{1.7}$$

Definition 1.0.11. *Any estimator* $\delta(X)$ *for a parameter* θ *is said to be an* ***unbiased estimator*** *if* $\mathbb{E}(\delta(X)) = \theta$.

Bayesian Inference

Let $\theta \in \Theta$ be parameter vector and Y be a random variable with p.d.f. (or p.m.f.) $f(y|\theta)$. We observe y, a realization of Y. Assume prior density $g(\theta)$ for θ. Then the joint probability density function for θ and y can be written as the product of the prior density $g(\theta)$ and $f(y|\theta)$ as follows:

$$\pi(y, \theta) = g(\theta) f(y|\theta)$$

Bayes rule is used to obtain expressions for the posterior density of θ ($\pi(\theta|y)$), the marginal density of y ($f(y)$), and the posterior predictive density ($f(\hat{y}|y)$) as follows:

- $f(y) = \int\limits_{\Theta} \pi(y, \theta) d\theta = \int\limits_{\Theta} f(y|\theta) g(\theta) d\theta$.
- $\pi(\theta|y) = \frac{\pi(\theta) f(y|\theta)}{f(y)}$.
- $f(\hat{y}|y) = \int\limits_{\Theta} f(\hat{y}|\theta) \pi(\theta|y) d\theta$.

Statistical Decision Theory

Consider a random variable X with p.d.f. (or p.m.f.) $f(x|\theta)$, where $\theta \in \Theta$. Θ is called as parameter space. Now, consider the problem of estimating θ based on the observations from X. Let $\mathscr{D}$ be the set of all estimators $(\delta(X))$ of θ.

Definition 1.0.12. *A* ***loss function*** *is any function* $L : \Theta \times \mathscr{D} \to \mathbb{R}_{\geq 0}$.

Example 1. *Quadratic loss function:* $L(\theta, \delta(X)) = (\theta - \delta(X))^2$.

Example 2. *Absolute deviation loss function:* $L(\theta, \delta(X)) = |\theta - \delta(X)|$.

Definition 1.0.13. *A* ***Risk function*** $R(\theta, \delta)$ *with respect to a loss function* $L(\theta, \delta(X))$ *is defined as function*

$$R(\theta, \delta) = \mathbb{E}(L(\theta, \delta(X))). \tag{1.8}$$

Definition 1.0.14. *A* ***Bayes Risk*** $r(\delta)$ *with respect to a loss function* $L(\theta, \delta(X))$ *and prior* $\pi(\theta)$ *is defined as function*

$$r_\pi(\delta) = \int_\Theta \mathbb{E}(L(\theta, \delta(X)))\pi(\theta)d\theta. \tag{1.9}$$

Definition 1.0.15. *A* ***Bayes estimator*** *is any estimator* $\delta(X)$ *which minimizes Bayes risk* $r_\pi(\delta)$.

Consider any two estimators δ_1 and δ_2 of θ.

Definition 1.0.16. *We say,* δ_1 ***is better than*** δ_2 *if*

$$R(\theta, \delta_1) \leq R(\theta, \delta_2), \forall \theta \in \Theta, \tag{1.10}$$

and $\exists \theta_0 \in \Theta$ *such that*

$$R(\theta_0, \delta_1) < R(\theta_0, \delta_2). \tag{1.11}$$

Definition 1.0.17. *An estimator* δ *is said to be* ***admissible*** *if there does not exist any other estimator which is better than* δ.

Chapter 2

Models for Multivariate Count data

Various different models for multivariate count data are available in the literature (see, for instance, Johnson, Kotz and Balkrishnan (2004) [16], Kocherlakota , S. and Kocherlakota, K (1992) [18]). In the literature, the more focus is given on constructing models for multivariate count data rather than parameter estimation problem due to computationally inefficient methods. In this chapter, we review different methods of constructing probability distribution models for multivariate count data.

2.1 Multivariate Count Data

The multivariate count data can be defined as counts of samples belonging to the different categories sampled from a population which is grouped in those categories. Suppose there are d categories. Each object in the population belongs to exactly one of the d categories. We are sampling from such populations and recording the counts in each category. The data for n such samples can be visualized as follows:

$$(m_{i1}, m_{i2}, \cdots, m_{id}; t_i), \quad i = 1, 2, \cdots, n, \quad t_i = \sum_{k=1}^{d} m_{ik},$$

where m_i represents the counts in the category i.

2.2 Construction pf multivariate discrete probability distributions

Johnson, Kotz and Balkrishnan (2004) have provided a book-length treatment on discrete multivariate distributions in [16] with the focus on strategies for the construction of multivariate discrete distributions. They also talk about parameter estimation problems for such models superficially. Due to complexity of the likelihood functions, not much improvement in terms of the estimation methods for such models has been noted in the literature. Kocherlakota , S. and Kocherlakota, K (1992) also provides a survey of generating methods for bivariate discrete distributions in [18]. Here, we provide a survey of some of these available methods of generating models for multivariate count data.

2.2.1 Mixing

If we have two or more multivariate discrete probability distributions with p.m.f. f_1 and f_2, the new distribution can be obtained by mixing as follows:

$$f(x_1, \dots, x_d) = \theta f_1(x_1, \dots, x_d) + (1-\theta) f_2(x_1, \dots, x_d)$$

for some $0 < \theta < 1$.

2.2.2 Compounding

Let $X_1 , \dots , X_d$ be discrete univariate random variables with p.m.f. $f_1(x_1|\theta_1) , \dots , f_d(x_d|\theta_d)$, respectively. Here, $\theta_1 , \dots , \theta_d$ are parameters associated with the f_i's.

$$f(x_1, \dots, x_d) = \int \cdots \int_{\theta_1 ,\dots ,\theta_d} f_1(x_1|\theta_1) \dots f_d(x_d|\theta_d) g(\theta_1 , \dots , \theta_d)\, d\theta_1 \cdots d\theta_d$$

where $g(\theta_1 , \dots , \theta_d)$ is the joint probability density of $\theta_1 , \dots , \theta_d$.

Example 3. *Let X_i be Poisson distributed random variables with means λ_i, for $i = 1, \dots, d$ i.e. $f(x_i) = \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}$. Suppose, $g(\lambda_1, \dots, \lambda_d)$ is a joint probability density for the λ_i's. Then,* *we can define*

$$f(x_i,\dots,x_d)=\underset{\lambda_1,\dots,\lambda_d}{\int\cdots\int} e^{-\sum\limits_i \lambda_i}\prod_i \frac{\lambda_i^{x_i}}{x_i!}\, g(\lambda_1,\dots,\lambda_d)\, d\lambda_1\dots d\lambda_d.$$

If the λ_i's are independent, then the X_i's are also independent. If λ_i's are independent and have gamma distribution, then the X_i's are also independent with a negative binomial distribution. Aitchison, et al. (1989) *considered a Poisson-log Normal model in which $g(\lambda_1,\dots,\lambda_d)$ is log-normal density in [2].*

2.2.3 Trivariate Reduction

Let $U_1,\dots,U_m$ be independent discrete univariate random variables. Then, we can construct $d-variate$ dependent discrete random variables as follows:

$$\begin{aligned} X_1 =& \tau_1(U_1,\dots,U_m)\\ &\vdots \\ X_d =& \tau_d(U_1,\dots,U_m)\end{aligned} \tag{2.1}$$

where τ_i's are functions from $\mathbb{N}^m \to \mathbb{N}$.

Example 4. *Let U_1,U_2,U_3 and U_4 be independent Poisson random variables with means $\lambda_1,\lambda_2,\lambda_3$ and λ_4, respectively. Suppose, X_i's are the linear combinations of U_i's are as follows:*

$$\begin{aligned} X_1 =& U_1+U_2+U_4\\ X_2 =& U_1+U_3+U_4\\ X_3 =& U_1+U_2+U_3\end{aligned} \tag{2.2}$$

Then, (X_1,X_2,X_3) has a multivariate discrete distribution with marginally distributed Poisson random variables with means $\lambda_1+\lambda_2+\lambda_4, \lambda_1+\lambda_3+\lambda_4$ and $\lambda_1+\lambda_2+\lambda_3$, respectively.

The covariance matrix is

$$\begin{bmatrix} \lambda_1+\lambda_2+\lambda_4 & \lambda_1+\lambda_4 & \lambda_1+\lambda_2 \\ \lambda_1+\lambda_4 & \lambda_1+\lambda_3+\lambda_4 & \lambda_1+\lambda_3 \\ \lambda_1+\lambda_2 & \lambda_1+\lambda_3 & \lambda_1+\lambda_2+\lambda_3 \end{bmatrix}$$

Note that all the correlations are positive.

Holgate ([13]) derived bivariate Poisson distribution using the method of trivariate reduction method of deriving bivariate distributions from three independent univariate discrete distributions. Let U_1, U_2, U_3 be independent univariate discrete random variables. Consider bivariate random variable (X, Y) obtained as follows:

$$(X, Y) = (U_1 + U_3, U_2 + U_3)$$

Hyunju Lee and Ji Hwan Cha ([20]) generalised this approach to generate two classes of bivariate discrete probability models based on minimum and maximum operator as

$$(X, Y) = (min(U_1, U_3), min(U_2, U_3))$$

and

$$(X, Y) = (max(U_1, U_3), max(U_1, U_3)),$$

respectively.

2.2.4 Copula Method

Definition 2.2.1. *A copula is a function $C : \boldsymbol{I}^n \to \boldsymbol{I}$ such that*

1. $C(0, \cdots, u_i, \cdots, 0) = 0, \forall u_i \in \boldsymbol{I}$
2. $C(1, \cdots, u_i, \cdots, 1) = u_i, \forall u_i \in \boldsymbol{I}$
3. *For any* $(u_1, \ldots, u_n)$ *and* $(v_1, \ldots, v_n) \in \boldsymbol{I}^n$ *such that* $u_i \leq v_i, \forall i \in \{1, 2, \ldots, n\}$,

$$\sum_{w_i \in u_i, v_i} (-1)^{\sum_{i=1}^{n} \boldsymbol{1}_{(v_i, w_i)}} C(w_1, \ldots, w_n) \geq 0,$$

$$\textit{where } \boldsymbol{1}_{(v_i,w_i)} = \begin{cases} 0 & if \;\; v_i = w_i \\ 1 & otherwise. \end{cases}$$

The following theorem by Sklar (1989) suggests the use of copula functions as links between the marginals and the joint distributions [28]. This enables to construct multivariate discrete distributions using univariate marginals and the copula function which provides a particular dependence structure.

Theorem 2.2.1 (Sklar's Theorem, 1959)**.** *Consider a joint distribution F with margins $F_1, \ldots, F_n$. Then there exist a copula C such that*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

If all the margins are continuous, then C is unique. In other cases, C is unique on $Range(F_1) \times \cdots \times Range(F_n)$.

Conversely, for any copula C and distribution functions F_i, $i = 1, 2, \ldots, n$, then F defined as

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

is a joint distribution function with margins F_i's.

Example 5. *Consider bivariate Frank's copula function*

$$C(u,v) = \frac{-1}{k} log \left(1 - \frac{(1 - e^{-ku})(1 - e^{-kv})}{(1 - e^{-k})} \right), \quad \forall (u,v) \in \boldsymbol{I}^2.$$

Then, a bivariate Negative binomial distribution with Frank's copula is given by

$$F_{X,Y}(x,y) = \frac{-1}{k} log \left(1 - \frac{\left(1 - e^{-k \sum\limits_{t=0}^{x} \frac{(a_1)_t}{t!} \theta_1^{a_1} (1-\theta_1)^t}\right) \left(1 - e^{-k \sum\limits_{t=0}^{y} \frac{(a_2)_t}{t!} \theta_2^{a_2} (1-\theta_2)^t}\right)}{(1 - e^{-k})} \right),$$

where $(x,y) \in \mathbb{N}^2$ and $-\infty < k < \infty$. This has marginal Negative binomial (a_i, θ_i) distributions and a dependence parameter k. The marginal distributions are independent for $k = 0$. Such models are considered by McHale, et al. in [27] used to model outcomes of the soccer matches.

Copula functions can be used to model any general dependency structure. But, it is well-known that all copula functions are bounded by Frechet bounds. In two-dimensions, the Frechet upper and lower bounds are both copula functions and we have, for any copula function C,

$$max(u+v-1,0) \leq C(u,v) \leq min(u,v), \quad \forall (u,v) \in \mathbf{I}^2 \tag{2.3}$$

This also gives bounds on the dependence structure induced by a copula.

Summary

As we have seen, all of the models generated using the above methods have restrictive covariance structure. Even for the bivariate cases, the whole range of correlation may not be covered using these models. Hence, this motivates us to find new strategies to generate multivariate discrete distributions. In Chapter 3, we study a novel strategy which, at least for bivariate cases, leads to models with both positive and negative correlation structures.

Chapter 3

Multivariate Sums and Share Model

This thesis builts on the multivariate sums and shares model proposed by Jones and Marchand. In the article [17], they introduce this model and provide various interesting statistical properties of the model. They also mention the connections of this model with the models available in the literature. In particular, they introduce a Negative binomial sums and Polya shares model. In this chapter, we summarize the results obtained by Jones and Marchand. This chapter also introduces the parameter estimation problem for the Negative binomial sums and Polya shares model.

3.1 Introduction

In the article by Jones and Marchand [17], they propose a novel strategy to generate or construct multivariate discrete distributions via sums and shares model. Consider $M = (M_1, \ldots, M_d)$ be d-variate discrete random variable. We denote $T = \sum_{i=1}^{d} M_i$ be the sum of the counts M_i. The sums and shares strategy consists of the following two steps:

1. First, model the sum of the counts T by the distribution with p.m.f. $p_T(t|\Theta_1)$ and conditioning on T, model the distribution of counts $M' = (M_1, M_2, \ldots, M_{d-1})$ in different categories as $M'|T = t$ having the distribution with p.m.f. $b_{[t]}(m_1, m_2, \ldots, m_{d-1}|\Theta_2)$. So that, using the definition of the conditional distribution, the resulting distribution

has probability mass function

$$p(M = (m_1, \dots, m_d)|\Theta_1, \Theta_2) = \quad p_T(t|\Theta_1) b_{[t]}(m_1, \dots, m_{d-1}|\Theta_2) \tag{3.1}$$

Here, Θ_1 and Θ_2 denotes the parameters involved in the probability mass functions p_T and $b_{[t]}$, respectively.

2. Then, averaging this mixing over the distributions of the parameters involved i.e. Θ_1, Θ_2. Let $F(\Theta_1, \Theta_2)$ be the joint distribution function for the parameters Θ_1 and Θ_2. We get the final distribution as follows:

$$p(M = (m_1, \dots, m_d)) = \int p(M = (m_1, \dots, m_d)|\Theta_1, \Theta_2) dF(\Theta_1, \Theta_2) \tag{3.2}$$

With different choices for the distributions of total sums T, shares $(M'|T)$ and parameters (Θ_1, Θ_2), this strategy gives rise to many interesting models for the multivariate count data.

3.2 Construction of the Model

Natural choices the distribution of sums and shares of counts are a Poisson distribution for the sum T and the multinomial distribution for the shares *i.e.*

$$p(T = t|\Lambda) = \frac{e^{-\Lambda}\Lambda^t}{t!}, \quad \Lambda > 0 \tag{3.3}$$

and

$$b_{[t]}(M = (m_1, \dots, m_{d-1})|t, U_1, \dots, U_{d-1}) = \frac{t!}{m_1! m_2! \dots m_d!} U_1^{m_1} \cdots U_d^{m_d} \tag{3.4}$$

where $t = m_1 + \cdots + m_d$ and $U_i \geq 0, \quad U_d = 1 - U_1 - \cdots - U_{d-1}$.

Then we get the p.m.f.,

$$\begin{aligned} p(m_1, \dots, m_d) = & \int \frac{e^{-\Lambda}\Lambda^{m_1+\cdots+m_d}}{m_1! m_2! \cdots m_d!} U_1^{m_1} \cdots U_d^{m_d} dF(\Lambda, U_1, \cdots, U_{d-1}) \\ = & \int \prod_{i=1}^{d} \frac{(\Lambda U_i)^{m_i} e^{-\Lambda U_i}}{m_i!} dF(\Lambda, U_1, \cdots, U_{d-1}) \end{aligned} \tag{3.5}$$

We obtain the following results related to marginal distributions of shares M_i and expectations of Λ and ΛU_i's conditioned on the observed value of $M = (m_1, \cdots, m_d)$.

Proposition 3.2.1. *The marginal distributions for M_i's are*

$$p(m_i) = \int \frac{(\Lambda U_i)^{m_i} e^{-\Lambda U_i}}{m_i!} dF(\Lambda, U_1, \cdots, U_{d-1}) \tag{3.6}$$

Proof. The marginal distribution of M_i is given as

$$\begin{aligned}
p(m_i) &= \sum_{m_j; j \neq i} p(m_1, \ldots, m_d) \\
&= \sum_{m_j; j \neq i} \int \prod_{k=1}^{d} \frac{(\Lambda U_k)^{m_k} e^{-\Lambda U_k}}{m_k!} dF(\Lambda, U_1, \cdots, U_{d-1}) \\
&= \int \sum_{m_j; j \neq i} \prod_{k=1}^{d} \frac{(\Lambda U_k)^{m_k} e^{-\Lambda U_k}}{m_k!} dF(\Lambda, U_1, \cdots, U_{d-1}) \quad \cdots (Fubini's\ theorem) \\
&= \int \frac{(\Lambda U_i)^{m_i} e^{-\Lambda U_i}}{m_i!} \prod_{k=1, k \neq i}^{d} \left(\sum_{m_k} \frac{(\Lambda U_k)^{m_k} e^{-\Lambda U_k}}{m_k!} \right) dF(\Lambda, U_1, \cdots, U_{d-1}) \\
&= \int \frac{(\Lambda U_i)^{m_i} e^{-\Lambda U_i}}{m_i!} dF(\Lambda, U_1, \cdots, U_{d-1}).
\end{aligned}$$

□

Proposition 3.2.2. *The conditional expectations of Λ and ΛU_i conditioned on $(m_1, \ldots, m_d)$ can be expressed recursively as follows:*

$$\mathbb{E}[\Lambda U_i | m_1, \ldots, m_d] = \frac{(m_i + 1) p(m_1, \cdots, m_i + 1, \cdots, m_d)}{p(m_1, \ldots, m_d)} \tag{3.7}$$

and

$$\mathbb{E}[\Lambda | m_1, \ldots, m_d] = \sum_{i=1}^{d} \frac{(m_i + 1) p(m_1, \cdots, m_i + 1, \cdots, m_d)}{p(m_1, \ldots, m_d)}. \tag{3.8}$$

Proof. Using Bayes rule, we get,

$$f(\lambda, u_1, \cdots, u_d | m_1, \cdots, m_d) = \frac{p(m_1, \ldots, m_d | \lambda, u_1, \cdots, u_d) dF(\lambda, u_1, \cdots, u_d)}{p(m_1, \ldots, m_d)}$$

where

$$p(m_1, \cdots, m_d | \lambda, u_1, \cdots, u_d) = \prod_{k=1}^{d} \frac{(\lambda u_k)^{m_k} e^{-\lambda u_k}}{m_k!}$$

Hence,

$$\begin{aligned}
\mathbb{E}[\Lambda U_i | m_1, \cdots, m_d] &= \frac{1}{p(m_1, \cdots, m_d)} \int \lambda u_i p(m_1, \cdots, m_d | \lambda, u_1, \cdots, u_d) dF(\lambda, u_1, \cdots, u_d) \\
&= \frac{1}{p(m_1, \cdots, m_d)} \int \lambda u_i \prod_{k=1}^{d} \frac{(\lambda u_k)^{m_k} e^{-\lambda u_k}}{m_k!} dF(\lambda, u_1, \cdots, u_d) \\
&= \frac{m_i + 1}{p(m_1, \cdots, m_d)} \int \frac{(\lambda u_i)^{m_i+1} e^{-\lambda u_i}}{(m_i+1)!} \prod_{k=1, k\neq i}^{d} \frac{(\lambda u_k)^{m_k} e^{-\lambda u_k}}{m_k!} dF(\lambda, u_1, \cdots, u_d) \\
&= \frac{(m_i+1) p(m_1, \cdots, m_i + 1, \cdots, m_d)}{p(m_1, \cdots, m_d)}.
\end{aligned}$$

Summing over all i's, we obtain the expression 3.8. □

If we assume that Λ and $U_i's$ are distributed independently, *i.e.* the distributions L and H of Λ and $U_1, \ldots, U_{d-1}$ are independent. We get the resulting distribution as follows:

$$\begin{aligned}
p(m_1, m_2, \cdots, m_d) = \int \cdots \int_{0 \leq U_1 + \cdots + U_{d-1} \leq 1} \frac{(m_1 + m_2 + \cdots + m_d)!}{m_1! m_2! \cdots m_d!} U_1^{m_1} \cdots (U_d)^{m_d} dH(U_1, \cdots, U_{d-1}) \\
\times \int_0^\infty \frac{e^{-\Lambda} \Lambda^{m_1 + \cdots + m_d}}{(m_1 + \cdots + m_d)!} dL(\Lambda) \quad (3.9)
\end{aligned}$$

By change of variables $R_i = \Lambda U_i$, we get

$$\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{d-1} \\ \Lambda \end{bmatrix} = \begin{bmatrix} \frac{1}{\Lambda} & 0 & \cdots\cdots & 0 \\ 0 & \frac{1}{\Lambda} & \ddots & \vdots \\ \vdots & & \ddots \; 0 & \vdots \\ 0 & \cdots\cdots & \frac{1}{\Lambda} & 0 \\ 1 & \cdots\cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_{d-1} \\ R_d \end{bmatrix} \quad (3.10)$$

Hence, the Jacobian of the transformation is $\left(\frac{1}{\Lambda}\right)^{d-1}$. Thus, we have the joint density function

$$f(r_1, r_2, \ldots, r_d) = \prod_{k=1}^{d} \frac{r_i^{m_i} e^{-r_i}}{m_i!} \frac{1}{\left(\sum_{i=1}^{d} r_i\right)^{d-1}} dH \left(\frac{r_1}{\sum_{i=1}^{d} r_i}, \ldots, \frac{r_{d-1}}{\sum_{i=1}^{d} r_i} \right) dL \left(\sum_{i=1}^{d} r_i \right). \tag{3.11}$$

Now, moments of M_i are given as

$$\mathbb{E}M_i = \mathbb{E}(\Lambda U_i) = \mathbb{E}(\Lambda)\mathbb{E}(U_i) \tag{3.12}$$

$$\mathbb{V}(M_i) = \mathbb{E}(\Lambda^2)\mathbb{V}(U_i) + \mathbb{V}(\Lambda)(\mathbb{E}(U_i))^2 + \mathbb{E}(\Lambda)\mathbb{E}(U_i) \tag{3.13}$$

$$Cov(M_i, M_j) = \mathbb{V}(\Lambda)\mathbb{E}(U_i U_j) + (\mathbb{E}(\Lambda))^2 Cov(U_i, U_j) \tag{3.14}$$

From (3.12) and (3.13), we can see that the model is inherently overdispersed *i.e.*

$$\mathbb{V}(M_i) > \mathbb{E}(M_i).$$

Using the Law of total covariance and $Cov(M_i, M_j | R_1, \ldots, R_d) = 0$, we get

$$\begin{aligned} Cov(M_i, M_j) = & \ \mathbb{E}[Cov(M_i, M_j | R_1, \cdots, R_d)] \\ & + \ Cov(\mathbb{E}[M_i | R_1, \cdots, R_d], \mathbb{E}[M_j | R_1, \cdots, R_d]) \\ = & \ Cov(R_i, R_j) \end{aligned} \tag{3.15}$$

3.3 Negative Binomial sums and Polya shares model

3.3.1 Definition and Construction

Now, if we consider H to be $Dirichlet(\alpha_1, \ldots, \alpha_d)$ and L as $Gamma(a, b)$, where $\alpha_1, \ldots, \alpha_d, a, b > 0$ in (3.9), *i.e.*, the probability density functions for Λ and U_i's are, respectively,

$$\ell(\lambda | a, b) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \tag{3.16}$$

and

$$h(u_1,\ldots,u_{d-1}) = \frac{\Gamma(\alpha)\prod_{i=1}^{d} u_i^{\alpha_i-1}}{\prod_{i=1}^{d}\Gamma(\alpha_i)}, \tag{3.17}$$

where $\alpha = \sum_{i=1}^{d}\alpha_i$.

