Marshall University
Marshall Digital Scholar

Theses, Dissertations and Capstones

2015

A Generalized Inflated Geometric Distribution

Ram Datt Joshi
joshi3@marshall.edu

Follow this and additional works at: http://mds.marshall.edu/etd

 Part of the Statistical Models Commons, and the Statistical Theory Commons

Recommended Citation
Joshi, Ram Datt, "A Generalized Inflated Geometric Distribution" (2015). *Theses, Dissertations and Capstones.* Paper 916.

This Thesis is brought to you for free and open access by Marshall Digital Scholar. It has been accepted for inclusion in Theses, Dissertations and Capstones by an authorized administrator of Marshall Digital Scholar. For more information, please contact zhangj@marshall.edu.

A GENERALIZED INFLATED GEOMETRIC DISTRIBUTION

A thesis submitted to
the Graduate College of
Marshall University
In partial fulfillment of
the requirements for the degree of
Master of Arts
in
Mathematics
by
Ram Datt Joshi
Approved by
Dr. Avishek Mallick, Committee Chairperson
Dr. Laura Adkins
Dr. Alfred Akinsete

Marshall University
May 2015

ACKNOWLEDGEMENTS

First, I would like to express my sincere thanks to my thesis adviser Dr. Avishek Mallick for his continuous support and guidance throughout the entire project. His infectious enthusiasm and unlimited zeal has always pushed me forward to the completion of the thesis. Without his support, this work would never have been in this form. Besides, he is the one who challenged me the most and pushed me the hardest from his courses in class from the beginning of my master's program at Marshall. I take the opportunity to say that he prepared me thoroughly for a doctoral program.

I would like to thank my committee members Dr. Alfred Akinsete and Dr. Laura Adkins. Dr. Akinsete is the one who always encouraged and guided me whenever I was in need of his guidance. He provided the easy and the best solution for whatever I used to reach him at. I am always grateful to his generosity and support. On the other hand, Dr. Adkins always has been eager to know about the progress and updates throughout my work. She always encouraged me for the completion of my thesis as well as my studies throughout my stay at Marshall. I am fortunate to be a student in one of her classes at Marshall where I learned a lot from her incredible Biostatistics class.

In addition, a special thank you to Dr. Rubin Gerald who taught me throughout my last year in his statistics classes. I learned many things from his classes which have been useful for the completion of this thesis. Moreover, he has been a constant source of inspiration towards the completion my masters degree at Marshall.

I would like to thank Dr. Basant Karna who as my mentor, helped and guided me throughout in every possible way. I am grateful for his care and guidance in every step I seek help from him throughout my time at Marshall. Without his support, life would not have been so easier. The other faculty that I never forget to mention is Dr. Ari Aluthge. He has not only prepared me in his Advanced Calculus course, but also helped me to go to the advanced doctoral program by writing letters of recommendation and giving me the courage to go ahead at the time of necessity.

A sincere thank goes to Dr. Carl Mummert for he has been the constant source of motivation towards the completion of my masters degree at Marshall. He is the professor who always listen every student carefully and he is fond of helping others.

I would like to thank the whole Department of Mathematics at Marshall University, for every member associated with the department has helped me directly or indirectly for the completion of my thesis in different possible ways.

At last but not the least, I would like to thank my parents and my wife-Saru, who always have been the source of motivation and inspiration throughout. They have been apprehensive about my study goals and my health condition. Without their blessings and prayers, this work would not have come into existence. I am thankful for my own health condition as well.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A count data that have excess number of zeros, ones, twos or threes are commonplace in experimental studies. But these inflated frequencies at particular counts may lead to over dispersion and thus may cause difficulty in data analysis. So, to get appropriate results from them and to overcome the possible anomalies in parameter estimation, we may need to consider suitable inflated distribution.

In this thesis, we have considered a Swedish fertility dataset with inflated values at some particular counts. Generally, Inflated Poisson or Inflated Negative Binomial distribution are the most common distributions for analyzing such data. Geometric distribution can be thought of as a special case of Negative Binomial distribution. Hence we have used a Geometric distribution inflated at certain counts, which we called Generalized Inflated Geometric distribution to analyze such data. The data set is analyzed, tested and compared using various tests and techniques to ensure the better performance of multi-point inflated Geometric distribution over the standard Geometric distribution.