We get the resulting distribution as:

$$p(m_1,\ldots,m_d) = \frac{\prod_{i=1}^{d}(\alpha_i)_{m_i}}{(\alpha)_{m_1+\cdots+m_d}}\frac{(a)_{m_1+\cdots+m_d}}{m_1!\cdots m_d!}\theta^a\,(1-\theta)^{m_1+\cdots+m_d}, \tag{3.18}$$

where $\theta = \frac{b}{1+b}$.

This is the p.m.f. for Negative Binomial sums and Polya shares distribution.

3.3.2 Moments

Moments of $M_i's$ for this model can be obtained using (3.12) as:

$$\mathbb{E}[M_i] = \frac{a\alpha_i}{b\alpha} = \frac{a(1-\theta)}{\theta}\frac{\alpha_i}{\alpha}, \tag{3.19}$$

and defining $Cov(M)$ as the covariance matrix of $(M_1,\ldots,M_d)$ with the entries $Cov(M_i,M_j)$, we obtain from (3.13), and (3.14):

$$Cov(M) = \frac{a}{b^2\alpha^2(1+\alpha)}\{(\alpha-a)\boldsymbol{\alpha}^t\boldsymbol{\alpha} + (\alpha(a+1+(1+\alpha)b))\boldsymbol{I}_d\} \tag{3.20}$$

where $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_d)$ and $\boldsymbol{I}_d$ is $d\times d$ identity matrix.

When $d=2$,

$$Cov(M_1,M_2) = \frac{a\alpha_1\alpha_2}{b^2(1+\alpha)\alpha^2}(\alpha-a),$$

and the correlation $\rho(M_1,M_2)$ between M_1 and M_2 is given as:

$$\rho(M_1,M_2) = \frac{\alpha_1\alpha_2(\alpha-a)}{\sqrt{\alpha^2(1+a+b+b\alpha)^2+(\alpha-a)^2\alpha_1\alpha_2+\alpha^2(\alpha-a)(1+a+b+b\alpha)}}$$

or

$$\rho(M_1, M_2) = \frac{\alpha f_1 f_2(\alpha - a)}{\sqrt{(1+a+b+b\alpha)^2 + (\alpha-a)^2 f_1 f_2 + (\alpha - a)(1+a+b+b\alpha)}},$$

where $f_1 = \frac{\alpha_1}{\alpha}$ and $f_2 = \frac{\alpha_2}{\alpha}$.

Hence,

1) if $\alpha > a$, $\rho(M_1, M_2) > 0$ *2)* if $\alpha < a$, $\rho(M_1, M_2) < 0$.

Hence, for bivariate cases, both positive and negative correlation are possible.

When $d > 2$, all correlations are either positive or negative.

3.3.3 Relations to other distributions

Many available distributions can be recovered as special cases of the Negative Binomial sums and Polya shares model. Jones and Marchand has provided connections to the following well-known discrete distributions:

Bivariate Bailey distribution

The expression (3.18) can be considered as the multivariate extension of bivariate Bailey distribution defined by Laurent [19]. The p.m.f. for Bailey distribution is given as

$$Bailey(m_1, m_2|\alpha_1, \alpha_2, \theta, a) = \frac{(\alpha_1)_{m_1}(\alpha_2)_{m_2}}{(\alpha_1+\alpha_2)_{m_1+m_2}} \frac{(a)_{m_1+m_2}}{m_1!m_2!} \theta^a (1-\theta)^{m_1+m_2}. \tag{3.21}$$

Schur Constant distribution

With $\alpha_1 = \cdots = \alpha_d = 1$, we get

$$p(m_1, \ldots, m_d) = \frac{(a)_{m_1+\cdots+m_d}}{(d)_{m_1+\cdots+m_d}} \theta^a (1-\theta)^{m_1+\cdots+m_d} \tag{3.22}$$

This is discrete Schur Constant distribution. Castaner, *et al.* [8] defined discrete Schur constant distributions as follows: A $d-variate$ random variable $X=(X_1,\ldots,X_d)$, is said to have discrete joint Schur Constant survival distribution if

$$P(X_1 \geq x_1,\ldots,X_d \geq x_d) = S(x_1+\cdots+x_d)$$

i.e., the survival function of the X depends on the argument $(X_1,\ldots,X_d)$ only through the sum $\sum_{i=1}^d X_i$.

Limiting model

Let $\alpha_1,\cdots,\alpha_d \to \infty$ such that $\frac{\alpha_i}{\alpha} \to \phi_i$. Then we have,

$$\frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{(\alpha)_{m_1+\cdots+m_d}} = \frac{\prod_{i=1}^d \alpha_i \cdots (\alpha_i+m_i-1)}{\alpha\cdots(\alpha+m_1+\cdots+m_d-1)} \to \frac{\prod_{i=1}^d \alpha_i^{m_i}}{\alpha^{m_1+\cdots+m_d}} = \prod_{i=1}^d \phi_i^{m_i},$$

and thus

$$p(m_1,\ldots,m_d) = \frac{(a)_{m_1+\cdots+m_d}}{m_1!\cdots m_d!}\left(\prod_{i=1}^d \phi_i^{m_i}\right)\theta^a\,(1-\theta)^{m_1+\cdots+m_d}\,. \tag{3.23}$$

The correlation for any two components M_i and M_j is given as

$$\rho(M_i,M_j) = (1-\theta)\sqrt{\frac{\phi_i}{(1-\theta)\phi_i+\theta}\,\frac{\phi_j}{(1-\theta)\phi_j+\theta}} > 0.$$

Multivariate Discrete Liouville Distribution

Jones and Marchand also studied the following model in great detail. They obtain very interesting results about marginal distributions and bounds on variances of the shares.

Assuming a general distribution $p_T(t)$ for T while $M^{'}|(T=t)$ is distributed as Dirichlet-Multinomial as before, we get

$$p(m_1,\ldots,m_d) = \frac{(m_1+\cdots+m_d)!}{m_1!\cdots m_d!}\frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{(\alpha)_{m_1+\cdots+m_d}} p_T(m_1+\cdots+m_d) \tag{3.24}$$

which can be rewritten as

$$\begin{aligned} p(m_1,\dots,m_d) =& \frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{m_1!\cdots m_d!}\frac{(m_1+\cdots+m_d)!}{(\alpha)_{m_1+\cdots+m_d}} p_T(m_1+\cdots+m_d) \\ =& \frac{\prod_{i=1}^d (\alpha_i)_{m_i}}{m_1!\cdots m_d!}\mathscr{F}(m_1+\cdots+m_d), \end{aligned} \tag{3.25}$$

where $\mathscr{F}(t) = \frac{t!}{(\alpha)_t}p_T(t)$. This is the discrete analogue of the continuous multivariate discrete Liouville distribution given by Lingappaiah [25].

As done in the previous section, considering $\alpha_1 = \cdots = \alpha_d = 1$, we get the p.m.f. as:

$$p(m_1,\dots,m_d) = \frac{(m_1+\cdots+m_d)!}{(d)_{m_1+\cdots+m_d}} p_T(m_1+\cdots+m_d) = \frac{p_T(m_1+\cdots+m_d)}{\binom{m_1+\cdots+m_d+d-1}{d-1}} \tag{3.26}$$

with above p.m.f., the multivariate marginals are obtained as

$$\begin{aligned} p_S(m_{i_1},\cdots,m_{i_k}) =& \sum_{t\geq m_{i_1}+\cdots+m_{i_k}} \frac{p_T(t)t!}{(d)_t}\binom{t-(m_{i_1}+\cdots+m_{i_k})+d-k-1}{d-k-1} \\ =& \quad (d-1)\cdots(d-k)\sum_{t\geq m_{i_1}+\cdots+m_{i_k}} p_T(t)\frac{(t-(m_{i_1}+\cdots+m_{i_k})+1)_{d-k-1}}{(t+1)_{d-1}}. \end{aligned}$$

Also, using the inclusion-exclusion principle, one obtains the following relationship:

$$p(m_1,\dots,m_d) = \sum_{j=0}^{d-1}(-1)^j\binom{d-1}{j}p_S(m_1+\cdots+m_d+j),$$

previously studied in the bivariate case ($d=2$) by Aoudia, *et al.* [4]. For the total sum T, we have

$$p_T(t) = \binom{t+d-1}{d-1}\sum_{j=0}^{d-1}(-1)^j\binom{d-1}{j}p_S(t+j).$$

Besides, we also have the factorial moments given as follows:

$$\mathbb{E}\left\{\prod_{i=1}^d\binom{M_i}{k_i}\right\} = \frac{\mathbb{E}\left\{\binom{T}{k_1+\cdots+k_d}\right\}}{\binom{k_1+\cdots+k_d+d-1}{d-1}}. \tag{3.27}$$

This proof uses Gould's identity which can be proved easily by a combinatorial argument (choosing k_1+k_2+1 objects out of $n+1$ objects in a special way gives the Gould's identity.) Identity (3.27) gives

$$\mathbb{E}(M_i) = \frac{\mathbb{E}(T)}{d} \quad , \mathbb{V}(M_i) = 2\frac{\mathbb{E}(T^2)-\mathbb{E}(T)}{d(d+1)} + \frac{\mathbb{E}(T)}{d} - \frac{(\mathbb{E}(T))^2}{d^2},$$

$$Cov(M_i, M_j) = \mathbb{E}\left\{\binom{M_i}{1}\binom{M_j}{1}\right\} - \mathbb{E}\left\{\binom{M_i}{1}\right\}\mathbb{E}\left\{\binom{M_j}{1}\right\} = \frac{\mathbb{E}(T^2-T)}{d(d+1)} - \frac{(\mathbb{E}(T))^2}{d^2},$$

or, again,

$$Cov(M_i, M_j) = \frac{1}{2}\left(Var(M_i) - \{\mathbb{E}(M_i)\}^2 - \mathbb{E}(M_i)\right).$$

Also,

$$\rho(M_i, M_j) = \frac{1}{2}\left\{1 - \frac{\mathbb{E}(M_i)^2 + \mathbb{E}(M_i)}{\mathbb{V}(M_i)}\right\} < \frac{1}{2}.$$

Since $\rho(M_i, M_j) \geq -1$, for the distribution with p.m.f. (3.26), we also obtain the inequality

$$\mathbb{V}(M_i) \geq \frac{1}{3}\mathbb{E}(M_i)\{\mathbb{E}(M_i)+1\}. \tag{3.28}$$

Johnson, *et al.* showed that for unimodal continuous distribution X, $\mathbb{V}(X) \geq \frac{1}{3}(\mathbb{E}(X))^2$ holds [14] . Inequality (3.28) is a discrete analogue of this result.

3.4 Parameter estimation problem

The main goal of the thesis is to provide methods to estimate the parameters for probability mass function (3.18), namely for negative binomial sums and Polya shares model obtained in Section 3.3. The pmf is

$$p(m_1, \ldots, m_d) = \frac{\prod_{i=1}^{d}(\alpha_i)_{m_i}}{(\alpha)_{m_1+\cdots+m_d}} \frac{(a)_{m_1+\cdots+m_d}}{m_1!\ldots m_d!}\theta^a\,(1-\theta)^{m_1+\cdots+m_d} \tag{3.29}$$

Here the parameters to be determined are $a\,,\theta\,\alpha_1\,,\cdots,\alpha_d$. The p.m.f. (3.29) can be rewritten as

$$p(m_1,\ldots,m_d) = \frac{(m_1+\cdots+m_d)!}{m_1!\ldots m_d!}\frac{\prod_{i=1}^{d}(\alpha_i)_{m_i}}{(\alpha)_{m_1+\cdots+m_d}}\frac{(a)_{m_1+\cdots+m_d}}{(m_1+\cdots+m_d)!}\theta^a\,(1-\theta)^{m_1+\cdots+m_d} \quad (3.30)$$

Hence, alternatively, model (3.29) can be decomposed as:

$$M^{'}|(T=t) \sim Polya(t,\alpha_1,\ldots,\alpha_d)$$

and

$$T \sim NegativeBinomial(a,\theta).$$

Hence, we study the problem of the parameter estimation for the model in two steps:

1. Parameter estimation for the distribution of sums - Negative Binomial distribution.

2. Parameter estimation for shares - Polya distribution.

Then, we combine both to obtain set of estimators for the parameters involved in the p.m.f. (3.29). In Chapter 4, we discuss the parameter estimation for the negative binomial distribution. Later, in Chapter 5, we discuss the parameter estimation for the Polya distribution.

Chapter 4

Parameter Estimation for the Negative Binomial Distribution

In this chapter, we study the problem of parameter estimation for the negative binomial distribution when both parameters are unknown. Here, we consider method of moments estimators, maximum likelihood estimators and Bayes estimators. We propose a family of distributions which is semiconjugate for the two-parameter unknown negative binomial distribution. Simulation studies are performed to compare the performance of the different estimators obtained.

4.1 Introduction

The negative binomial distribution is a well-known univariate discrete distribution. It has been applied to model diverse count-data generating processes. We will refer to a Negative binomial distribution T with parameters a and θ, if the probability mass function is given as

$$p(T = t|a, \theta) = \frac{(a)_t}{t!} \theta^a (1 - \theta)^t \, \mathbb{I}_{\mathbb{N}}(t) \ , \tag{4.1}$$

where $a > 0$ and $0 < \theta < 1$. We denote this as $T \sim NB(a, \theta)$. The mean and variance of

T are given as follows:

$$\mu = \mathbb{E}[T] = \frac{a(1-\theta)}{\theta}, \quad \sigma^2 = \mathbb{V}[T] = \frac{a(1-\theta)}{\theta^2}.$$

The moment generating function for the negative binomial distribution with p.m.f. (4.1) is

$$\mathbb{M}_X(s) = \frac{\theta^a}{[1-(1-\theta)e^s]^a} \tag{4.2}$$

Consider $t_1, t_2, \ldots, t_n$ be n independent observations from Negative binomial $NB(a, \theta)$. In this chapter, we consider the problem of estimating both the parameters a and θ based on these n observations. In the section 4.3 and 4.4, we summarize some available results in the literature related to method of moments and maximum likelihood estimators. There is not much work done on Bayesian inference for such distribution. There are some studies in the literature regarding Bayesian inference for negative binomial distribution. Bradlow, *et al.* (2002) have suggested Bayesian inference for the Negative Binomial Distribution via Polynomial Expansions [7]. Lio (2009) elaborates on Bayesian approach for parameters estimation in case of Negative binomial distribution using MCMC techniques [26]. In this Chapter, we provide a more general setup of Bayesian inference for the negative binomial distribution. We suggest prior choices for the distribution of a and θ which forms semiconjugate family of priors for the two parameter negative binomial distribution. We also derive some results related to posterior distributions and predictive density distributions. Simulation studies are conducted to compare the performances of the estimators under consideration. All the methods developed are illustrated using numerical examples. Applications to some real datasets are also provided.

4.2 Unbiasedness of the Estimators

If the size parameter a is known, we get an unbiased estimator for θ. But, when a is unknown, no estimator for a is unbiased. These results are reported in the following theorems:

Theorem 4.2.1. *If $a > 1$ is known, then the estimator*

$$\delta(t_1, \cdots, t_n) = \frac{na - 1}{na + \sum_{i=1}^{n} t_i - 1} \tag{4.3}$$

of θ is an unbiased estimator of θ.

Proof. If a is known, $T = \sum_{i=1}^{n} t_i$ is sufficient statistics and is distributed as $NB(na, \theta)$. Now,

$$\begin{aligned}
\mathbb{E}\left(\frac{T}{na + T - 1}\right) &= \sum_{T \geq 0} \left(\frac{T}{na + T - 1}\right) \frac{(na)_T}{T!} \theta^{na} (1-\theta)^T \\
&= \sum_{T \geq 1} \frac{(na)_{(T-1)}}{(T-1)!} \theta^{na} (1-\theta)^T \\
&= (1-\theta) \sum_{T \geq 0} \frac{(na)_T}{T!} \theta^{na} (1-\theta)^T \\
&= 1 - \theta.
\end{aligned}$$

Hence,

$$\mathbb{E}\left(\frac{na - 1}{na + T - 1}\right) = \theta$$

□

Theorem 4.2.2. *(Wang [34]) There does not exist an unbiased estimator of a.*

Proof. Suppose there exist an unbiased estimator of a, say $\delta(T)$, where $T = (t_1, \cdots, t_n)$. Then we have,

$$\sum_{T} \left[\prod_{i=1}^{n} \frac{(a)_{t_i}}{t_i!}\right] \theta^{na} (1-\theta)^{\sum t_i} \delta(T) = a \quad , \forall a \in (0, \infty) \text{ and } \forall \theta \in (0,1) \tag{4.4}$$

Now, consider $a = 1$ and $\theta \to 1$, we get $\delta(0, \cdots, 0) = 1$. Also, for $a = 2$ and $\theta \to 1$, we get $\delta(0, \cdots, 0) = 2$. Thus we get the contradiction. Hence, no estimator of a is unbiased. □

4.3 Method of moments estimators

Fisher (1941), in [11], has provided an elaborate discussion on the method of moments estimators for Negative binomial distribution as well as their efficiencies. Dropkin (1959) used method of moments to fit negative binomial model to data consisting of number of accidents [9]. Savani, *et al.* (2006) have also discussed the method of moment estimators and their asymptotic properties [31].

Comparing first sample moment $\left(\hat{\mu_1} = \frac{\sum_{i=1}^n t_i}{n}\right)$ and second sample moment $\left(\hat{\mu_2} = \frac{\sum_{i=1}^n t_i^2}{n}\right)$ with their theoretical counterparts, we get point estimates for a and θ as follows:

$$\hat{a} = \frac{\hat{\mu_1}^2}{\hat{\mu_2} - \hat{\mu_1} - \hat{\mu_1}^2}, \quad \hat{\theta} = \frac{\hat{\mu_1}}{\hat{\mu_2} - \hat{\mu_1}^2}. \tag{4.5}$$

This estimators may lie outside of the parameter space. Note that these estimators are valid only if the data is overdispersed *i.e.* sample variance is greater than sample mean $(\hat{\mu_2} - \hat{\mu_1}^2 > \hat{\mu_1})$. This issue is explored by Nkingi and Vrbik (2016) [29]. They have talked about the probability of the occurrence of under dispersed samples (sample variance < sample mean) and constructed the confidence regions for the method of moments estimators. According to the simulation study, when θ is large the percentage of occurrence of under dispersion is very high. But the trend is more evident (Refer Table 4.1 and Table 4.2) with the following parametrizations used by Nkingi and Vrbik [29]:

$$\mu = \frac{a(1-\theta)}{\theta}, \qquad P = \frac{1}{\theta} - 1.$$

$\mu \backslash P$	0.1	0.3	1	10
0.1	0.69	0.60	0.41	0.12
0.3	0.55	0.42	0.27	0.16
1	0.45	0.28	0.08	0.04
10	0.42	0.26	0.04	0.00

Table 4.1: Fraction of samples with under dispersion, sample size $N = 20$ and number of samples 10000.