The various tests and techniques used include comparing the parameters obtained through method of moment estimators and maximum likelihood estimators. The two types of estimators obtained from method of moment estimations and maximum likelihood estimation method, were compared using simulation study, and it is found after the analysis that the maximum likelihood estimators perform better.

CHAPTER 1

INTRODUCTION

A random variable X that counts the number of trials to obtain the r^{th} success in a series of independent and identical Bernoulli trials, is said to have a Negative Binomial distribution whose probability mass function (pmf) is given by

$$P(X = k) = P(k|p) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \tag{1.1}$$

where $r = 1, 2, 3, \ldots$; $k = r, r+1, \ldots$ and $p > 0$.

The above distribution is also the "Generalized Power Series distribution" as mentioned in Johnson et al. (2005)[7]. Some writers, for instance Patil et al. (1984)[9], called this the "P*ó*lya-Eggenberger distribution", as it arises as a limiting form of Eggenberger and P*ó*lya's (1923)[3] urn model distribution. A special case of Negative Binomial Distribution is the Geometric distribution which can be defined in two different ways

Firstly, the probability distribution for a Geometric random variable X (where X being the number of independent and identical trials to get the first success) is given by

$$P(X = k|p) = \begin{cases} p(1-p)^{k-1} & \text{if } k = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

However, instead of counting the number of trials, if the random variable X counts the number of failures before the first success, then it will result in the second type of Geometric distribution which again is a special case of Negative Binomial distribution when $r = 1$ (first success) and its pmf is given by

$$P(X = k) = P(k \mid p) = \begin{cases} p(1-p)^{k} & \text{if } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

The support set of this random variable is {0,1,2,...} which makes it different from the

distribution in (1.2). The above model in (1.3), henceforth referred to as "Geometric(p)" has mean $\frac{1-p}{p}$ and variance $\frac{1-p}{p^2}$ and is the only distribution with non-negative integer support that can be characterized by the "Memory-less property" or the "Markovian property". Many other characterizations of this distribution can be found in Feller (1968, 1969) [4] [5]. The distribution occurs in many applications and some of them are indicated in the references below:

- The famous problem of Banach's match boxes (Feller (1968)[4]);
- The runs of one plant species with respect to another in transects through plant populations (Pielou (1962, 1963)[10][11]);
- A ticket control problem (Jagers (1973)[6]);
- A survillance system for congenital malformations (Chen (1978)[2]);
- The number of tosses of a fair coin before the first head (success) appears;
- The number of drills in an area before observing the first productive well by an oil prospector.(Wackerly et al. (2008)[13]).

The Geometric model in (1.3) which is widely used for modeling count data may be inadequate for dealing with overdispersed as well as underdispersed count data. One such instance is the abundance of zero counts in the data, and (1.3) may be an inefficient model for such cases due to the presence of heterogeneity, which usually results in undesired over dispersion. Therefore, to overcome this situation, i.e., to explain or capture such heterogeneity, we consider a 'two-mass distribution' by giving mass π to 0 counts, and mass $(1-\pi)$ to the other class which follows Geometric(p). The result of such a 'mixture distribution' is called the 'Zero-Inflated Geometric' (ZIG) distribution with the probability mass function

$$P(k \mid p, \pi) = \begin{cases} \pi + (1-\pi)p & \text{if } k = 0 \\ (1-\pi)P(k \mid p) & \text{if } k = 1, 2, \ldots \end{cases} \tag{1.4}$$

where, $p > 0$, and $P(k \mid p)$ is given in (1.3). However the mixing parameter π is chosen such that $P(k=0) \in (0,1)$ in (1.4), i.e., it ranges over the interval $-\frac{p}{1-p} < \pi < 1$. This allows the

distribution to be well defined for certain negative values of π, depending on p. Although the mixing interpretation is lost when $\pi < 0$, these values have a natural interpretation in terms of zero-deflation, relative to a Geometric(p) model. Correspondingly, $\pi > 0$ can be regarded as zero inflation as discussed in Johnson et al. (2005)[7].

A further generalization of (1.4) can be obtained by inflating/deflating the Geometric distribution at several specific values. To be precise, if the discrete random variable X is thought to have inflated probabilities at the values $k_1, ..., k_m \in \{0, 1, 2, ...\}$, then the following general probability mass function can be considered:

$$P(k \mid p, \pi_i, 1 \leq i \leq m) = \begin{cases} \pi_i + \left(1 - \sum_{i=1}^{m} \pi_i\right) P(k \mid p) & \text{if } k = k_1, k_2, \ldots, k_m \\ \left(1 - \sum_{i=1}^{m} \pi_i\right) P(k \mid p) & \text{if } k \neq k_i; 1 \leq i \leq m \end{cases} \tag{1.5}$$

where $k = 0, 1, 2, \ldots$; $p > 0$ and π_i's are chosen in such a way that $P(k_i) \in (0, 1)$ for all $i = 1, 2, ..., m$ in (1.5). For the remainder of this work, we will refer to (1.5) as the Generalized Inflated Geometric (GIG) distribution which is the main focus of this work.

We will consider some special cases of the (GIG) distribution such as Zero-One-Inflated Geometric (ZOIG) distribution in the case $k = 2$ with $k_1 = 0$ and $k_2 = 1$ or Zero-One-Two Inflated Geometric (ZOTIG) models. Similar type of Generalized Inflated Poisson (GIP) models have been considered by Melkersson and Rooth (2000)[8] to study a women's fertility data of 1170 Swedish women of the age group 46-76 years (Table 1.1). This data set consists of the number of child(ren) per woman, who have crossed the childbearing age in the year 1991. They justified the Zero-Two Inflated Poisson distribution was the best to model it. However recently in his Master's Thesis, Stewart (2014)[12] studied the same data set and found that a Zero-Two-Three Inflated Poisson (ZTTIP) distribution was a better fit.

Instead of using an Inflated Poisson model, we will consider fitting appropriate Inflated Geometric models to the data in Table 1.1. Now, which model is the best fit whether a GIG model with focus on counts (0, 1), i.e., ZOIG or a GIG model focusing on some other set $\{k_1, k_2, ..., k_m\}$ is appropriate for the above data will be eventually decided by different model selection criteria in Chapter 4. In the next chapter, we discuss different techniques of parameter

Table 1.1: Observed number of children (=count) per woman

Count	Frequency	Proportion
0	114	.097
1	205	.175
2	466	.398
3	242	.207
4	85	.073
5	35	.030
6	16	.014
7	4	.003
8	1	.001
10	1	.001
12	1	.001
Total	1,170	1.000

estimation namely, the method of moments (MME), and the maximum likelihood estimation (MLE). In Chapter 3, we compare the performances of MMEs and MLEs for different GIG model parameters using simulation studies.

CHAPTER 2

ESTIMATION OF MODEL PARAMETERS

In this chapter, we estimate the parameters by two well known methods of parameter estimation namely the method of moment estimations and the method of maximum likelihood estimations.

2.1 Method of Moments Estimation (MME)

The easiest way to obtain estimators of the parameters is through the method of moments estimation (MME). The r^{th} raw moment of a random variable X following a GIG in (1.5) with parameters $\pi_1, \ldots, \pi_m$ and p can be obtained from the following expression:

$$\begin{aligned} E[X^r] &= \sum_{k=0}^{\infty} k^r P(k|p, \pi_i, 1 \leq i \leq m) \\ &= \sum_{i=1}^{m} k_i^r \pi_i + (1 - \sum_{i=1}^{m} \pi_i) \sum_{k=0}^{\infty} k^r P(k|p) \\ &= \sum_{i=1}^{m} k_i^r \pi_i + (1 - \sum_{i=1}^{m} \pi_i) \mu_r'(p). \end{aligned} \tag{2.1}$$

where $\mu_r'(p)$ is the the r^{th} raw moment of Geometric(p) and can be calculated easily by differentiating its moment generating function (MGF) given by $\dfrac{p}{1-(1-p)e^t}$, i.e., $\mu_r'(p) = \frac{d^r}{dt^r}\left(\dfrac{p}{1-(1-p)e^t}\right)\Big|_{t=0}$.