$\mu \backslash P$	0.1	0.3	1	10
0.1	0.46	0.27	0.12	0.05
0.3	0.31	0.10	0.03	0.01
1	0.29	0.06	0.02	0.01
10	0.27	0.04	0.01	0.01

Table 4.2: Fraction of samples with under dispersion, sample size $N = 100$ and number of samples 10000.

4.4 Maximum likelihood estimators

The log-likelihood function is given by

$$logL(a,\theta|\boldsymbol{t}) = nalog(\theta) + \sum_{i=1}^{n} t_i log(1-\theta) + \sum_{i=1}^{n}\left(\sum_{j=0}^{t_i-1} log(a+j)\right) - \sum_{i=1}^{n} log(t_i!) \tag{4.6}$$

Johnson, *et al.* (2005) have discussed the method of maximum likelihood estimators for Negative Binomial distribution with both parameters unknown (page 216 of [15]). A closed form expression of the MLE is not easy to obtain. But, MLE's of a and θ are given as a solution to the following system of equations:

$$\hat{a} = \frac{\sum_{i=1}^{n} t_i}{n}\frac{\hat{\theta}}{1-\hat{\theta}} = \hat{\mu}_1 \frac{\hat{\theta}}{1-\hat{\theta}},$$

$$log\hat{\theta} = -\sum_{j=0}^{max\{t_j\}} \frac{1}{\hat{a}+j} \sum_{i=j+1}^{max\{t_j\}} f_i,$$

where f_i is the observed frequency of i. We can use numerical root finding methods such as the secant method or bisection method to find a solution for the above system. Another way is to use gradient accent algorithm for the log-likelihood function.

It is seen that MLE estimator for a may not be exist as the likelihood function may be increasing as $a \to \infty$. Anscombe (1950) commented on the existence of MLE in [3]. He has argued that in case of overdispersion, you have at least one finite root to the above mentioned system of equations. Proof for the argument of having at least one root when sample is overdispersed is given by Willson, *et al.* [35], Levin and Reeds [23] and Aragon, *et al.* [5]. Willson, *et al.*(1986) have also shown that the order statistic is minimal sufficient statistic but not complete for negative binomial distribution when the sample size is greater than three [35] . Aragon, *et al.* (1992) have shown that MLE exists iff sample is overdispersed. Simonsen has also proved the same theorem in [32]. We report this result in the following theorem and provide a proof by Aragaon, *et al.* [5] :

Theorem 4.4.1. *Maximum likelihood estimator for a and θ exists if and only if the sample is overdispersed.*

Proof. The maximum likelihood estimator of a is a root of the score function

$$g(a) = \sum_{j=0}^{max\{t_j\}} \frac{1}{\hat{a}+j} \sum_{i=j+1}^{max\{t_j\}} f_i - log\left(1+\frac{\hat{\mu}_1}{a}\right)$$

and that of θ is $\theta = \frac{\hat{a}}{\hat{a}+\bar{t}}$. Defining the survival functions as $F_j = \sum_{i=j}^{max\{t_j\}} f_i$, we get,

$$g(a) = \sum_{j=1}^{max\{t_j\}} \frac{1}{a+j-1} F_j - log\left(1+\frac{\hat{\mu}_1}{a}\right)$$

Note that $a.g(a) \to F_1$ as $a \to 0$. Using change of variable, $z = \frac{1}{a}$, we get,

$$G(z) = g\left(\frac{1}{a}\right) = \sum_{j=1}^{max\{t_j\}} \frac{z}{1+(j-1)z} F_j - log\left(1+z\hat{\mu}_1\right)$$

As noted earlier, $\frac{G(z)}{z} \to F_1 > 0$ as $z \to \infty$.

On differentiation, we get,

$$G^{'}(z) = \sum_{j=1}^{max\{t_j\}} \frac{1}{1+(j-1)z} F_j - \sum_{j=1}^{max\{t_j\}} \frac{z(j-1)}{(1+(j-1)z)^2} F_j - \frac{\hat{\mu}_1}{1+z\bar{t}}$$

and

$$G^{''}(z) = -2\sum_{j=1}^{max\{t_j\}} \frac{(j-1)}{(1+(j-1)z)^2} F_j + 2\sum_{j=1}^{max\{t_j\}} \frac{z(j-1)^2}{(1+(j-1)z)^3} F_j + \frac{\hat{\mu}_1^2}{(1+z\hat{\mu}_1)^2}$$

Note that $G(0) = G^{'}(0) = 0$ and

$$G^{''}(0) = -2 \sum_{j=1}^{max\{t_j\}} (j-1)F_j + \hat{\mu}_1^2 = \hat{\mu}_1 - (\hat{\mu_2} - \hat{\mu_1}^2)$$

If the sample is overdispersed *i.e.* $(\hat{\mu_2} - \hat{\mu_1}^2) > \hat{\mu}_1$, $G^{''}(0) < 0$. Hence, G will have the solution. □

Moreover, we have the following result by Levis and Reeds [23]:

Theorem 4.4.2. *(Levis and Reeds [23]) If maximum likelihood estimator of a and θ exists, then it is unique.*

4.5 Bayesian inference

In Section 4.3 and Section 4.4, we noted that there are limitations of method of moment estimators and maximum likelihood estimators in case of underdispersed data. We have developed Bayesian setup to overcome this parameter estimation problem. Bradlow, *et al.* (2002) have suggested Bayesian inference for the Negative Binomial Distribution via Polynomial Expansions [7]. They have used polynomial approximations for ratios of gamma functions appearing in the posterior distribution. With beta prime prior for success probability θ and Pearson Type VI prior for size parameter a, they provided closed from approximations to the posterior moments of a and θ. Lio (2009) elaborates on Bayesian approach for parameters estimation in case of Negative binomial distribution [26].

Besides, in the literature, we can find Bayesian model for negative binomial distribution in case of single unknown parameter (θ) using beta prior which is conjugate prior. We propose a Bayesian model when both the parameters are unknown. Also, the prior choice is shown to be semiconjugate.

4.6 Priors for a and θ

We consider the prior of the form $\pi(a,\theta) = f(a)g(\theta)$, where $f(a)$ is *Gammapoly* density defined as follows:

4.6.1 Gammapoly density

Gamma distribution is widely studied and has been used to model continuous data in the literature. We propose a new density function for a of the following form which we call *Gammapoly* density function with shape parameter γ:

$$f(a) \propto \Phi(a) exp(-a\gamma),$$

where $\Phi(a)$ is a polynomial in a, say, $\Phi(a) = c_0 + c_1 a + \cdots + c_d a^d$ for some non-negative integer d. Note that this is a proper density function. We have

$$f(a) = \frac{1}{C} \sum_{i=0}^{d} c_i a^i e^{-a\gamma}, \tag{4.7}$$

where $C = \sum_{i=0}^{d} \frac{c_i \Gamma(i+1)}{\gamma^{i+1}}$. The *Gammapoly* density can be considered as finite mixture of *Gamma* distributions ($Gamma(i+1,\gamma)$, $i = 1,\ldots,d$) with weights $w_i = \frac{c_i \Gamma(i+1)}{C\gamma^{i+1}}$.

When $\Phi(a) = 1$, $f(a)$ reduces to an exponential distribution. For $\Phi(a) = c_i a^i$ for some i, $f(a)$ gives a *Gamma* distribution. We also note that this can be considered as a generalization of Lindley distribution given by Lindley (1958) [24]. The Lindley distribution is a mixture of Gamma and exponential distribution with probability density function,

$$f(a|\gamma) = \frac{\gamma^2}{1+\gamma}(1+a)e^{-\gamma a}.$$

The Lindley distribution is special case of *Gammapoly* with $\Phi(a) = 1 + a$. The large number of parameters in *Gammapoly* provides flexibility to cover large class of distributions which was not available with Lindley distribution due to single parameter.

We also note that, if $\Phi(a) = a^{\alpha-1}(a+c)^d$, $c > 0$, $\alpha, d \in \mathbb{N}$, we get

$$\begin{aligned} f(a) &= \frac{a^{\alpha-1}(a+c)^d e^{-\gamma a}}{\int a^{\alpha-1}(a+c)^d e^{-\gamma a}\, da} \\ &= \frac{a^{\alpha-1}(a+c)^d e^{-\gamma a}}{\Gamma(\alpha) c^{\alpha+d} U\left(\alpha, d+\alpha+1, c\gamma\right)} \end{aligned} \tag{4.8}$$

where $U(x, y, z)$ is confluent hypergeometric function of second kind, defined as follows,

$$U(\alpha, y, z) = \frac{1}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1}(1+t)^{y-\alpha-1} e^{-tz}\, dt.$$

Equation (4.8) is the density function of Kummer distribution of type 2 given by Hamza and Vallois (2016) in [12].

Moment Generating Function

Proposition 4.6.1. *The moment generating function for a random variable X with the density function*

$$f(x) = \frac{1}{C} \sum_{i=0}^{d} c_i x^i e^{-x\gamma} \tag{4.9}$$

is

$$\mathbb{M}_X(t) = \frac{1}{C} \sum_{i=0}^{d} \frac{c_i \Gamma(i+1)}{\gamma^{i+1}} \left(\frac{1}{1 - \frac{t}{\gamma}} \right)^{i+1}.$$

Proposition 4.6.2. *The r^{th} moments of a random variable with the density (4.9) are given as*

$$\mathbb{E}(X^r) = \frac{d^r}{dt^r} \mathbb{M}_X(t)\Big|_{t=0} = \frac{1}{C} \sum_{i=0}^{d} \frac{c_i \Gamma(i+1)}{\gamma^{i+1}} \frac{(i+r)!}{i! \gamma^r}, \quad r = 1, 2, \ldots$$

In particular,

$$\mathbb{E}(X) = \frac{1}{C} \sum_{i=0}^{d} \frac{c_i \Gamma(i+1)}{\gamma^{i+1}} \frac{i+1}{\gamma}$$

and

$$\mathbb{E}(X^2) = \frac{1}{C} \sum_{i=0}^{d} \frac{c_i \Gamma(i+1)}{\gamma^{i+1}} \frac{(i+1) + (i+1)^2}{\gamma^2}.$$

With *Gammapoly* prior for a, we consider Beta prior for θ, *i.e.*

$$g(\theta|\beta_1,\beta_2) = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\,\theta^{\beta_1-1}(1-\theta)^{\beta_2-1}\,\mathbb{I}_{(0,1)}(\theta), \quad \textit{with known } \beta_1,\beta_2>0. \tag{4.10}$$

Hence the joint prior density for a and θ is

$$\pi(a,\theta) = \frac{1}{C}\sum_{i=0}^{d} c_i a^i e^{-a\gamma}\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\,\theta^{\beta_1-1}(1-\theta)^{\beta_2-1}, \tag{4.11}$$

where $C=\sum_{i=0}^{d}\frac{c_i\Gamma(i+1)}{\gamma^{i+1}}$, $a>0,\ \theta\in(0,1)$.

With these choices of priors for a and θ, the c.d.f. of the mean $\mu=\frac{a(1-\theta)}{\theta}$ is given as follows:

$$F(\mu) = \int_0^\infty\int_{\frac{a}{a+\mu}}^{1} f(a)g(\theta)\,d\theta\,da.$$

In particular, if $a\sim Gamma(\alpha,\gamma)$, we get the following result for the distribution of μ.

Proposition 4.6.3. *Suppose a is distributed as $Gamma(\alpha,\gamma)$ and θ is distributed as $Beta(\beta_1,\beta_2)$. Then, the probability density function for $\mu=\frac{a(1-\theta)}{\theta}$ is given by*

$$p(\mu) = \frac{\gamma^\alpha\mu^{\alpha-1}}{Beta(\beta_1,\beta_2)}U(\alpha+\beta_1,\alpha-\beta_2+1,\gamma\mu). \tag{4.12}$$

Proof. Since $\mu=\frac{a(1-\theta)}{\theta}$ or $\theta=\frac{a}{a+\mu}$, the *c.d.f* of μ is

$$\begin{aligned}
F(\mu) &= \int_0^\infty\int_{\frac{a}{a+\mu}}^{1} f(a)\pi(\theta)\,d\theta\,da \\
&= \int_0^\infty\int_{\frac{a}{a+\mu}}^{1} \frac{\gamma^\alpha a^{\alpha-1}e^{-\gamma a}}{\Gamma(\alpha)}\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\theta^{\beta_1-1}(1-\theta)^{\beta_2-1}\,d\theta\,da \\
&= 1-\int_0^\infty\int_0^{\frac{a}{a+\mu}} \frac{\gamma^\alpha a^{\alpha-1}e^{-\gamma a}}{\Gamma(\alpha)}\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\theta^{\beta_1-1}(1-\theta)^{\beta_2-1}\,d\theta\,da \\
&= 1-\int_0^\infty \frac{\gamma^\alpha a^{\alpha-1}e^{-\gamma a}}{\Gamma(\alpha)}\frac{Beta(\frac{a}{a+\mu},\beta_1,\beta_2)}{Beta(\beta_1,\beta_2)}\,da,
\end{aligned}$$

where $Beta(z, \beta_1, \beta_2)$ is the incomplete beta function defined as

$$Beta(z, \beta_1, \beta_2) = \int_0^z \theta^{\beta_1 - 1}(1-\theta)^{\beta_2 - 1}\, d\theta$$

Then the $p.d.f$ of μ is

$$\begin{aligned} p(\mu) &= -\int_0^\infty \frac{\gamma^\alpha a^{\alpha-1} e^{-\gamma a}}{\Gamma(\alpha)} \frac{\frac{d}{d\mu} Beta(\frac{a}{a+\mu}, \beta_1, \beta_2)}{Beta(\beta_1, \beta_2)}\, da \\ &= \int_0^\infty \frac{\gamma^\alpha a^{\alpha-1} e^{-\gamma a}}{\Gamma(\alpha) Beta(\beta_1, \beta_2)} \frac{a^{\alpha+\beta_1-1}}{(a+\mu)^{\beta_1+\beta_2}}\, da \\ &= \frac{\gamma^\alpha \mu^{\alpha-1}\Gamma(\alpha+\beta_1)}{Beta(\beta_1, \beta_2)\Gamma(\alpha)} U(\alpha+\beta_1, \alpha-\beta_2+1, \gamma\mu), \end{aligned}$$

where we have used the fact that

$$\frac{d}{dz} Beta(\alpha, \beta_1, \beta_2) = z^{\beta_1-1}(1-z)^{\beta}.$$

□

4.6.2 Posterior Density Distribution

Let $t_1, t_2, \ldots, t_n$ be n independent observations from Negative Binomial given in (4.1). We will use $\boldsymbol{t}$ to denote $(t_1, t_2, \ldots, t_n)$. The likelihood function is given as

$$L(a, \theta | \boldsymbol{t}) = \prod_{i=1}^n \frac{(a)_{t_i}}{t!} \theta^{na}(1-\theta)^{\sum_{i=1}^n t_i}. \tag{4.13}$$

We consider the prior

$$\pi(a, \theta) = \frac{1}{C} \sum_{i=0}^d c_i a^i e^{-a\gamma} \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} \theta^{\beta_1-1}(1-\theta)^{\beta_2-1} \tag{4.14}$$

With respect to this prior, the marginal density for $\boldsymbol{t}$ is

$$
\begin{aligned}
p(\boldsymbol{t}) &= \iint L(a,\theta|\boldsymbol{t})\pi(a,\theta)\,da\,d\theta \\
&= \frac{1}{C}\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\iint \prod_{i=1}^{n}\frac{(a)_{t_i}}{t!}\theta^{na+\beta_1-1}(1-\theta)^{\sum_{i=1}^{n}t_i+\beta_2-1}\sum_{i=0}^{d}c_i a^i exp(-a\gamma)\,da\,d\theta \\
&= \frac{1}{C}\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\frac{\Gamma(\sum_{i=1}^{n}t_i+\beta_2)}{t!}\int \frac{\prod_{i=1}^{n}(a)_{t_i}\sum_{i=0}^{d}c_i a^i}{(na+\beta_1)_{\sum_{i=1}^{n}t_i+\beta_2}}exp(-a\gamma)\,da,
\end{aligned}
$$

and the joint posterior density of a and θ is

$$
\begin{aligned}
\pi(a,\theta|\boldsymbol{t}) &= \frac{L(a,\theta|\boldsymbol{t})f(a)\pi(\theta)}{p(\boldsymbol{t})} \\
&= \frac{\prod_{i=1}^{n}(a)_{t_i}\sum_{i=0}^{d}c_i a^i e^{-\gamma a}\theta^{na+\beta_1-1}(1-\theta)^{\sum_{i=1}^{n}t_i+\beta_2-1}}{\Gamma(\sum_{i=1}^{n}t_i+\beta_2)K},
\end{aligned}
\tag{4.15}
$$

where

$$
K = \int \frac{\prod_{i=1}^{n}(a)_{t_i}\sum_{i=0}^{d}c_i a^i e^{-\gamma a}}{(na+\beta_1)_{\sum_{i=1}^{n}t_i+\beta_2}}\,da.
$$

The posterior marginal densities of a and θ are :

$$
\begin{aligned}
\pi(a|\boldsymbol{t}) &= \int_0^1 \pi(a,\theta|\boldsymbol{t})\,d\theta \\
&= \frac{\prod_{i=1}^{n}(a)_{t_i}\Gamma(na+\beta_1)\sum_{i=0}^{d}c_i a^i e^{-\gamma a}}{\Gamma(na+\beta_1+\sum_{i=1}^{n}t_i+\beta_2)K},
\end{aligned}
\tag{4.16}
$$

and

$$
\begin{aligned}
\pi(\theta|\boldsymbol{t}) &= \int_0^\infty \pi(a,\theta|\boldsymbol{t})\,da \\
&= \frac{\theta^{\beta_1-1}(1-\theta)^{\sum t_i+\beta_2-1}}{\Gamma(\sum t_i+\beta_2)K}\int_0^\infty \prod(a)_{t_i}\sum c_i a^i e^{-a(\gamma-nlog\theta)}\,da
\end{aligned}
\tag{4.17}
$$

Semiconjugate priors are defined for more than one dimensional parameter space. It is a family of probability distributions on parameter space such that for each member of that family, the full posterior marginals belong to the same family.

Proposition 4.6.4. *The family of distributions with density*

$$\pi(a,\theta) = \frac{1}{C}\sum_{i=0}^{d} c_i a^i e^{-a\gamma}\frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)}\theta^{\beta_1-1}(1-\theta)^{\beta_2-1}, \quad a>0, \theta\in(0,1) \tag{4.18}$$

is semiconjugate for Negative binomial distribution (4.1).

Proof. From (4.15), the full marginal posterior distribution for a is

$$\pi(a|\theta,\boldsymbol{t}) \propto \prod_{i=1}^{n}(a)_{t_i}\sum_{i=0}^{d} c_i a^i e^{-(\gamma-nlog\theta)a}, \tag{4.19}$$

which is again *Gammapoly* distribution with polynomial part $\prod_{i=1}^{n}(a)_{t_i}\sum_{i=0}^{d} c_i a^i$ and shape parameter $\gamma - nlog\theta$.

Also, the full marginal distribution of θ is $Beta(na+\beta_1, \sum_{i=1}^{n} t_i+\beta_2)$ as

$$\pi(\theta|a,\boldsymbol{t}) = \frac{\Gamma(na+\beta_1+\sum_{i=1}^{n} t_i+\beta_2)}{\Gamma(na+\beta_1)\,\Gamma(\sum_{i=1}^{n} t_i+\beta_2)}\,\theta^{na+\beta_1-1}(1-\theta)^{\sum_{i=1}^{n} t_i+\beta_2-1}. \tag{4.20}$$

$\square$

4.6.3 Predictive Density Analysis

Let Y be a new observation from $NB(a,\theta)$ in (4.1). Then, predictive probability mass function for Y given $t_1, t_2, \ldots, t_n$ is given as

$$\begin{aligned} q(y|\boldsymbol{t}) &= \iint p(y|a,\theta)\pi(a,\theta|\boldsymbol{t})\,d\theta\,da \\ &= \frac{(\sum_{i=1}^{n} t_i+\beta_2)_y}{y!K}\int \frac{(a)_y\prod_{i=1}^{n}(a)_{t_i}\sum_{i=0}^{d} c_i a^i e^{-\gamma a}}{((n+1)a+\beta_1)_{(\sum_{i=1}^{n} t_i+\beta_2+y)}}\,da \end{aligned} \tag{4.21}$$

Proposition 4.6.5. *When $\beta_1 = 1$, the expected value of Y given* $\boldsymbol{t}$ *is independent of the choice of prior for a.*

Before proceeding further with the proof, we will recall some properties of ascending factorials which follows from the properties of *Gamma* function.