Given a random sample $X_1, ..., X_n$, i.e. independent and identically distributed (iid) observations from the GIG distribution, we equate the sample moments with the corresponding population moments to get a system of (m+1) equations involving the (m+1) model parameters $p, \pi_1, \ldots, \pi_m$ of the form

$$m_r' = \sum_{i=1}^{m} k_i^r \pi_i + \left(1 - \sum_{i=1}^{m} \pi_i\right) \mu_r'(p) \tag{2.2}$$

where $r = 1, 2, \ldots, (m+1)$. Note that $m_r' = \sum_{j=1}^{n} X_j^r / n$ is the r^{th} raw sample moment. The values of $\pi_i, i = 1, 2, ..., m$, and p obtained by solving the system of equations (2.2) are denoted by $\hat{\pi}_{i(MM)}$ and $\hat{p}_{MM}$ respectively. The subscript "(MM)" indicates the MME approach. Note that the parameter p is non-negative and hence the estimate must obey this restriction. But there is

no such guarantee, as such we propose the corrected MME's to ensure non-negativity of this moment estimator as

$$\hat{p}^c_{MM} = \hat{p}_{MM} \text{ truncated at 0 and 1 and } \hat{\pi}^c_{i(MM)} = \hat{\pi}_{i(MM)} \tag{2.3}$$

where $\hat{\pi}^c_{i(MM)}$ is the solution of π_i in (2.2) obtained after substituting $\hat{p}_{MM}$.

Consider the case of ZIG distribution when $m = 1$ i.e., $k_1 = 0$, resulting into only two parameters to estimate, i.e. π_1 and p. The population mean and population variance in this special case are obtained as:

$$E(X) = (1-\pi_1)\frac{(1-p)}{p} \text{ and } Var(X) = (1-\pi_1)\{1+\pi_1(1-p)\}\frac{(1-p)}{p^2} \tag{2.4}$$

Now, in order to obtain the method of moments estimators, we can equate the preceding mean and variance with sample mean ($\bar{X}$) and sample variance (s^2) respectively. This is an alternative approach to estimate the parameters instead of dealing with the sample raw moments. Hence we obtain,

$$\begin{aligned} (1-\pi_1)\frac{(1-p)}{p} &= \bar{X} \\ (1-\pi_1)\{1+\pi_1(1-p)\}\frac{(1-p)}{p^2} &= s^2 \end{aligned} \tag{2.5}$$

Solving the above equations simultaneously for p and π_1, we get, $\hat{p}^c_{MM} = \dfrac{2\bar{X}}{(s^2+\bar{X}+\bar{X}^2)}$ and $\hat{\pi}^c_{1(MM)} = \dfrac{s^2-\bar{X}-\bar{X}^2}{s^2-\bar{X}+\bar{X}^2}$.

Now let us consider another special case of GIG, the Zero-One Inflated Geometric (ZOIG) distribution, i.e., $m = 2$ and $k_1 = 0$ and $k_2 = 1$. It has three parameters π_1, π_2 and p and to estimate them we need to equate the first three raw sample moments with the corresponding

population moments and we thus obtain the following system of equations:

$$\begin{aligned} m_1' &= \pi_2 + (1-\pi_1-\pi_2)\frac{(1-p)}{p} \\ m_2' &= \pi_2 + (1-\pi_1-\pi_2)\frac{(p^2-3p+2)}{p^2} \\ m_3' &= \pi_2 + (1-\pi_1-\pi_2)\frac{(1-p)(p^2-6p+6)}{p^3} \end{aligned} \tag{2.6}$$