Lemma 4.6.6. *For $a, t, r > 0$ and $a > r$, we have,*

1. $(a-r)_{r+t} = (a)_t(a-r)_r$
2. *In particular, for $r = 1$, we get $(a-1)(a)_t = (a-1)_{t+1}$*

Proof. (Proposition 4.6.5) We have, from the definition of conditional expectations,

$$\begin{aligned}
E_q[Y|\boldsymbol{t}] =& E_{\pi(a,\theta|\boldsymbol{t})}\left[\frac{a(1-\theta)}{\theta}\right] \\
=& \iint \frac{a(1-\theta)}{\theta}\pi(a,\theta|\boldsymbol{t})\,da\,d\theta \\
=& \frac{\sum_{i=1}^n t_i + \beta_2}{K}\int \frac{a\prod_{i=1}^n (a)_{t_i}\sum_{i=0}^d c_i a^i e^{-\gamma a}}{(na+\beta_1-1)_{(\sum_{i=1}^n t_i+\beta_2+1)}}\,da.
\end{aligned}$$

Using property (2) of Lemma 4.6.6 in the denominator, we get, when $\beta_1 = 1$,

$$E_q[Y|\boldsymbol{t}] = \frac{\sum_{i=1}^n t_i + \beta_2}{n} \tag{4.22}$$

which is independent of the c_i's and γ. □

Moreover, we get a general result as follows:

Proposition 4.6.7. *For any $\beta_1, \beta_2 > 0$,*

$$E_{\pi(a,\theta|\boldsymbol{t})}\left[(na+\beta_1-1)\frac{(1-\theta)}{\theta}|\boldsymbol{t}\right] = \sum_{i=1}^n t_i + \beta_2 \tag{4.23}$$

When $\beta_1 = 1$, equation (4.23) reduces to (4.22). Note that, uniform prior choice for θ leads to

$$E_q[Y|\boldsymbol{t}] = \frac{\sum_{i=1}^n t_i + 1}{n}. \tag{4.24}$$

Proposition 4.6.8. *For any $\beta_1, \beta_2 > 0$, , $m_2 > m_1 > 0$*

$$E_{\pi(a,\theta|\boldsymbol{t})}\left[(na+\beta_1-m_1)_{m_1}\left(na+\beta_1+\sum_{i=1}^n t_i+\beta_2\right)_{m_2-m_1}\frac{(1-\theta)^{m_2}}{\theta^{m_1}}|\boldsymbol{t}\right] = \frac{\Gamma(\sum_{i=1}^n t_i+\beta_2+m_2)}{\Gamma(\sum_{i=1}^n t_i+\beta_2)}.$$

4.7 Computational aspects

The implementation of the above-mentioned method of Bayesian analysis of negative binomial distributed data requires sample generating methods from Gammapoly distribution (4.7) and posterior density distribution (4.15). In this section, we provide such methods which will enable us the employment of the new method of inference.

4.7.1 Generating samples from Gammapoly distribution

The characterization of *Gammapoly* distribution as a finite mixture of *Gamma* distributions, given in Section 4.6, is useful in sampling from *Gammapoly* distribution. To generate samples from the density function (4.7),

1. Compute the weights $w_i = \frac{c_i \Gamma(i+1)}{C \gamma^{i+1}}$, $i = 0, 1, \dots, d$;
2. Choose $i \in \{0, 1, \dots, d\}$ with probability w_i ;
3. Generate sample from $Gamma(i+1, \gamma)$.

4.7.2 Generating samples from posterior density distribution

Further, Gibbs sampler can be used to draw samples from posterior density (4.15) as we can sample from full marginal densities given in (4.19) and (4.20) as follows:

1. Choose prior distributions for a and θ, *i.e.* to fix values for coefficients $c_0, \dots, c_d$ in $\Phi(a)$, γ, β_1 and β_2 ;
2. Compute polynomial part in full marginal posterior density of a, which is necessary to compute weights to sample a from (4.19);
3. Start with an arbitrary value for θ;
4. Using this value of θ, sample a_{new} from (4.19);
5. Sample a new value for θ from $Beta(na_{new} + \beta_1, \sum t_i + \beta_2)$;

6. Repeat steps 4 and 5 till convergence of Gibbs sampler;

7. Throw away first few samples, also called as burn-in period of Gibbs sampler to get independent samples a_i and $\theta_i, i = 1, 2, \ldots, N$ from (4.15).

The above mentioned algorithm can be computationally inefficient in some cases, for example, large number of samples (n), large values of the observations, etc. But, this can be improved upon by avoiding computations of weights via long polynomial computations in step 2 by considering ratios of consecutive coefficients and normalizing them with respect to particular coefficient, set to be unity. In step 4, note that, the coefficients of $a^i, i = 0, 1, \cdots, n - n_0 + \delta - 1$, (where δ is the lowest non-zero degree in $\Phi(a)$) are all zero and the remaining coefficients are non-zero which further facilitates the sampling from the target distribution by the method described with improvement.

4.7.3 Estimators of a and θ

With respect to quadratic loss function, Bayes estimators for a and θ are posterior means of a and θ. These can be computed using posterior averages of samples generated using method presented in the previous section.

$$\hat{a} = \frac{\sum_{i=1}^{N} a_i}{N}, \quad \hat{\theta} = \frac{\sum_{i=1}^{N} \theta_i}{N} \tag{4.25}$$

We can also use the generated samples to estimate the median or quantiles of the posterior distributions. The median of a and θ are the Bayes estimators with respect to absolute error loss.

4.7.4 Empirical estimators of predictive density

To get empirical estimates for predictive probability density function (4.21) and expected value of new observation given $\boldsymbol{t}$ *i.e.* $E_q[Y|\boldsymbol{t}]$, for each sampled pair of a_i and θ_i sampled using Gibbs sampler, we compute $\frac{(a)_y}{y!}\theta_i^{a_i}(1-\theta_i)^y$ for $y = 0, 1, 2, \ldots$ and $\frac{a_i(1-\theta_i)}{\theta_i}$, and then take averages.

$$\hat{q}(y|\boldsymbol{t}) = \frac{\sum_{i=1}^{N} \frac{(a)_y}{y!}\theta_i^{a_i}(1-\theta_i)^y}{N} \quad , y = 0, 1, 2, \ldots \tag{4.26}$$

and

$$\hat{E}[Y|\boldsymbol{t}] = \frac{\sum_{i=1}^{N} \frac{a_i(1-\theta_i)}{\theta_i}}{N} \tag{4.27}$$

4.8 Numerical examples

4.8.1 Application to a tiny dataset

We consider a dataset $t_1 = 0, t_2 = 1, t_3 = 4$ of three independent observations from $NB(a, \theta)$ for some a and θ. For this particular example, we explicitly derived the expressions for marginal posterior densities of a and θ as well as joint posterior density , predictive density and expected predictive value for following prior values and compare the results of simulations with the analytically obtained expressions. Considering prior values as $\Phi(a) = 1$, $\gamma = 2$, $\beta_1 = \beta_2 = 1$, which corresponds to the prior $\pi(a, \theta) = 2e^{-2a}$, we get the joint posterior density

$$\pi(a, \theta|\boldsymbol{t}) = \frac{a^2(a+1)(a+2)(a+3)e^{-2a}\theta^{3a}(1-\theta)^5}{0.013215} \quad \cdots from\,(4.15).$$

The marginal posterior densities are

$$\pi(a|\boldsymbol{t}) = \frac{1008.97}{9}\left(\frac{2}{81}\frac{e^{-2a}}{3a+1} - \frac{14}{81}\frac{e^{-2a}}{3a+2} + \frac{40}{81}\frac{e^{-2a}}{3a+4} - \frac{25}{81}\frac{e^{-2a}}{3a+5}\right),$$

and

$$\pi(\theta|\boldsymbol{t}) = \frac{(1-\theta^5)}{12}\left\{\frac{120}{(2+3ln(\theta))^6} + \frac{144}{(2+3ln(\theta))^5} + \frac{66}{(2+3ln(\theta))^4} + \frac{12}{(2+3ln(\theta))^3}\right\}.$$

Figure (4.1) shows the comparison of analytic expressions and the results obtained via simulations using Gibbs sampler. With respect to quadratic loss, Bayes estimates for unknown parameters are

$$\hat{a}_{Bayes} = 0.7404, \quad \hat{\theta}_{Bayes} = 0.3353$$

The predictive density is given as

$$q(y|\boldsymbol{t}) = \frac{\Gamma(6+y)}{0.013215y!}\int \frac{(a)_y a^2(a+1)(a+2)(a+3)e^{-2a}}{(4a+1)\dots(4a+6+y)}\,da$$

Table 4.3 provides few values of predictive density.

y	0	1	2	3	4	5	6
$q(y\|\boldsymbol{t})$	0.5004	0.1817	0.1014	0.0621	0.0402	0.0271	0.0189

Table 4.3: Predictive density for a tiny dataset

Comparison of empirical estimates of predictive density with the values in Table 4.3 is shown in Figure 4.9. We can see that the empirical means, obtained using (4.26) and (4.27) matches well with the analytically derived values in Table 4.3.

4.8.2 Application to Fisher's data of the number of ticks on sheep

Consider the dataset (Table A.3) representing a sample of sheep classified according to the number of ticks found on each, as given by Fisher in [11].

Let prior be as follows:

$$a \sim Gamma(3,1), \quad \theta \sim Beta(1,1)$$

The posterior distributions for a and θ are given in Figure 4.3. Predictive density estimates are obtained (Figure 4.9).

	MoM	MLE	Bayes Estimates(Posterior Means)
a	4.10859	3.75254	3.625036
θ	0.558339	0.535902	0.514822

Table 4.4: Comparison of different estimators for a and θ

4.9 Risk comparison of estimators

We conducted a simulation study to compare Bayes estimate under quadratic loss with method of moments estimators and maximum likelihood estimators under different choices for a and θ. The values chosen for a and θ are $1, 2, 3$ and $0.25, 0.50, 0.75$, respectively. For each pair, $N = 1000$ samples of size $n = 50$ are generated and estimates for a and θ are computed. Method of moments (MoM) estimators are calculated using (4.5). The following prior is considered : $\pi(a, theta) = \frac{1}{2}e^{-\frac{1}{2}a}$, hence the prior mean of a is 2 and θ is 0.5. Figures (4.5), (4.6) and (4.7) shows boxplots for $|\hat{a}-a|$ and $|\hat{\theta}-\theta|$ where $\hat{a}$ and $\hat{\theta}$ denote the estimates for a and θ (namely, MoM estimators and posterior expectations of a and θ) , respectively. We have considered samples which are overdispersed as we know that for underdispersed samples, MoM and MLE estimators either lie outside the parameter space or do not exist. All the computations are done using **R**.

From the simulation study, we can see that, in almost all the cases studied, Bayes estimators performs better than or equal to other estimates, when absolute deviation from the true values of parameters is considered. The variability in Bayes estimators is less as compared to other estimates.

Risk functions with respect to quadratic loss functions are plotted (4.9) for a and θ for $(a, \theta) \in (0, 10) \times (0, 1)$. The samples of size 5 are considered for this simulation study. It can be seen that, except for the extreme values of θ, the frequentist risk with respect to quadratic loss function of Bayes estimator is better than that of the MoM estimator, whenever they exists.

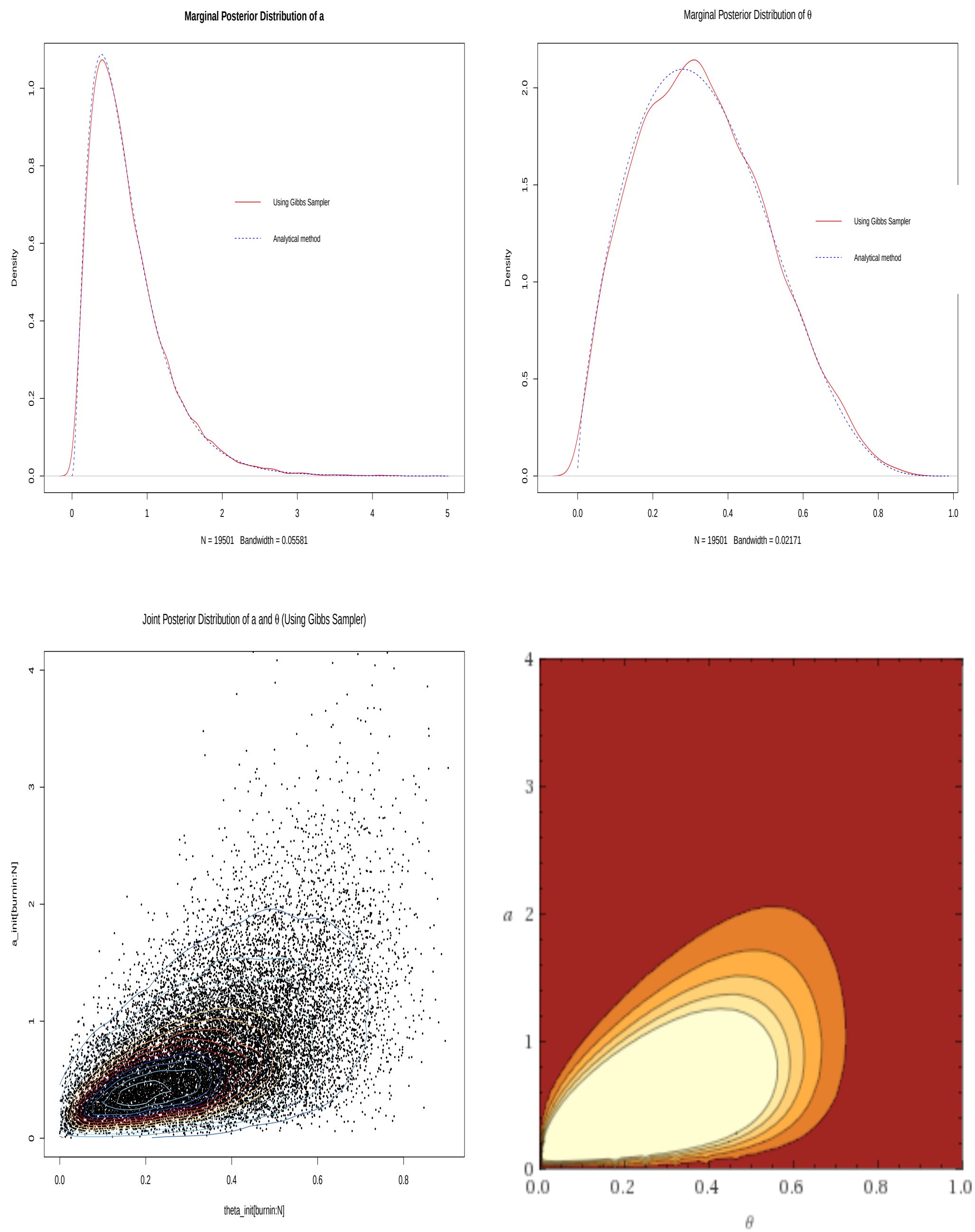

Figure 4.1: Posterior summaries for Tiny data $t_1 = 0, t_2 = 1, t_3 = 4$

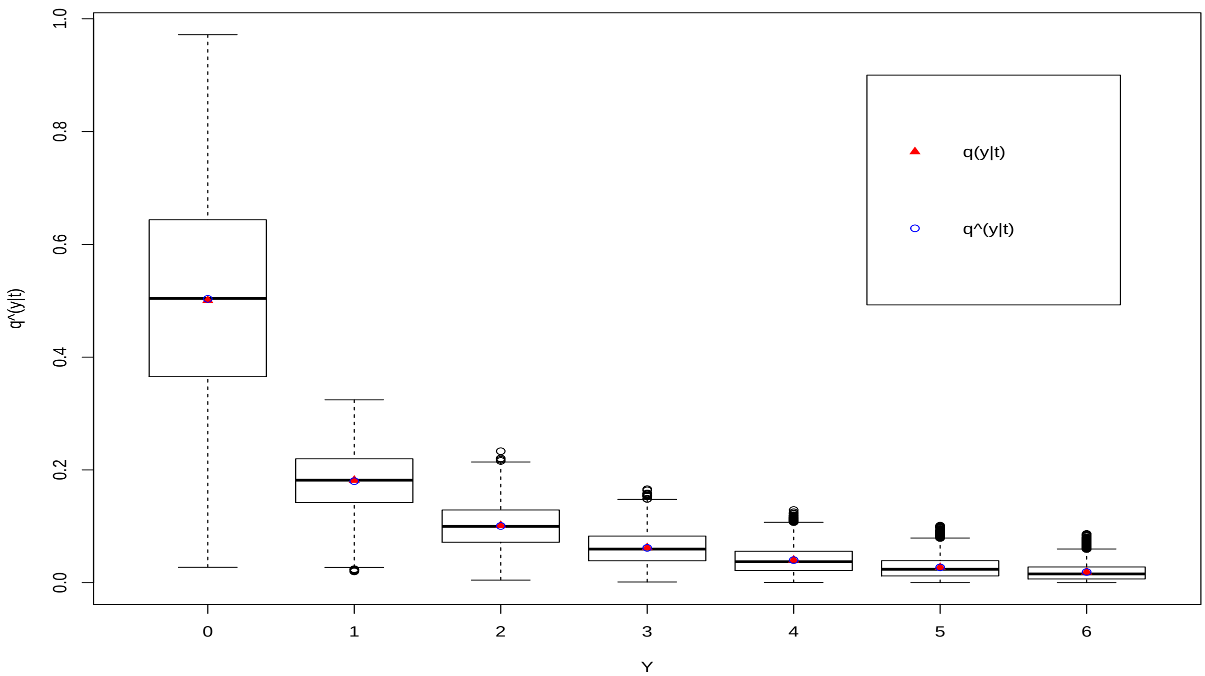

Figure 4.2: Predictive densities for tiny data set example obtained analytically and estimated using simulations

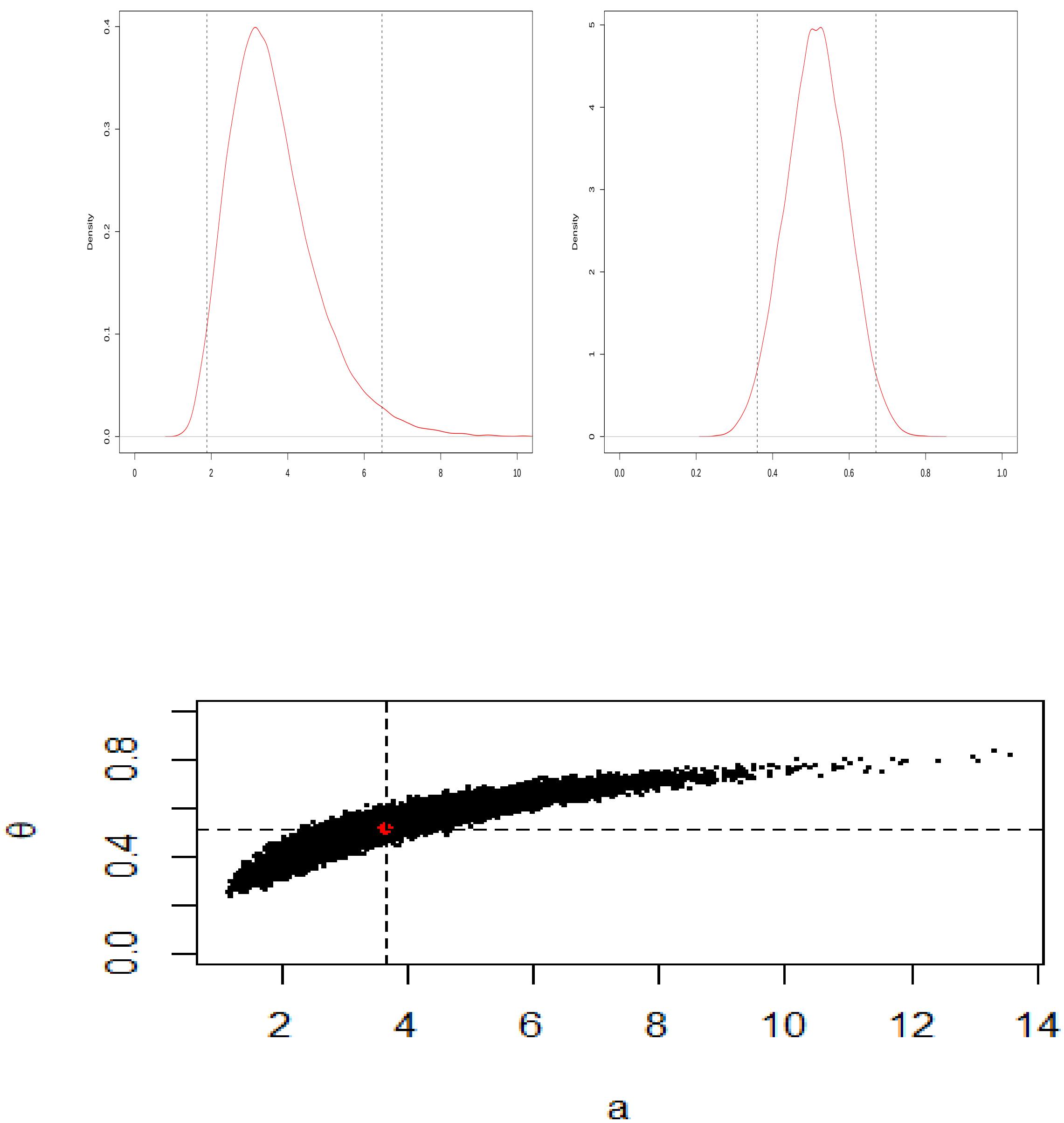

Figure 4.3: Fisher Data 1) Marginal Posterior density for a with 95% Credible interval. 2)Marginal Posterior density for θ with 95% Credible interval .3) Joint Posterior density of a and θ

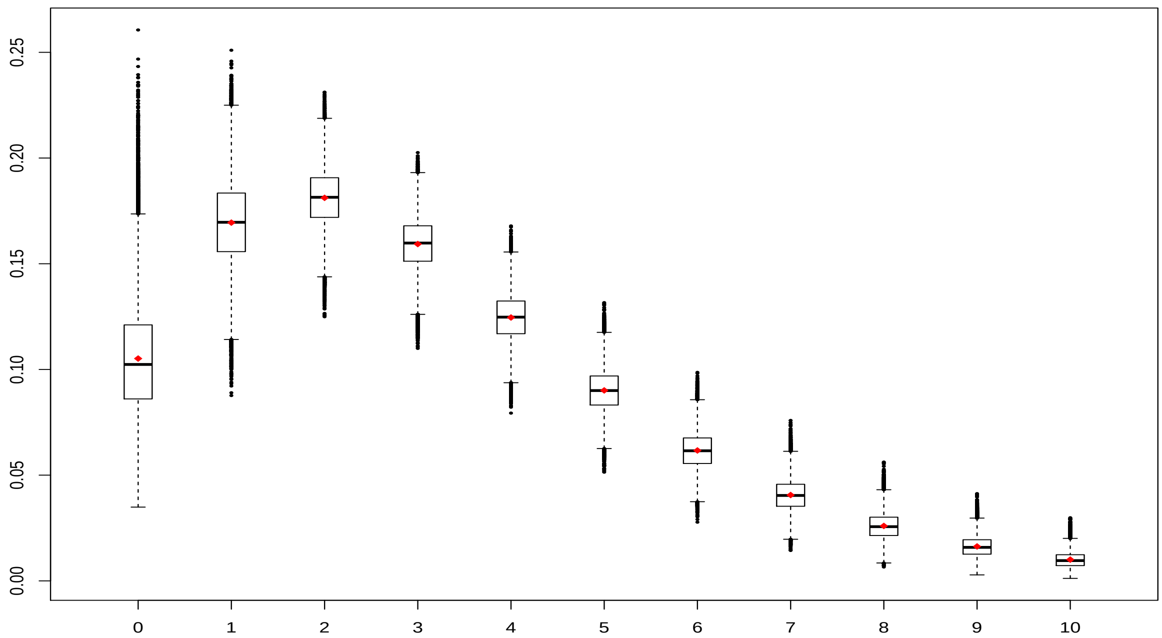

Figure 4.4: Predictive density for Fishers' data

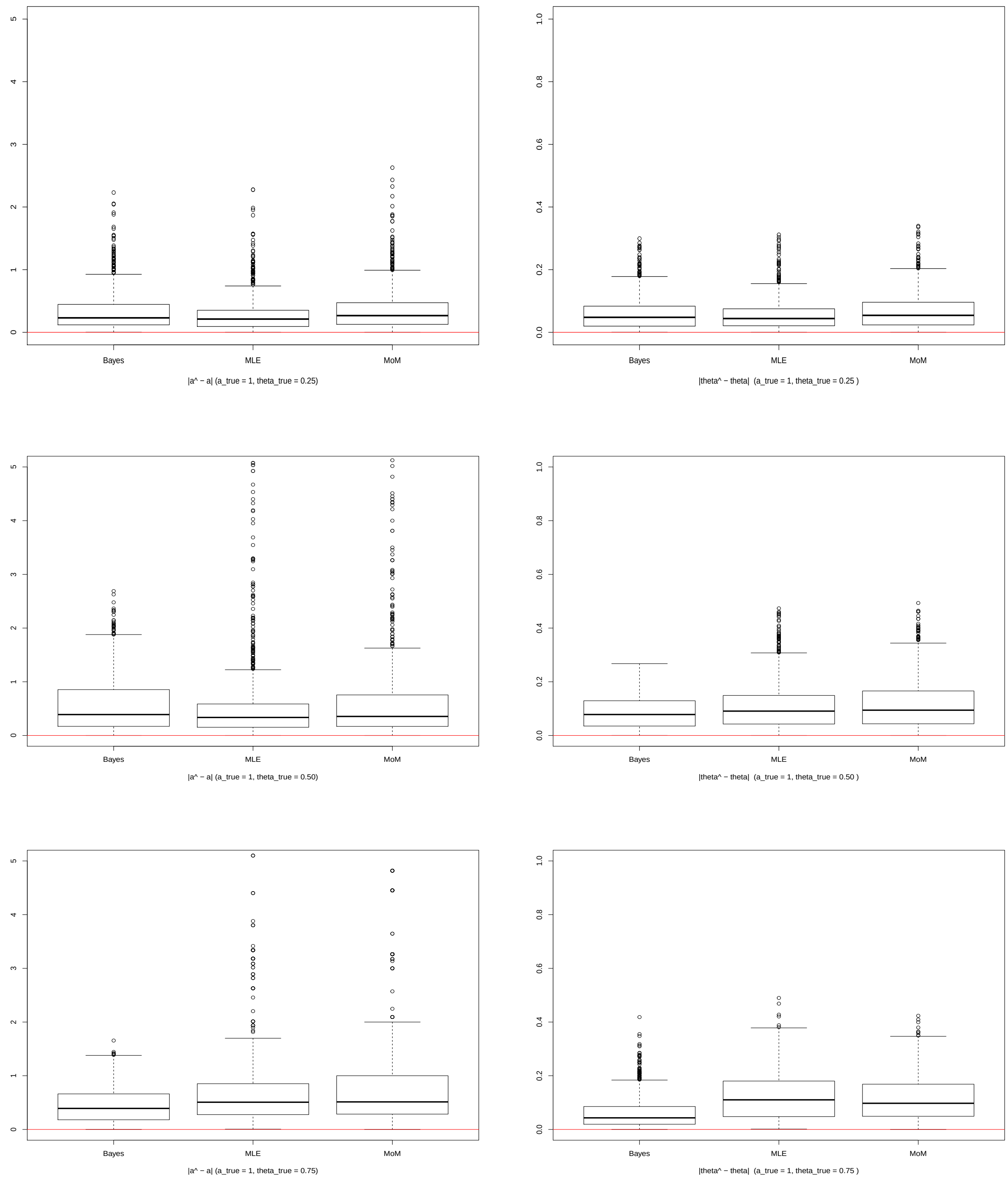

Figure 4.5: $|\hat{a} - a|$ and $|\hat{\theta} - \theta|$ plots for $a = 1$ and $\theta = 0.25, 0.50, 0.75$

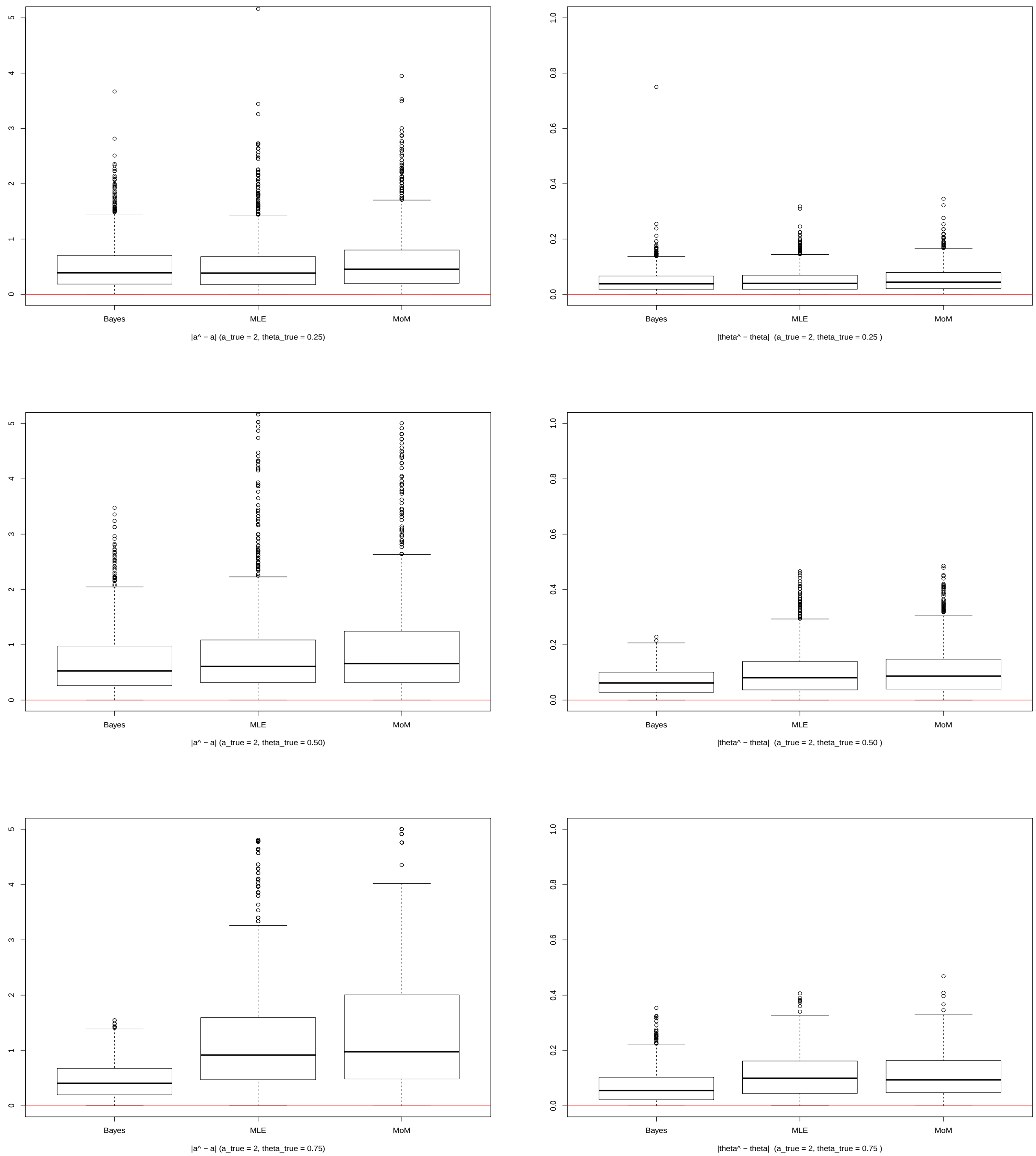

Figure 4.6: $|\hat{a} - a|$ and $|\hat{\theta} - \theta|$ plots for $a = 2$ and $\theta = 0.25, 0.50, 0.75$

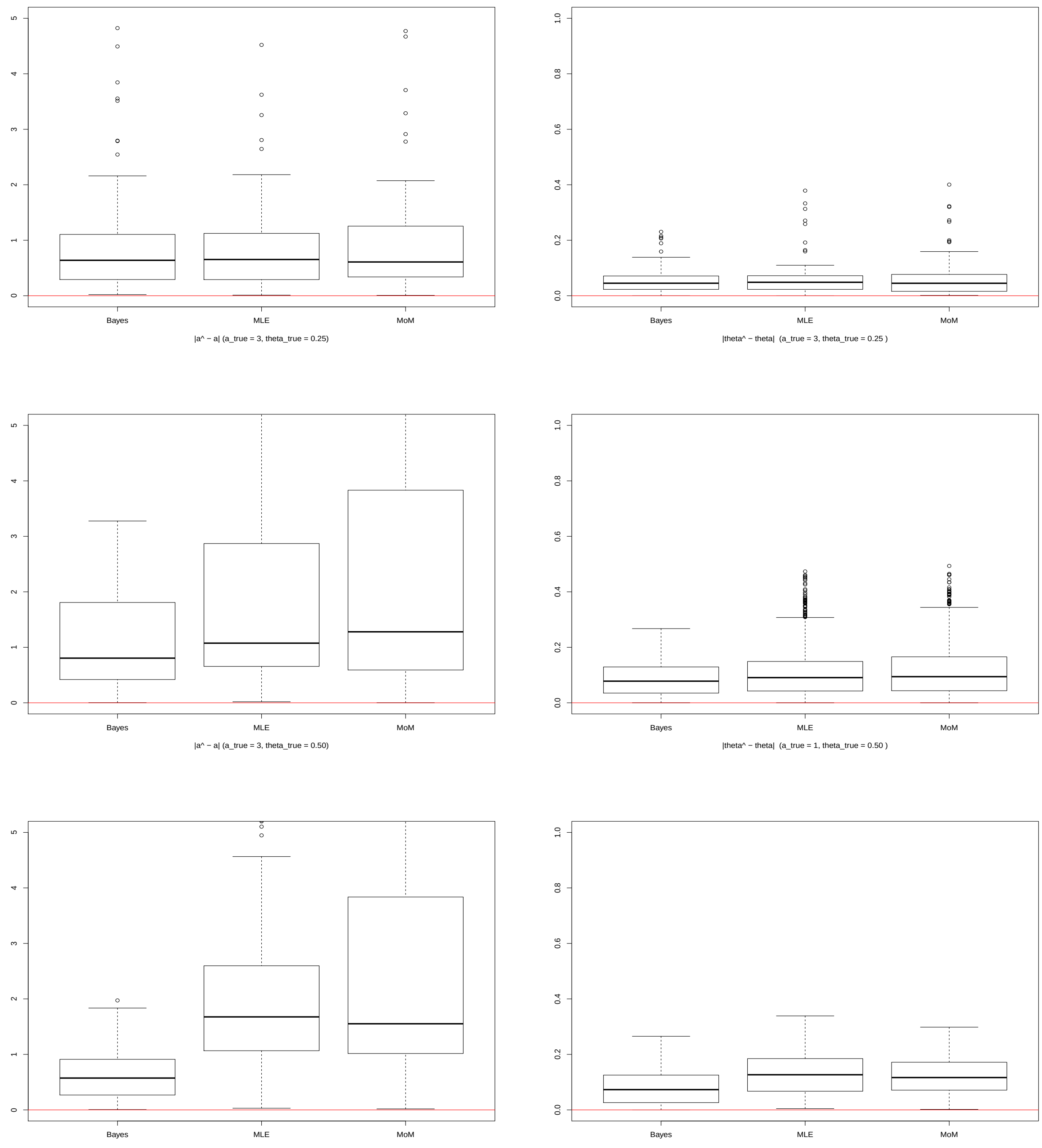

Figure 4.7: $|\hat{a} - a|$ and $|\hat{\theta} - \theta|$ plots for $a = 3$ and $\theta = 0.25, 0.50, 0.75$

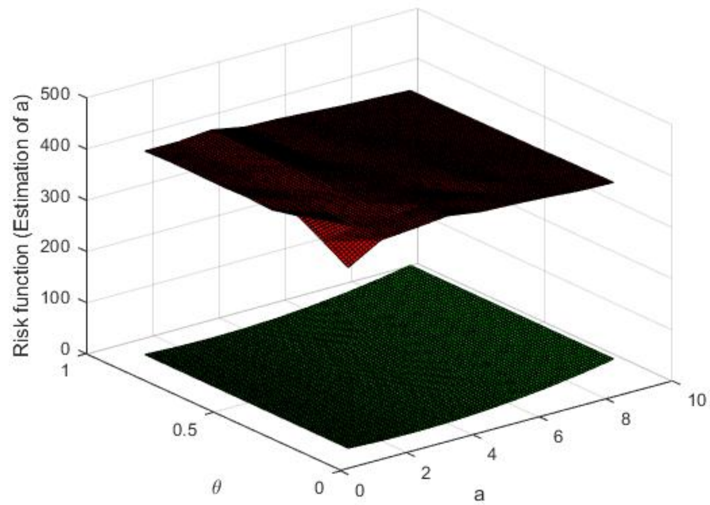

Figure 4.8: Risk comparison for Bayes estimator (*green*) and MoM estimator (*red*) of a

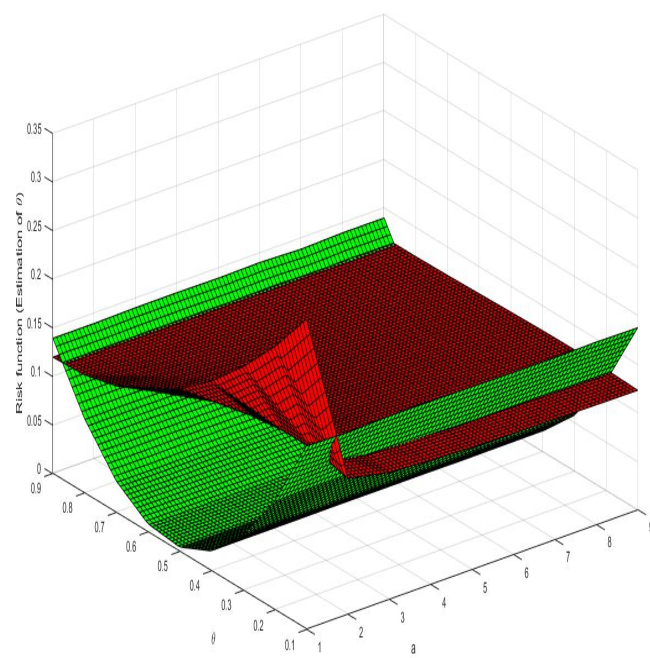

Figure 4.9: Risk comparison of Bayes estimator (*green*) and MoM estimator (*red*) of θ

Chapter 5

Parameter Estimation for the Polya Distribution

In this Chapter, we consider the problem of the parameter estimation for the Polya distribution. We propose the use of a data cloning method to obtain MLE for such model. A Bayesian approach is also suggested. All methods developed are illustrated using real dataset.

5.1 Polya Distribution

Let $X = (X_1, \ldots, X_d)$ be a random vector where $X_i \in \{0, 1, \ldots, t\}$ and $\sum_{i=1}^{d} X_i = t$. The probability mass function for a $Polya(t; \alpha_1, \ldots, \alpha_d)$ distribution is

$$p(x_1, \ldots, x_d | \alpha_1, \ldots, \alpha_d, t) = \frac{t!}{(\alpha)_t} \prod_i \frac{(\alpha_i)_{x_i}}{x_i!}, \tag{5.1}$$

where, $\alpha = \sum\limits_i \alpha_i, \quad t = \sum\limits_i x_i$. Let $(x_{1j}, \ldots, x_{dj}), j = 1, 2, \ldots, n$ be n independent observations from $Polya(t_j; \alpha_1, \ldots, \alpha_d)$ distribution, $j = 1, 2, \ldots, n$, respectively. The likelihood function is given as

$$\mathbb{L}(\alpha_1, \ldots, \alpha_d) = \prod_{j=1}^{n} \frac{t_j!}{(\alpha)_{t_j}} \prod_i \frac{(\alpha_i)_{x_{ij}}}{x_{ij}!} \tag{5.2}$$

Here, we consider the problem of estimating α_i's.

5.2 Maximum likelihood estimation

Maximum likelihood estimators for α_i's are the solutions for the following score equations:

$$\sum_{j=1}^{n}\sum_{l=0}^{x_{ij}-1}\frac{1}{\alpha_i+l}=\sum_{j=1}^{n}\sum_{l=0}^{t_j-1}\frac{1}{\alpha+l}, \quad i=1,2,\ldots,d \tag{5.3}$$

Levin and Reeds (1977) showed that the MLE may not exist for Polya distribution in [23]. No closed form solutions can be obtained for the MLE. Numerical methods are needed to be implemented to obtain solutions for the system of equations (5.3). Here we provide an alternative method to obtain MLE using data cloning.

5.2.1 Data Cloning

A well known result in asymptotic Bayes theory that under certain regularity conditions, the posterior distribution $\pi(\theta|x_1,\ldots,x_n)$ asymptotically approaches to $N_d(\hat{\theta}_{MLE}, I^{-1}(\hat{\theta}_{MLE}))$ as $n\to\infty$, where $\hat{\theta}_{MLE}$ is MLE of the parameter θ and $I(\theta)$ is the information matrix based on n observations $x_1,\ldots,x_n$. A rigorous proof of such result is provided by Walker (1969) in [33]. Here is the version of the result for a multidimensional parameter spaces:

Theorem 5.2.1. *Let $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_d)\in\Theta$ be $d-$ dimensional parameter vector. Let $X_1,\ldots,X_n$ have density $\prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$. Assume that following regularity conditions are satisfied:*

1. *$\mathbb{S}=\{x: f(x|\boldsymbol{\theta})\}$ does not depend on $\boldsymbol{\theta}$.*
2. *Identification : The distributions corresponding to two distinct values of $\boldsymbol{\theta}$ are different.*
3. *Consider $x\in\mathbb{S}$ and $\boldsymbol{\theta}'\in\Theta$. Then there exist $\delta>0$ such that $\forall\boldsymbol{\theta}$ with $|\boldsymbol{\theta}-\boldsymbol{\theta}'|<\delta$, we have,*

$$|log(f(x|\boldsymbol{\theta}))-log(f(x|\boldsymbol{\theta}'))|<H_\delta(x,\boldsymbol{\theta}')$$

and

$$\lim_{\delta \to 0} H_\delta(x, \boldsymbol{\theta}') = 0$$

.

Also, for any $\boldsymbol{\theta}_0 \in \Theta$,

$$\lim_{\delta \to 0} \int_{\mathbb{S}} H_\delta(x, \boldsymbol{\theta}') f(x|\boldsymbol{\theta}_0) d\mu = 0.$$

In the below conditions, let $\boldsymbol{\theta}_0$ be any interior point in Θ.

4. *For unbounded Θ and any $\boldsymbol{\theta}_0 \in \Theta$,*

$$log(f(x|\boldsymbol{\theta})) - log(f(x|\boldsymbol{\theta}_0)) < K_\Delta(x, \boldsymbol{\theta}_0), \quad \forall \boldsymbol{\theta} \;\; such \;\; that \;\; ||\boldsymbol{\theta}|| > \Delta,$$

where

$$\lim_{\Delta \to \infty} \int_{\mathbb{S}} K_\Delta(x, \boldsymbol{\theta}_0) f(x|\boldsymbol{\theta}_0) d\mu < \infty.$$

5. *$Logf(x|\boldsymbol{\theta})$ is a twice differentiable function of $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$.*

6. *The $d \times d$ information matrix $I(\boldsymbol{\theta}_0)$ should be finite and positive definite. Here,*

$$I_{ij}(\boldsymbol{\theta}_0) = \int_{\mathbb{S}} f(x|\boldsymbol{\theta}_0) \left(\frac{\partial f(x|\boldsymbol{\theta})}{\partial \theta_i} \right)_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} \left(\frac{\partial f(x|\boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta} = \boldsymbol{\theta}_0} d\mu.$$

7. *The prior density $\pi(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_0$ and $\pi(\boldsymbol{\theta}_0) > 0$.*

Then, the posterior distribution $\pi(\boldsymbol{\theta}|x_1, \ldots, x_n)$ converges in distribution to $N_d(\hat{\boldsymbol{\theta}}_{MLE,n}, \frac{1}{n} I^{-1}(\hat{\boldsymbol{\theta}}_{MLE,n}))$, where $\hat{\boldsymbol{\theta}}_{MLE,n}$ is MLE based on the sample $x_1, \ldots, x_n$.

Using this result, Lele, *et al.* (2010) proposed method of data cloning to obtain MLE in case of random effect models [22]. The idea of the data cloning is as follows :

Suppose we have n observations $x_1, \ldots, x_n$ with the density $\prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$. For any prior $\pi(\boldsymbol{\theta})$, the posterior density for $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}|x_1, \ldots, x_n) \propto \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \pi(\boldsymbol{\theta})$$

Now the cloned data is obtained by considering k copies of the observed data. Then the likelihood function for the cloned data is

$$\mathbb{L}(\boldsymbol{\theta}) = \left(\prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})\right)^k .$$

Here, note that the MLE's based on the observed data and the cloned data are same. The posterior distribution based on the cloned data is

$$\pi(\boldsymbol{\theta}|(x_1, \ldots, x_n)^k) \propto \left(\prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})\right)^k \pi(\boldsymbol{\theta}).$$

Invoking above theorem 5.2.1 for the cloned data, we get, the posterior distribution $\pi(\boldsymbol{\theta}|(x_1, \ldots, x_n)^k)$ converges in distribution to $N_d(\hat{\boldsymbol{\theta}}_{MLE,n}, \frac{1}{nk}I^{-1}(\hat{\boldsymbol{\theta}}_{MLE,n}))$. Hence, for fixed n, if we let $k \to \infty$, the variance matrix will shrink to $\mathbf{0}$ and the samples from the posterior density based on the cloned data can be used as as estimator for $\hat{\boldsymbol{\theta}}_{MLE,n}$. Further, we can use this normal approximation to construct confidence intervals for the estimate of $\hat{\boldsymbol{\theta}}_{MLE,n}$.

5.2.2 Using Data Cloning for estimating the MLE for a Polya Distribution

Let $(x_{1j}, \ldots, x_{dj})\, j = 1, 2, \ldots, n$ be n independent observations from Polya distribution (5.1). The likelihood function is given in (5.2).

Levin and Reeds (1977) showed that the MLE may not be finite for some samples [23]. They proved that the MLE is finite if and only if

$$\sum_{j=1}^{n}\sum_{i=1}^{d} \frac{(x_{ij} - t)^2}{t} > \sum_{i=1}^{d} \frac{\sum_{j=1}^{n} x_{ij}}{t} - n. \tag{5.4}$$

Assuming that the MLE exists for a given data, here is an algorithm to estimate the MLE using a data cloning approach:

1. We consider independent priors for α_i's.

$$\pi(\alpha_1, \ldots, \alpha_d) = \pi(\alpha_1) \cdots \pi(\alpha_d).$$

2. Fix k. Then, the posterior distribution of α_i's is

$$\pi(\alpha_1, \ldots, \alpha_d | (x_{1j}, \ldots, x_{dj})^k) \propto (\mathbb{L}(\alpha_1, \ldots, \alpha_d))^k \, \pi(\alpha_1) \cdots \pi(\alpha_d). \tag{5.5}$$

3. Then, we have the full posterior marginal distributions for α_i's as follows:

$$\pi(\alpha_i | \alpha_{j \neq i}, (x_{1j}, \ldots, x_{dj})^k) \propto \pi(\alpha_i) \left(\prod_{j=1}^{n} \frac{(\alpha_i)_{x_{ij}}}{(\alpha)_{t_j}} \right)^k.$$

4. Use Gibbs sampler to sample $(\alpha_1, \ldots, \alpha_d)$ from the joint posterior density (5.5).

5. Repeat steps 2, 3 and 4 with increasing values for k till samples start to converge.

6. The mean of the samples converges to the MLE for α_i's.

In cases where finite MLE solutions does not exist for α_i's, it is observed that $\frac{\alpha_i}{\alpha}$ converges for large k. If $\alpha_1, \ldots, \alpha_d \to \infty$ such that $\frac{\alpha_i}{\alpha} \to \phi_i$, the probability mass function for this limiting case of Polya distribution is

$$p(x_1, \ldots, x_d | \phi_1, \ldots, \phi_d) = d! \prod_i \frac{\phi_i^{x_i}}{x_i!} \tag{5.6}$$

Then, the data cloning method can also be used to estimate ϕ_i's by considering $\frac{\alpha_i}{\alpha}$ for the samples generated using large values of k.

5.2.3 Examples

Here, we present examples of cases where

1) the MLE solution exists and

2) the MLE solution does not exist.

Shunter's Accident Data

Conditional on the sum T, we fit a Polya disribution$(T; \alpha_1, \alpha_2)$ to the bivariate Shunter Accidents data. This dataset does not admit finite MLE solution. As expected, the α_i's don't converge for any k (Fig. 5.1). But, if we consider ϕ_i's as defined in earlier section, it can be seen that ϕ_i's converge (Fig. 5.2). Table 5.1 shows estimates for ϕ_i's with standard errors.

k	$\hat{\phi}_{1MLE}$	$\hat{\phi}_{2MLE}$
100	$0.434212(8.95 \times 10^{-6})$	$0.565788(8.95 \times 10^{-6})$
1000	$0.4340187(1.48 \times 10^{-6})$	$0.5659813(1.48 \times 10^{-6})$

Table 5.1: MLEs of ϕ_i's for Shunter's accident bivariate data in Table A.1 using data cloning for different values of k. The bracketed quantities denote the standard error.

Aitchison's Trivariate Bacterial Data Count

Conditional on the sum T, we fit a Polya disribution$(t; \alpha_1, \alpha_2, \alpha_3)$ to the trivariate bacterial data given in Table A.2. This dataset does admit finite MLE solution. Fig. 5.3 shows sample iterates obtained for $\alpha_1, \alpha_2, \alpha_3$ from the posterior density using Gibbs sampler for cloned Aitchison's trivariate data with $k = 100, 1000$ and 10000. Table 5.2 shows estimates for MLE of α_i's with standard errors. We can see that, as k becomes large, the standard error decreases which is suggested by Theorem 5.2.1.

k	$\hat{\alpha_1}$	$\hat{\alpha_2}$	$\hat{\alpha_3}$
100	2.0881(0.0461)	2.8949(0.0602)	2.6497(0.0560)
1000	2.0839(0.0133)	2.8334(0.0194)	2.6471(0.0183)
10000	2.0781(0.0072)	2.8248(0.0101)	2.6402(0.0101)

Table 5.2: MLE of α_i's for Aitchison's trivariate data using data cloning for different values of k. The bracketed quantities denote the standard error.

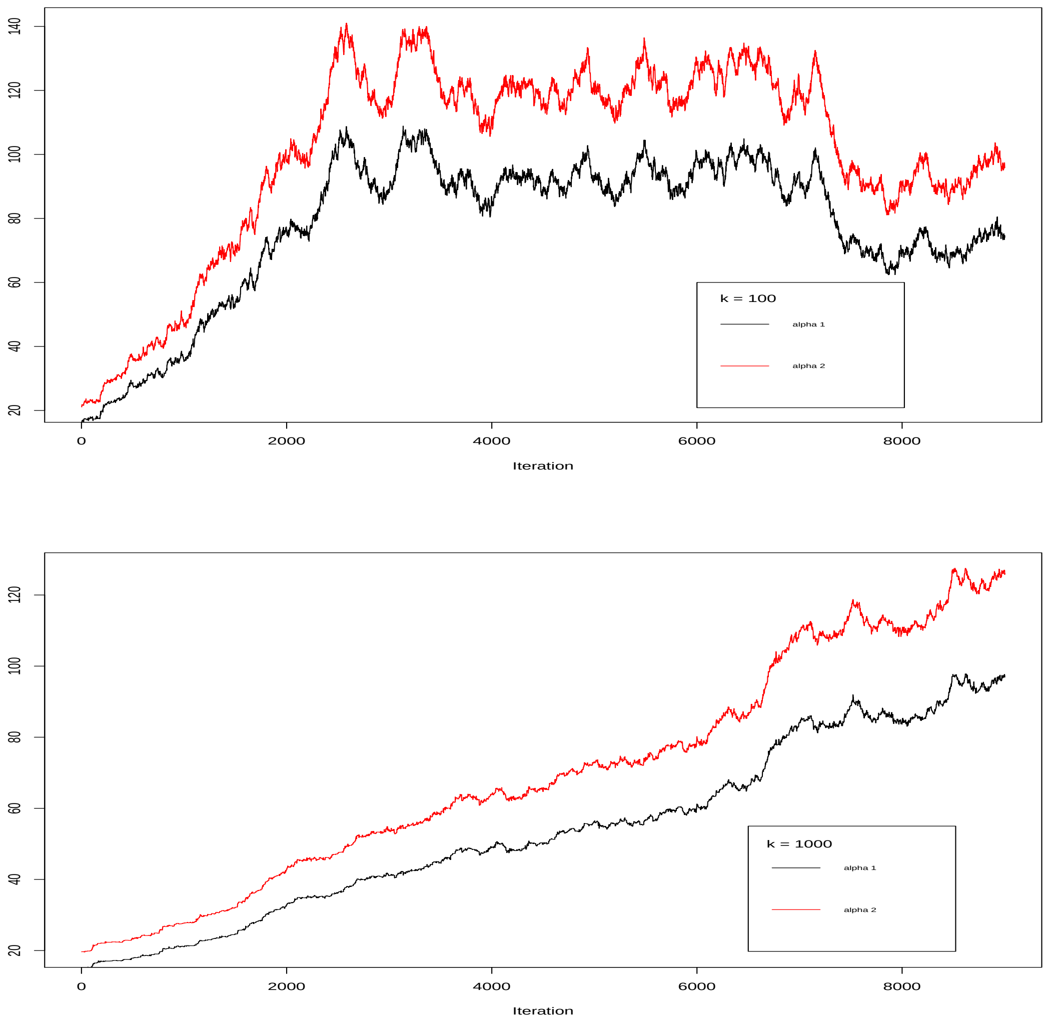

Figure 5.1: Iterates obtained for α_1, α_2 from the posterior density using Gibbs sampler for cloned Shunter's accident data with $k = 100, 1000$

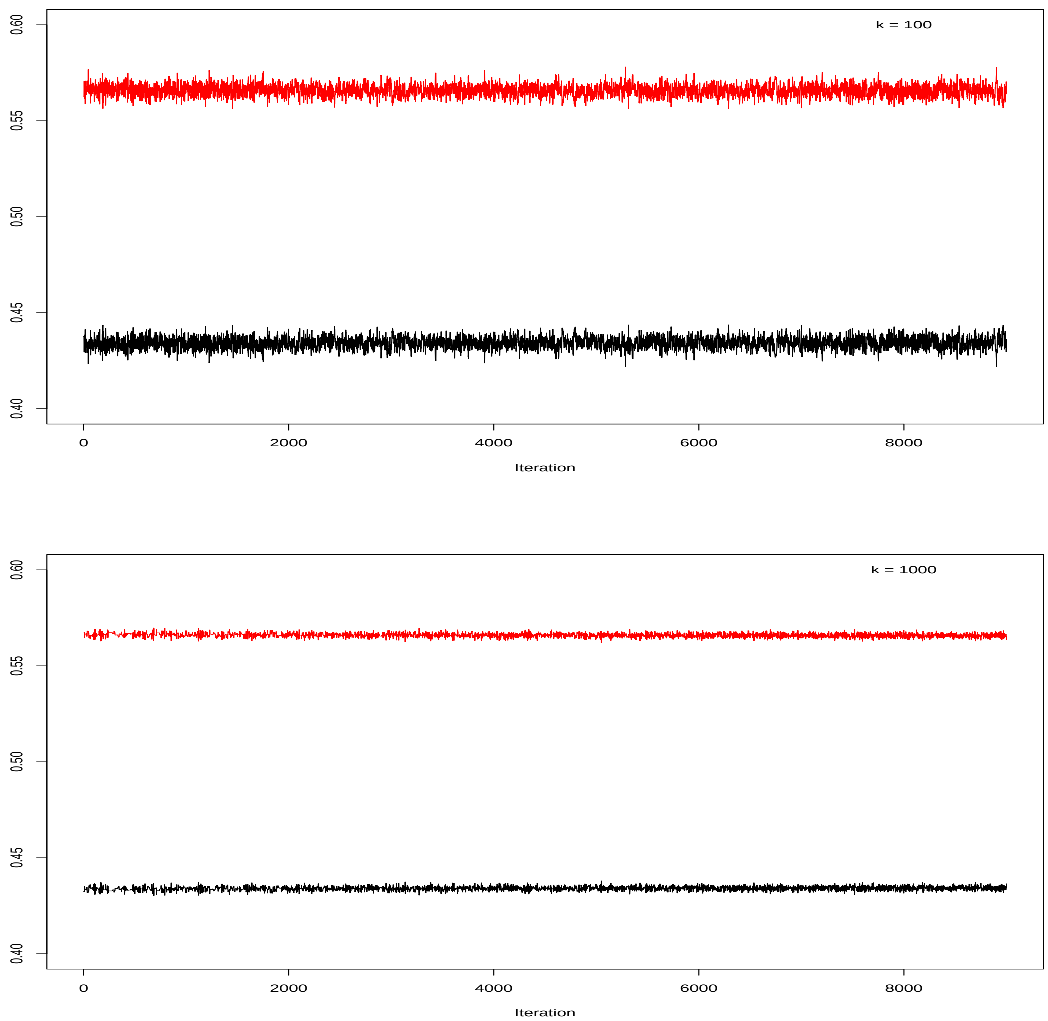

Figure 5.2: Plots for $\phi_1 = \frac{\alpha_1}{\alpha}, \phi_2 = \frac{\alpha_1}{\alpha}$ from samples generated using Gibbs sampler for cloned Shunter's accident data with $k = 100, 1000$

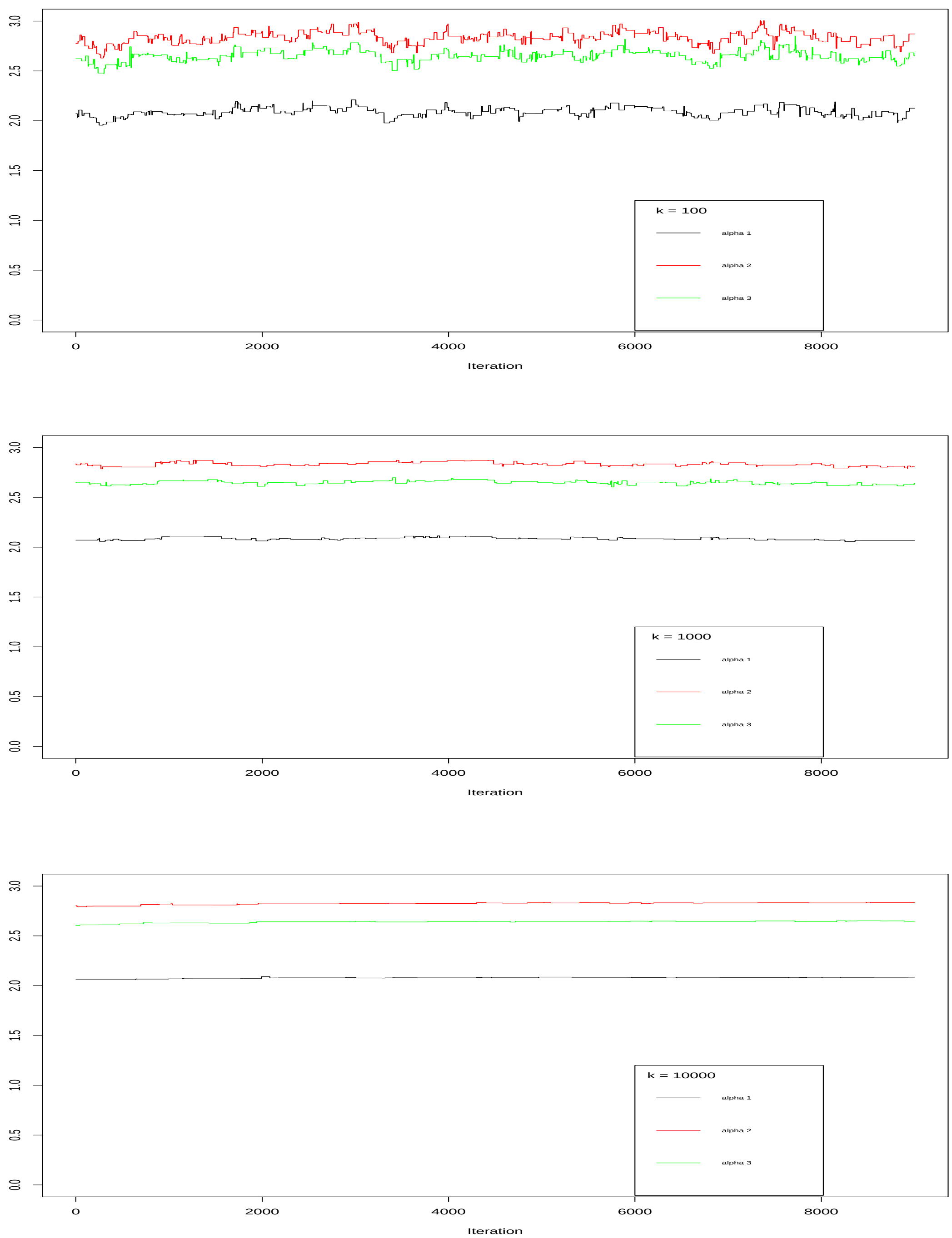

Figure 5.3: Iterates obtained for $\alpha_1, \alpha_2, \alpha_3$ from the posterior density using Gibbs sampler for cloned Aitchison's trivariate data with $k = 100, 1000, 10000$

5.3 Bayesian Inference

We consider the following priors for the α_i's :

$$\pi(\alpha_i) \propto \frac{P_i(\alpha_i)}{Q_i(\alpha_i)} e^{-\beta\alpha_i}, \quad i = 1, 2, \ldots, d, \tag{5.7}$$

where $P_i(\alpha_i)$ and $Q_i(\alpha_i)$ are polynomials in α_i's and β is any positive real number.

Proposition 5.3.1. *The class of priors given in (5.7) is semiconjugate prior for the Polya$(t; \alpha_1, \ldots, \alpha_d)$ distribution.*

Proof. The joint posterior density of the α_i's is

$$\pi(\alpha_1, \ldots, \alpha_d | (x_{1j}, \ldots, x_{dj})_{j=1}^n) \propto \pi(\alpha_i) \left(\prod_{j=1}^n \frac{\prod_{i=1}^d (\alpha_i)_{x_{ij}}}{(\alpha)_{t_j}} \right).$$

Now, the full posterior marginal density of α_i is

$$\begin{aligned} \pi(\alpha_i | \alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_d, (x_{1j}, \ldots, x_{dj})_{j=1}^n) &\propto \pi(\alpha_i) \left(\prod_{j=1}^n \frac{(\alpha_i)_{x_{ij}}}{(\alpha)_{t_j}} \right) \\ &\propto \frac{P_i(\alpha_i)}{Q_i(\alpha_i)} e^{-\beta\alpha_i} \left(\prod_{j=1}^n \frac{(\alpha_i)_{x_{ij}}}{(\alpha)_{t_j}} \right), \\ &\propto \frac{P_i(\alpha_i) \prod_{j=1}^n (\alpha_i)_{x_{ij}}}{Q_i(\alpha_i) \prod_{j=1}^n (\alpha)_{t_j}} e^{-\beta\alpha_i} \end{aligned}$$

Hence, we have,

$$\pi(\alpha_i | \alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_d, (x_{1j}, \ldots, x_{dj})_{j=1}^n) \propto \frac{P_i'(\alpha_i)}{Q_i'(\alpha_i)} e^{-\beta\alpha_i}, \tag{5.8}$$

where $P_i'(\alpha_i) = P_i(\alpha_i) \prod_{j=1}^n (\alpha_i)_{x_{ij}}$ and $Q_i'(\alpha_i) = Q_i(\alpha_i) \prod_{j=1}^n (\alpha)_{t_j}$ are polynomials in α_i. □

We can use Gibbs sampling scheme to sample from the joint posterior density of the α_i's.

The samples generated, then, can be used to estimate functions of α_i's such as posterior moments of α_i's *i.e.* $\mathbb{E}(\alpha_i^r|(x_{1j},\ldots,x_{dj})_{j=1}^n)$. We can thus obtain Bayes point estimators for α_i's.

5.3.1 Illustration

We apply the Bayesian methodology suggested above to the Shunter's Accident data to obtain posterior means of α_i's as Bayes point estimators.

We consider $\pi(\alpha_i) = Gamma(3,1)\,, i = 1,2,\ldots,d$. The samples from the joint posterior density are generated using Gibbs sampler. Fig. 5.4 shows the marginal posterior densities for α_1 and α_2. Table 5.3 gives posterior expectations of α_1 and α_2.

	α_1	α_2
Posterior mean	4.0546	5.2617
Posterior s.d.	1.1891	1.5226

Table 5.3: Estimates of α_i's for Shunter's Accident data

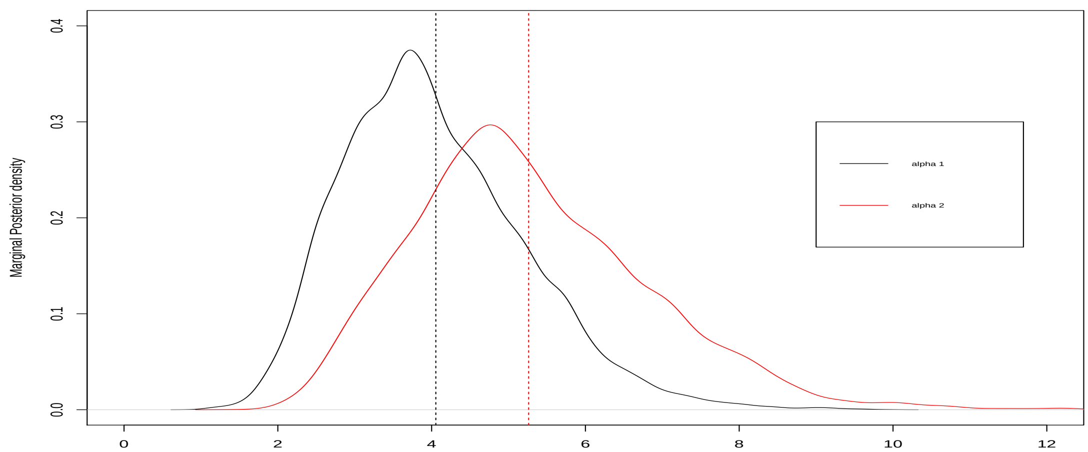

Figure 5.4: Marginal Posterior densities for α_1, α_2 for Shunter's Accident bivariate data

Chapter 6

Parameter estimation for the Negative Binomial Sums and Polya Shares Model

This chapter addresses the main goal of the thesis. Based on the estimators obtained in Chapter 4 and Chapter 5, here we provide the estimators for the parameters in Negative binomial sums and Polya shares model 3.18. The limiting distribution obtained in 3.23 is also considered. We also provide applications of the model to real datasets.

6.1 Introduction

We return to the original problem of the parameter estimation for the Negative binomial sums and Polya shares model with the probability mass function

$$p(m_1, m_2, \ldots, m_d) = \frac{\prod_{i=1}^{d}(\alpha_i)_{m_i}}{(\alpha)_{m_1+\cdots+m_d}} \frac{(a)_{m_1+\cdots+m_d}}{m_1!m_2!\cdots m_d!} \theta^a (1-\theta)^{m_1+\cdots+m_d} \tag{6.1}$$

where $\alpha = \sum_{i=1}^{d} \alpha_i$. We also consider the limiting distribution obtained as $\alpha_1, \ldots, \alpha_d \to \infty$ such that $\frac{\alpha_i}{\alpha} \to \phi_i$ with the probability mass function

$$p(m_1, m_2, \ldots, m_d) = \frac{(a)_{m_1+\cdots+m_d}}{m_1!\cdots m_d!} \left(\prod_{i=1}^{d} \phi_i^{m_i}\right) \theta^a (1-\theta)^{m_1+\cdots+m_d} \tag{6.2}$$

Let $(m_{11}, \cdots, m_{d1}), \cdots, (m_{1n}, \cdots, m_{dn})$ be n independent observations from 6.1. We denote these observations as $\boldsymbol{M} = \{(m_{11}, \ldots, m_{d1}), \cdots, (m_{1n}, \ldots, m_{dn})\}$. Let $m_{i.} = \frac{1}{n}\sum_{j=1}^{n} m_{ij}$, $s_{ij} = \frac{1}{n}\sum_{k=1}^{n}(m_{ik} - m_{i.})(m_{jk} - m_{j.})$ and $t_j = \sum_{i=1}^{d} m_{ij}$.

6.2 Method of Moments estimators

There are many challenges in obtaining method of moments estimators for the model (6.1) due to large number of parameters. Besides, as seen in Section 4.3, the method of moments estimators may not give a feasible solution for Negative binomial distribution. As mentioned by Jones and Marchand in [17], here we provide estimators when $d = 2$. Let $(m_{11}, m_{21}), \cdots, (m_{1n}, m_{2n})$ be n independent observations from bivariate discrete distributions with probability mass function

$$p(m_1, m_2) = \frac{(\alpha_1)_{m_1}(\alpha_2)_{m_2}}{(\alpha_1 + \alpha_2)_t}\frac{(a)_t}{m_1!m_2!}\theta^a(1-\theta)^t, \quad t = m_1 + m_2 \tag{6.3}$$

We denote $t_j = m_{1j} + m_{2j}, \quad j = 1, 2, \cdots, n$. Let $\hat{\mu_1} = \frac{\sum_{i=1}^{n} t_i}{n}$ and $\hat{\mu_2} = \frac{\sum_{i=1}^{n} t_i^2}{n}$. From (4.5), we have

$$\hat{a} = \frac{\hat{\mu_1}^2}{\hat{\mu_2} - \hat{\mu_1} - \hat{\mu_1}^2}, \quad \hat{\theta} = \frac{\hat{\mu_1}}{\hat{\mu_2} - \hat{\mu_1}^2}. \tag{6.4}$$

Comparing theoretical (refer to eq. 3.19 and eq. 3.20) and sample moments of m_i's ,

$$m_{1.} = \frac{\hat{a}(1-\hat{\theta})}{\hat{\theta}}\frac{\hat{\alpha}_1}{\hat{\alpha}} = \hat{\mu}_1\frac{\hat{\alpha}_1}{\hat{\alpha}}, \quad m_{2.} = \frac{\hat{a}(1-\hat{\theta})}{\hat{\theta}}\frac{\hat{\alpha}_2}{\hat{\alpha}} = \hat{\mu}_1\frac{\hat{\alpha}_2}{\hat{\alpha}}, \quad s_{12} = \frac{\hat{a}(1-\hat{\theta})^2}{\hat{\theta}^2\hat{\alpha}^2(1+\hat{\alpha})}\{(\hat{\alpha} - \hat{a})\hat{\alpha}_1\hat{\alpha}_2\},$$

we get,

$$\hat{\alpha_1} = \frac{\hat{\alpha}m_{1.}}{\hat{\mu_1}} \quad , \quad \hat{\alpha_2} = \frac{\hat{\alpha}m_{2.}}{\hat{\mu_1}} \quad , \quad \hat{\alpha} = \frac{\hat{a}^2(1-\hat{\theta})^2 m_{1.}m_{2.} + s_{12}\hat{\theta}^2\hat{\mu}_1^2}{\hat{a}(1-\hat{\theta})^2 m_{1.}m_{2.} - s_{12}\hat{\theta}^2\hat{\mu}_1^2}$$

Note that $\hat{\alpha} > 0 \quad$ if and only if $\quad \hat{a}(1-\hat{\theta})^2 m_{1.}m_{2.} > s_{12}\hat{\theta}^2\hat{\mu}_1^2$.

We can't generalize this method for higher dimensions as there are less number of parameters than the moments equations to compare. For instance, when $d = 3$, one can estimate

a and θ in a similar way. But, to estimates $\alpha_1, \alpha_2, \alpha_3$, we have six independent terms in the covariance matrix. It is not known which equations should be preferred for the estimation purpose.

For the limiting distribution (6.2), the method of moments estimators are

$$\hat{\phi}_i = \frac{m_{i.}}{\hat{\mu}_1} \quad ,i = 1, 2, \cdots, d \tag{6.5}$$

6.3 Maximum Likelihood Estimators

The log-likelihood function based on n independent observations from the Negative binomial sums and Polya shares distribution is given as

$$\begin{aligned} logL(a, \theta, \alpha_1, \cdots, \alpha_d | M) &= nalog(\theta) + \sum_{j=1}^{n} t_j log(1-\theta) + \sum_{j=1}^{n} \left(\sum_{k=0}^{t_j - 1} log(a+k) \right) \\ &+ \sum_{i=1}^{d} \sum_{j=1}^{n} \left(\sum_{k=0}^{m_{ij}-1} log(\alpha_i + k) \right) - \sum_{j=1}^{n} \left(\sum_{k=0}^{t_j-1} log(\alpha + k) \right) - \sum_{i=1}^{d} \sum_{j=1}^{n} log(m_{ij}!) \end{aligned} \tag{6.6}$$

Note that the terms involving a, θ and $\alpha_i, \cdots, \alpha_d$ are separated in the log-likelihood equation. Hence, the score functions are independent. As seen in the Chapter 4 and Chapter 5, MLE for a as well as α_i's may not exist. Whenever the solutions exist, methods developed in the Chapter 4 and Chapter 5 can be used.

6.4 Bayesian Inference

Here, we suggest a general setup for the Bayesian inference for the negative binomial sums and shares model. We consider the priors for the parameters as follows:

1. Polya parameter a is distributed as a *Gammapoly* distribution introduced in (4.7) *i.e.*, $\pi(a) \propto \Phi(a) e^{-\gamma a}$;
2. $\theta \sim Beta(\beta_1, \beta_2)$ *i.e.* $\pi(\theta) = \frac{1}{B(\beta_1, \beta_2)} \theta^{\beta_1 - 1}(1-\theta)^{\beta_2 - 1} \mathbb{I}_{(0,1)}(\theta)$;

3. $\pi(\alpha_i) \propto \frac{P_i(\alpha_i)}{Q_i(\alpha_i)} e^{-\beta\alpha_i} \quad , \quad i = 1, 2, \ldots, d.$

Using the algorithms proposed in previous chapters, we can obtain estimates for the posterior distribution of the parameters.

Bayesian Inference in case of the limiting distribution 6.2:

We consider the following priors:

1. $\pi(a) \propto \Phi(a)e^{-\gamma a}$;
2. $\theta \sim Beta(\beta_1, \beta_2)$ *i.e.* $\pi(\theta) = \frac{1}{B(\beta_1,\beta_2)}\, \theta^{\beta_1-1}(1-\theta)^{\beta_2-1}\mathbb{I}_{(0,1)}(\theta)$;
3. $\phi_1, \cdots, \phi_d \sim Dirichlet(\delta_1, \delta_2, \cdots, \delta_d)$ *i.e.*

$$\pi(\phi_1, \cdots, \phi_d) = \frac{\Gamma(\delta)\prod_{i=1}^{d} \phi_i^{\delta_i-1}}{\prod_{i=1}^{d}\Gamma(\delta_i)},$$

   where $\delta = \sum_{i=1}^{d} \delta_i$.

Then the joint posterior density is

$$\begin{aligned}\pi(a, \theta, \phi_1, \phi_2, \cdots, \phi_d|\boldsymbol{M}) &\propto \prod_{j=1}^{n} p(m_{1j}, \ldots, m_{dj})\Phi(a)e^{-\gamma a}\theta^{\beta_1-1}(1-\theta)^{\beta_2-1}\prod_{i=1}^{d}\phi_i^{\phi_i-1} \\ &\propto \prod_{j=1}^{n}(a)_{t_j}\Phi(a)e^{-\gamma a}\theta^{na+\beta_1-1}\,(1-\theta)^{\sum\limits_{i=1}^{d} m_{ij}+\beta_2-1}\prod_{i=1}^{d}\phi_i^{\sum_{j=1}^{n} m_{ij}+\delta_i-1},\end{aligned}$$

and thus

$$\pi(\phi_1, \phi_2, \cdots, \phi_d|\boldsymbol{M}) \propto \prod_{i=1}^{d}\phi_i^{\sum_{j=1}^{n} m_{ij}+\delta_i-1}.$$

Hence,

$$\phi_1, \phi_2, \cdots, \phi_d|\boldsymbol{M} \sim Dirichlet\left(\sum_{j=1}^{n} m_{ij} + \delta_1, \ldots, \sum_{j=1}^{n} m_{dj} + \delta_d\right). \tag{6.7}$$

6.5 Shunter's Accident Data

We fit the limiting model 6.2 to bivariate Shunters' accident data (Table A.1). Here are the Bayes estimators withh respect to the following prior:

$$a \sim Gamma(3,1), \quad \theta \sim Beta(1,1), \quad \phi_1, \phi_2 \sim Dirichlet(1,1).$$

Parameters	Posterior means	Posterior s.d.	95% HPD intervals
a	3.5539	1.0876	(1.9161 , 6.2798)
θ	0.6004	0.0712	(0.4591 , 0.7380)
ϕ_1	0.4652	0.4329	(0.0822 , 0.8847)

Table 6.1: Fit of limiting distribution to Shunters' accident data using Bayes estimators.

$m_1 \backslash m_2$	0	1	2	3	4	5	6	7+
0	19.90	15.12	7.35	2.91	1.02	0.33	0.10	0.03
1	13.15	12.80	7.59	3.55	1.43	0.52	0.18	0.06
2	5.57	6.61	4.63	2.49	1.14	0.46	0.17	0.06
3	1.92	2.68	2.17	1.32	0.67	0.30	0.13	0.05
4	0.58	0.94	0.86	0.59	0.33	0.16	0.07	0.03
5	0.16	0.30	0.31	0.23	0.14	0.08	0.04	0.02
6	0.04	0.09	0.10	0.08	0.06	0.03	0.02	0.01
7+	0.01	0.02	0.03	0.03	0.02	0.01	0.01	0.00

Table 6.2: Expected counts under fitting of limiting model to Shunters' accident data

6.6 Aitchison's Trivariate Bacterial Count Data

We fit the Negative binomial sums and Polya shres model (6.1) to the trivariate bacterial dataset (Table A.2). Here are the estimators for the parameters obtained using maximum likelihood estimation and Bayes inference:

For Bayesian inference, we consider following priors:

$$a \sim Gamma(5,1), \quad \theta \sim Beta(1,1), \quad \alpha_i \sim Gamma(3,1), \quad i = 1,2,3.$$

Parameters	MLE
a	16.0029
θ	0.4734
α_1	2.0781(0.0072)
α_2	2.8248(0.0101)
α_3	2.6402(0.0101)

Table 6.3: Fit Negative Binomial Sums and Polya Shares model to Aitchison's trivariate data using MLE. The bracketed quantities denote the standard error.

Parameters	Posterior means	Posterior s.d.	95% HPD intervals
a	9.3426	1.9677	(6.0445 , 13.6789)
θ	0.3409	0.0479	(0.2501 , 0.4384)
α_1	2.2195	0.4329	(1.4617 , 3.1715)
α_2	2.9945	0.5813	(2.0141 , 4.2269)
α_3	2.8104	0.5602	(1.8638 , 4.0865)

Table 6.4: Fit Negative binomial Sums and Polya shares model to Aitchison's trivariate data using Bayes estimators.

Chapter 7

On a proper Bayes, but inadmissible estimator

We present an example of a proper Bayes point estimator which is inadmissible. It occurs for a negative binomial model with shape parameter a, probability parameter p, prior densities of the form $\pi(a,p) = \beta\, g(a)\, (1-p)^{\beta-1}$, and for estimating the population mean $\mu = a(1-p)/p$ under squared error loss. Other intriguing features are exhibited, one such feature is the constancy of the Bayes estimator with respect to the choice of g, including degenerate or known a cases.

7.1 Introduction

Bayesian methods are intimately linked to statistical decision theory, they possess desirable theoretical properties, such as coherence and, in general, good frequentist risk properties. However, proper Bayes estimators need not be admissible. So much and more is known (see for instance Berger, 1985 [6]; Lehmann, 1983 [21]; and the discussion following Theorem 7.2.1 below). Nonetheless, such examples remain surprising and instructive, especially when they occur in simple situations that are also relevant in practice. We report and comment on such a situation that occurred recently in studying Bayesian posterior analysis for a negative binomial model. Moreover, the example which we present exhibits other intriguing features, one such feature is the constancy of a Bayes point estimator with respect to a large and dispersed class of priors.

As implied by the following well-known result (e.g., Ferguson, 1968 [10]; Lehmann, 1983 [21]), the inadmissibility of a proper Bayes estimator can only occur when the Bayes risk is infinite.

Lemma 7.1.1. *Consider model $X \sim p_\theta$, $\theta \in \Theta$, and the problem of estimating $\tau(\theta)$ under loss $L(\theta, \delta)$. Let $\delta_\pi(X)$ be a unique* [1] *Bayes estimator for a proper prior density π. Then, the estimator $\delta_\pi(X)$ is admissible when $r_\pi(\delta_\pi) < \infty$, with $r_\pi(\delta)$ the Bayes risk of δ given by*

$$r_\pi(\delta) \;=\; \int_\Theta \mathbb{E}\, L\left(\theta, \delta(x)\right) \, \pi(\theta)\, d\theta\,.$$

We point out that Lemma 1.1 holds more generally, in particular for cases where π is a probability mass function.

Proof. Suppose $\delta_\pi(X)$ is not admissible and dominated by another estimator $\delta_\pi^{'}(X)$. Hence, we have

$$r_\pi(\delta_\pi) \le r_\pi(\delta_\pi^{'}) = \int_\Theta \mathbb{E}\, L\left(\theta, \delta(x)^{'}\right) \, \pi(\theta)\, d\theta \;\le \int_\Theta \mathbb{E}\, L\left(\theta, \delta_\pi(x)\right) \, \pi(\theta)\, d\theta \;= r_\pi(\delta_\pi) \tag{7.1}$$

But, Bayes estimator is unique. Hence, $\delta_\pi(X)$ is admissible. $\square$

7.2 The example

Let $X_1, \ldots, X_n$, $n \ge 2$, be independently distributed $NB(a, p)$, $a > 0, p \in (0,1)$, with common marginal probability mass function

$$\mathbb{P}(X_i = t) \;=\; \frac{(a)_t}{t!}\, p^a \,(1-p)^t \,\mathbb{I}_{\mathbb{N}}(t)\,, i = 1, \ldots, n, \tag{7.2}$$

with ascending factorial $(a)_t \;=\; \prod_{j=0}^{t-1}(a+j)$ for $t \ge 1$, $(a)_0 = 1$. We take both a and p to be unknown. The negative binomial model is one of the better known and appealing models for count data, in particular for over-dispersed data, with mean lower than the variance, as

$$\mathbb{E}(X_1) \;= \mu \;= a\,\frac{1-p}{p} < \sigma^2 \;=\; \mathbb{V}(X_1) \;=\; a\,\frac{1-p}{p^2}\,,$$

for all $a > 0, p \in (0,1)$. An alternative and appealing representation of (2.1), expressed in terms of the mean μ and the overdispersion parameter $k = a$, with $\sigma^2 = \mu(1 + \mu/k)$, is given by

[1] Up to an equivalence

$$\mathbb{P}(X_i = t) \;=\; \frac{(k)_t}{t!}\,(\frac{k}{\mu+k})^k\,(\frac{\mu}{\mu+k})^t\,\mathbb{I}_{\mathbb{N}}(t)\,,$$

Now, consider estimating μ under squared error loss $(\delta-\mu)^2$ with Bayesian estimators given by $\mathbb{E}(\mu|x_1,\ldots,x_n)$ as long as $\mathbb{E}(\mu^2|x_1,\ldots,x_n) < \infty$ for all $x_1,\ldots,x_n$. Our main example is the following and relates to a joint prior for (a,p) which factorizes into independent components $a \sim g$ and $p \sim Beta(1,\beta)$.

Theorem 7.2.1. *Consider $X_1,\ldots,X_n \sim$ NB(a,p) as in (7.2) with prior density*

$$\pi_g(a,p) \;=\; \beta\, g(a)(1-p)^{\beta-1} I_{(0,1)}(p)\,,$$

with $\beta > 0$ and g being a density (or a probability mass function) such that $\int_1^\infty g(a)da \;=\; 1$ (or $\sum_{j=1}^\infty g(j) = 1$). Then, the Bayes estimator of $\mu \;=\; a\,(1-p)/p$ for loss $(\delta-\mu)^2$, and with respect to prior $\pi_g(a,p)$ is given by $\delta_{\pi_g}(x_1,\ldots,x_n) \;=\; \frac{\sum_{i=1}^n x_i+\beta}{n}$, irrespective of g. Furthermore, δ_{π_g} is inadmissible and dominated by the unbiased estimator $\delta_0(x_1,\ldots,x_n) \;=\; \frac{\sum_{i=1}^n x_i}{n}$.

Proof. In terms of frequentist risk $R((a,p),\delta) \;=\; \mathbb{E}_{a,p}\left\{\delta(X_1,\ldots,X_n)-\mu\right\}^2$ of an estimator $\delta(X_1,\ldots,X_n)$, we have

$$\begin{aligned} R((a,p),\delta_0) \;&=\; \mathbb{V}(\delta_0) \;=\; \sigma^2/n \\ R((a,p),\delta_{\pi_g}) \;&=\; \mathbb{V}(\delta_0) + (\text{Bias of } \delta_{\pi_g})^2 \;=\; \sigma^2/n \,+\beta^2/n^2\,, \end{aligned} \tag{7.3}$$

which shows indeed that δ_0 dominates δ_{π_g}.

Now, for evaluating δ_{π_g}, we have the likelihood

$$p(x_1,\ldots,x_n|a,p) \;=\; \left\{\prod_i \frac{(a)_{x_i}}{x_i!}\right\}\; p^{na}\,(1-p)^{\sum_i x_i}\,\prod_i \mathbb{I}_{\mathbb{N}}(x_i)\,,$$

yielding the posterior density

$$(a,p)|x_1,\ldots,x_n \;\propto\; \prod_i (a)_{x_i}\, g(a)\, p^{na}(1-p)^{\sum_i x_i+\beta-1}\,;a\geq 1, 0<p<1\,.$$

Hereafter, we pursue for g being a density. The discrete case (i.e., g is a probability mass function) can be studied analogously. From the above posterior density, we obtain

$$
\begin{aligned}
\mathbb{E}\left(\mu | x_1, \ldots, x_n\right) &= \mathbb{E}\left(a \left.\frac{1-p}{p}\right| x_1, \ldots, x_n\right) \\
&= \frac{\int_1^\infty a \prod_i (a)_{x_i}\, g(a) \int_0^1 p^{na-1}(1-p)^{\sum_i x_i+\beta} dp\, d\nu(a)}{\int_1^\infty \prod_i (a)_{x_i}\, g(a) \int_0^1 p^{na}(1-p)^{\sum_i x_i+\beta-1} dp\, d\nu(a)} \\
&= \frac{\int_1^\infty a \prod_i (a)_{x_i}\, g(a) \frac{\Gamma(na)\,\Gamma(\sum_i x_i+\beta+1)}{\Gamma(na+\sum_i x_i+\beta+1)}\, d\nu(a)}{\int_1^\infty \prod_i (a)_{x_i}\, g(a) \frac{\Gamma(na+1)\,\Gamma(\sum_i x_i+\beta)}{\Gamma(na+\sum_i x_i+\beta+1)}\, d\nu(a)} \\
&= \left(\frac{\sum_i x_i \,+\, \beta}{n}\right) \frac{\int_1^\infty \prod_i (a)_{x_i}\, g(a) \frac{\Gamma(na+1)}{\{\Gamma(na+\sum_i x_i+\beta+1)\}}\, d\nu(a)}{\int_1^\infty \prod_i (a)_{x_i}\, g(a) \frac{\Gamma(na+1)}{\Gamma(na+\sum_i x_i+\beta+1)}\, d\nu(a)} \qquad (7.4) \\
&= \frac{\sum_i x_i \,+\, \beta}{n},
\end{aligned}
$$

as stated.

For the above, an application of Stirling's formula, i.e., $\Gamma(z+1) \sim z^{z+1/2} e^{-z}\sqrt{2\pi}$, with $a(z) \sim b(z)$ meaning $\lim_{z\to\infty} a(z)/b(z) = 1$, implies that the integrals converge. Indeed, we have

$$
\frac{\Gamma(na+1)}{\Gamma(na+1+\beta+\sum_i x_i)} \sim (\frac{e}{na})^t \left(\frac{na}{na+t+\beta}\right)^{na+t+1/2} \left(\frac{e}{na+t+\beta}\right)^{\beta},
$$

with $t = \sum_i x_i$. With $\prod_i (a)_{x_i} \sim a^t$, we obtain

$$
\prod_i (a)_{x_i} \frac{\Gamma(na+1)}{\Gamma(na+1+\beta+\sum_i x_i)} \sim (\frac{e}{n})^t \left(\frac{na}{na+t+\beta}\right)^{na+t+1/2} \left(\frac{e}{na+t+\beta}\right)^{\beta},
$$

Since the above is bounded for large a, in fact converging to 0 as $a \to \infty$, we infer the (2.3) is well defined.

To conclude the proof, we require the posterior variance $\mathbb{V}(\mu|x_1, \ldots, x_n)$ to exist, i.e., $\mathbb{E}(\mu^2|x_1, \ldots, x_n) < \infty$, and we show that this to be the case under the given conditions on g. Indeed, proceeding with a decomposition as above, we have

$$\begin{aligned}
\mathbb{E}(\mu^2|x_1,\ldots,x_n) &= \mathbb{E}(a^2\,\frac{(1-p)^2}{p^2}|x_1,\ldots,x_n) \\
&= \frac{\int_1^\infty a^2\,\prod_i (a)_{x_i}\,g(a)\int_0^1 p^{na-2}(1-p)^{\sum_i x_i+\beta+1}dp\,da}{\int_1^\infty \prod_i (a)_{x_i}\,g(a)\int_0^1 p^{na}(1-p)^{\sum_i x_i+\beta-1}dp\,da} \\
&= (\sum_i x_i+\beta)\,(\sum_i x_i+\beta+1)\,\frac{\int_1^\infty a^2\,\prod_i (a)_{x_i}\,g(a)\,\frac{\Gamma(na-1)}{\Gamma(na+\sum_i x_i+\beta+1)}\,da}{\int_1^\infty \prod_i (a)_{x_i}\,g(a)\,\frac{\Gamma(na+1)}{\Gamma(na+\sum_i x_i+\beta+1)}\,da}\,,
\end{aligned}$$

Finally, another use of Stirling's formula as above establishes that the integrals above converge. This proves that $\mathbb{E}(\mu^2|x_1,\ldots,x_n) < \infty$ and completes the proof. $\square$

Remark 7.2.1. *As implied by the above Theorem and Lemma 1.1., the Bayes risk of δ_g must be equal to $+\infty$. Indeed, we have for the continuous case and using risk expression (2.2):*

$$\begin{aligned}
r_{\pi_g}(\delta_{\pi_g}) &= \int_1^\infty\int_0^1 \left(\frac{a(1-p)}{np^2}+\frac{\beta^2}{n^2}\right) g(a)\,\beta(1-p)^{\beta-1}dp\,da \\
&= \frac{\beta^2}{n^2}+\frac{\beta}{n}\left(\int_1^\infty ag(a)da\right)\int_0^1\frac{(1-p)^\beta}{p^2}\,dp \\
&= +\infty\,.
\end{aligned}$$

Theorem 7.2.1 exhibits the motivating purpose and the main feature of this note. It adds to a small collection of known examples where a proper Bayes estimator is inadmissible. Earlier examples appear in Lehmann (1983, page 270) with a Gamma model and inverse-Gamma prior, as well as in Berger (1985) (see Robert, 2001, Section 8.2, who reports on both of these examples). In the latter case, one takes $X|\theta \sim N(\theta,1)$, prior $\theta \sim N(0,1)$ and weighted squared-error loss $e^{3\theta^2/4}\,(\delta-\theta)^2$ for estimating θ. Calculations yield the Bayes estimator $\delta_\pi(X) = \mathbb{E}(e^{3\theta^2/4}\,\theta|X)/\mathbb{E}(e^{3\theta^2/4}\,|X) = 2X$. Clearly, with larger bias in absolute value and larger variance than the unbiased estimator $\delta_0(X) = X$, the frequentist risk of $\delta_\pi(X)$ is quite poor, as represented by the ratio of risks $\frac{R(\theta,\delta_\pi)}{R(\theta,\delta_0)} = (4+\theta^2)$.

Theorem 7.2.1 does exhibit other surprises though. Although both the posterior distribution and variance of $\mu = a(1-p)/p$ do depend on the choice of g for the prior $(a,p) \sim \beta\,g(a)(1-p)^{\beta-1}$, the posterior expectation obtained is independent of g, and whether or not the prior is discrete or continuous for instance. Moreover, it has to be the case for degenerate a, in other words cases where a is known and $p \sim Beta(1,\beta)$. Indeed, this can be verified directly by deriving the posterior $p|x_1,\ldots,x_n \sim Beta(na+1,\sum_i x_i+\beta)$, from which one obtains $\mathbb{E}(a(1-p)/p\,|\,x_1,\ldots,x_n) = \frac{\sum_i x_i+\beta}{n}$. Despite its greater simplicity, to our knowledge, this known a case result has not been reported on before. The Bayes estimator here is still inadmissible and dominated by $\bar{X}$ as an estimate of

$\mu = a(1-p)/p$. Furthermore, the unbiased estimator $\bar{X}$ is itself inadmissible under squared error loss and dominated by $\frac{na}{na+1}\bar{X}$ (Ferguson, 1967, [10], problem 12, page 86).

Remark 7.2.2. *If one happens to derive the result of Theorem 7.2.1 for the known $a > 1$ case, one then can see that Theorem 7.2.1's expression for the Bayes estimator will hold since*

$$\begin{aligned}
\mathbb{E}(\mu|X_1,\ldots,X_n) &= \mathbb{E}^{a|X_1,\ldots,X_n}\left[\mathbb{E}\left(\frac{a(1-p)}{p}\middle| X_1,\ldots,X_n,a\right)\right] \\
&= \mathbb{E}^{a|X_1,\ldots,X_n}\left[\frac{\sum_i X_i + \beta}{n}\right] \\
&= \frac{\sum_i X_i + \beta}{n}\,.
\end{aligned}$$

7.3 Concluding Remarks

We have provided an original example of a proper Bayes estimator which is inadmissible and which arises for both a two-parameter and one-parameter negative binomial model. More specifically, we came about the finding in an ongoing work aiming to provide Bayesian inference results for the $NB(a,p)$ model. Although, the negative binomial model arises in many applications, our ongoing work originates with recent work of Jones and Marchand (2019) who introduce a Sums-and-Shares model for multivariate discrete data, consisting for instance of a negative binomial distributed sum $T \in \mathbb{N}$, and then randomly allocated shares $M_1,\ldots,M_d$ into d categories such that $\sum_{j=1}^{d} M_j = T$ [17]. In such a case, as well as for the earlier known examples in the literature, the Bayes estimators are obtained in a coherent manner from a proper prior and by making trustworthy inferences from the posterior distribution. As described by Berger (1985, Section 4.8.1) [6], such an unsettling or paradoxical situation, which can only possibly happen when the Bayes risk is infinite (Lemma 7.1.1), is alleviated with the use of a bounded loss function for which the Bayes risk cannot be infinite.

Still, the surprise persists given the deep connections in statistical decision theory between complete classes of estimators and Bayes estimators. In the other direction, it has long been known that the collection of proper Bayes estimators is not large enough to generally contain all admissible estimators and, typically, one requires the inclusion of some generalized Bayes estimators. A well known example arises for the multivariate normal model with X for $X \sim N_p(\theta, I_p)$ and squared error loss $\|d-\theta\|^2$ for estimating θ. Here, X is generalized Bayes for the improper prior density $\pi(\theta) = 1$, admissible for $p = 1, 2$, but inadmissible for $p \geq 3$. There exist many deep findings related to the admissibility of generalized Bayes estimators (e.g., Rukhin, 1995 [30], as well as the references below, among others).

Chapter 8

Conclusion

This thesis addresses an important problem, namely the parameter estimation for the Negative binomial sums and Polya shares model. This model has a wide range of applications, for instance, analysis of insurance data, accidents' count data, joint species distribution models in community ecology, etc. The methods developed in this work can facilitate the use of the model in such situations for the estimation purpose.

Besides the development of the methods of parameter estimation for Negative binomial sums and Polya shares , the work has led to many intriguing outcomes. In Chapter 4, we constructed a more general Bayesian setup for a two-parameter Negative binomial distribution. There is not much literature available on this problem. As we have seen, the method of moments and maximum likelihood estimaton may lead to infeasible estimators. Bayesian inference for Negative binomial is a challenging problem due to the complexity of the likelihood function and hence the intractability of the posterior densities. A Bayesian approach proposed in this work uses a Gibbs sampler algorithm to sample from the joint posterior density. Our methodology can be used to provide closed-form expressions for full marginal densities. This also enables us to estimate predictive densities, which is itself an interesting problem. Besides, as seen in an illustrative example consisting of a tiny dataset, interesting probability density functions emerge in the form of marginal posterior densities. For instance, marginal distribution of a in (4.8.1) is a mixture of densities with the normalization constants as exponential integrals. In Chapter 5, for the Polya distribution, we suggest the use of a data cloning method to obtain MLE when they exist. Interestingly, even when the MLE solution is not finite, outcomes of the data cloning method can be used to estimate MLE solution for the limiting distribution.

We only consider a Negative binomial sums and Polya shares model obtained using sums and share strategy. The results are very promising for the application of the model. There are many other constructions possible using the strategy. For instance, if we consider distribution L in (3.9) to be *Gammapoly* distribution, we get the mixture of Negative binomial sums and Polya shares model. This approach provides more flexible way to model multivariate count data. The work presented in the thesis can be extended to explore the properties of this mixture model.

In Chapter 7, an original example of a proper Bayes estimator which is inadmissible is provided. It arises for both a two-parameter and one-parameter negative binomial model. This adds to the small list of such instances. Also, Bayes estimators have deep connections to the complete classes of estimators in statistical decision theory.

Bibliography

[1] A. M. Adelstein. Accident Proneness: A Criticism of the Concept Based Upon an Analysis of Shunters' Accidents. 115(3):354–410, 1952.

[2] J. Aitchison and C. H. Ho. The Multivariate Poisson-Log Normal Distribution. *Biometrika*, 76(4):643–653, 1989.

[3] F. J. Anscombe. Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. *Biometrika*, 37(3-4):358–382, 1950.

[4] Djilali Ait Aoudia and Éric Marchand. On a Simple Construction of a Bivariate Probability Function With a Common Marginal. *The American Statistician*, 68(3):170–173, 2014.

[5] Jorge Aragón, David Eberly, and Shelly Eberly. Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution. *Statistics and Probability Letters*, 15(5):375 – 379, 1992.

[6] James O Berger. *Statistical Decision Theory and Bayesian Analysis; 2nd ed.* Springer Series in Statistics. Springer, New York, 1985.

[7] Eric Bradlow, Bruce G. S Hardie, and Peter Fader. Bayesian inference for the negative binomial distribution via polynomial expansions. 11:189–201, 03 2002.

[8] A. Castañer, M.M. Claramunt, C. Lefèvre, and S. Loisel. Discrete Schur-constant models. *Journal of Multivariate Analysis*, 140:343 – 362, 2015.

[9] Lester B Dropkin. Some considerations on automobile rating systems utilizing individual driving records. *PCAS XLVI*, page 165, 1959.

[10] T. S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, Boston, 1967.

[11] R. A. Fisher. The Negative Binomial Distribution. *Annals of Eugenics*, 11(1):182–187, 1941.

[12] Marwa Hamza and Pierre Vallois. On Kummer's distribution of type two and a generalized beta distribution. *Statistics Probability Letters*, 118(C):60–69, 2016.

[13] P. Holgate. Estimation for the bivariate poisson distribution. *Biometrika*, 51(1-2):241–287, 1964.

[14] N. L. Johnson and C. A. Rogers. The moment problem for unimodal distributions. *Ann. Math. Statist.*, 22(3):433–439, 09 1951.

[15] N.L. Johnson, A.W. Kemp, and S. Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley, 2005.

[16] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete Multivariate Distributions*. Wiley series in probability and statistics. John Wiley & Sons, 2004.

[17] M.C. Jones and Eric Marchand. Multivariate discrete distributions via sums and shares. *Journal of Multivariate Analysis*, 171:83 – 93, 2019.

[18] S Kocherlakota and K Kocherlakota. *Bivariate Discrete Distributions*. Marcel Dekker Inc, 1992.

[19] Stéphane Laurent. Some Poisson mixtures distributions with a hyperscale parameter. *Braz. J. Probab. Stat.*, 26(3):265–278, 08 2012.

[20] Hyunju Lee and Ji Hwan Cha. On Two General Classes of Discrete Bivariate Distributions. *The American Statistician*, 69(3):221–230, 2015.

[21] E.L. Lehmann. *Theory of point estimation*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley, 1983.

[22] Subhash R. Lele, Khurram Nadeem, and Byron Schmuland. Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning. *Journal of the American Statistical Association*, 105(492):1617–1625, 2010.

[23] Bruce Levin and James Reeds. Compound Multinomial Likelihood Functions are Unimodal: Proof of a Conjecture of I. J. Good . *Ann. Statist.*, 5(1):79–87, 01 1977.

[24] D. V. Lindley. Fiducial Distributions and Bayes' Theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):102–107, 1958.

[25] G. S. Lingappaiah. Discrete generalized Liouville-Type distribution and related multivariate distributions. *Trabajos de Estadistica y de Investigacion Operativa*, 35(3):319–330, Oct 1984.

[26] Y. L. Lio. A Note On Bayesian Estimation for the Negative-Binomial Model. http://http://www.math.bas.bg/pliska/Pliska-19/Pliska-19-2009-207-216.pdf, 2009.

[27] Ian McHale and Phil Scarf. Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, 61(4):432–445, 2007.

[28] Roger B. Nelsen. *An Introduction to Copulas (Springer Series in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[29] E. Nkingi and J. Vrbik. Confidence Regions for Parameters of Negative Binomial Distribution. *ArXiv e-prints*, December 2016.

[30] Andrew L. Rukhin. Admissibility: Survey of a concept in progress. *International Statistical Review / Revue Internationale de Statistique*, 63(1):95–115, 1995.

[31] V. Savani and A. A. Zhigljavsky. Efficient Estimation of Parameters of the Negative Binomial Distribution. *Communications in Statistics - Theory and Methods*, 35(5):767–783, 2006.

[32] W. Simonsen. On the solution of a maximum-likelihood equation of the negative binomial distribution. *Scandinavian Actuarial Journal*, 1976(4):220–231, 1976.

[33] A. M. Walker. On the Asymptotic Behaviour of Posterior Distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(1):80–88, 1969.

[34] Y. Wang. Estimation problems for the two-parameter negative binomial distribution. *Statistics and Probability Letters*, 26(2):113 – 114, 1996.

[35] L. J. Willson, J. L. Folks, and J. H. Young. Complete sufficiency and maximum likelihood estimation for the two-parameter negative binomial distribution. *Metrika*, 33(1):349–362, Dec 1986.

Appendix A

Datasets

Here are the details of the datasets used in the thesis:

Shunters' Accident Data

This is a bivariate data set given by Adelstein (1952) in Table 11A in [1]. There are $n = 122$ observations of (m_1, m_2), where m_1 and m_2 denote the number of accidents for shunters on South African Railways during $1937 - 1942$ and $1943 - 1947$, respectively.

$m_1 \backslash m_2$	0	1	2	3	4	5	6	7+
0	21	18	8	2	1	0	0	0
1	13	14	10	1	4	1	0	0
2	4	5	4	2	1	0	1	0
3	2	1	3	2	0	1	0	0
4	0	0	1	1	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7+	0	1	0	0	0	0	0	0

Table A.1: Shunters' accident data

Aitchison's Trivariate Bacterial Count Data

The triplet of bacterial colony counts in three air samplers located in each of 50 different sterile rooms is obtained. This dataset can be found in Aitchison, *et al.* (1989, Table 4 in [2]).

1	2	11	8	6	0	2	13	5	2	8	1
5	6	5	14	1	7	3	9	2	7	6	8
3	4	12	1	9	7	3	6	6	3	9	14
4	2	25	9	7	3	5	4	8	4	4	7
7	3	2	1	14	6	2	13	0	14	9	5
3	8	2	1	1	30	4	5	15	7	6	3
8	10	4	3	2	10	6	8	5	2	3	10
1	7	3	2	9	12	7	10	5	2	2	8
3	15	3	1	8	2	4	6	0	8	7	3
6	6	6	4	14	7	3	3	14	6	8	3
22	9	6	5	2	4	2	0	6	2	1	1
4	6	4	4	9	2	8	4	6	3	10	6
4	7	10	2	4	6						

Table A.2: Aitchison's trivariate bacterial count data

Fisher's Data

This is an univariate count dataset (Table A.3) representing a sample of sheeps classified according to the number of ticks found on each, as given by Fisher in [11]. The total number of observations is $n = 60$.

Number of ticks	0	1	2	3	4	5	6	7	8	9	10
Number of sheeps	7	9	8	13	8	5	4	3	0	1	2

Table A.3: Classification of sheeps based on number of ticks