Similarly, another special case of GIG is the Zero-One-Two Inflated Geometric (ZOTIG) distribution. Here we have four parameters π_1, π_2, π_3 and p to estimate. This is done by solving the following system of four equations obtained by equating the first four raw sample moments with their corresponding population moments to have,

$$\begin{aligned} m_1' &= \pi_2 + 2\pi_3 + (1-\pi_1-\pi_2-\pi_3)\frac{(1-p)}{p} \\ m_2' &= \pi_2 + 4\pi_3 + (1-\pi_1-\pi_2-\pi_3)\frac{(1-p)(2-p)}{p^2} \\ m_3' &= \pi_2 + 8\pi_3 + (1-\pi_1-\pi_2-\pi_3)\frac{(1-p)(p^2-6p+6)}{p^3} \\ m_4' &= \pi_2 + 16\pi_3 + (1-\pi_1-\pi_2-\pi_3)\frac{(1-p)(2-p)(p^2-12p+12)}{p^4} \end{aligned} \tag{2.7}$$

Algebraic solutions to these systems of equations in (2.6) and (2.7) are obtained by using Mathematica and are given in Appendix (A). We note that these solutions may not fall in the feasible regions of the parameter space, so we put appropriate restrictions to these solutions as discussed for the ZIG distribution to obtain the corrected MMEs.

2.2 Maximum Likelihood Estimation (MLE)

In this section, we discuss the approach of estimating our parameters by the method of Maximum Likelihood Estimation(MLE). Based on the random sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we define the likelihood function $L = L(p, \pi_i, 1 \leq i \leq m| \ \mathbf{X})$ as follows. Let Y_i = the number of observations at k_i with inflated probability, i.e., if I is an indicator function, then $Y_i = \sum_{j=1}^{n} I(X_j = k_i)$, $1 \leq i \leq m$, which means Y_i is the total number of observed counts at k_i. Also, let $Y_\bullet = \sum_{i=1}^{m} Y_i =$ total number of observations with inflated observations, $n =$ total number of observations and,

$(n - Y.)$ is the total number of non-inflated observations. Then,

$$\begin{aligned} L &= \prod_{i=1}^{m}\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i \mid p)\}^{Y_i} \prod_{X_j \neq k_i}\{(1-\sum_{l=1}^{m}\pi_l)P(X_j|p)\} \\ &= \prod_{i=1}^{m}\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i \mid p)\}^{Y_i}(1-\sum_{l=1}^{m}\pi_l)^{(n-Y.)} \prod_{X_j \neq k_i} P(X_j|p) \end{aligned} \tag{2.8}$$

The log likelihood function $l = \ln L$ is given by

$$l = \sum_{i=1}^{m} Y_i \ln\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i \mid p)\} + (n - Y.)\ln(1-\sum_{l=1}^{m}\pi_l) + \sum_{X_j \neq k_i} \ln P(X_j \mid p)$$

But we have,

$$\sum_{X_j \neq k_i} \ln P(X_j \mid p) = (n - Y.)\ln p + \ln(1-p)(\sum_{j=1}^{n} X_j - \sum_{l=1}^{m} k_l Y_l)$$

hence the log likelihood function becomes

$$\begin{aligned} l = &\sum_{i=1}^{m} Y_i \; ln\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i \mid p)\} + (n - Y.)\ln(1-\sum_{l=1}^{m}\pi_l) \\ &+ (n - Y.)\ln p + \ln(1-p)(\sum_{j=1}^{n} X_j - \sum_{l=1}^{m} k_l Y_l) \end{aligned} \tag{2.9}$$

Now to obtain the MLEs, we maximize l in (2.9) with respect to the parameters π_i, $1 \leq i \leq m$, and p over the appropriate parameter space. Differentiating l partially w.r.t the parameters and setting them equal to zero yields the following system of likelihood equations or score equations.

$$\begin{aligned} \frac{\partial l}{\partial \pi_i} &= \frac{Y_i}{\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i|p)\}} - \sum_{i=1}^{m}\frac{Y_i P(k_i|p)}{\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i|p)\}} \\ &\qquad - \frac{(n - Y.)}{(1-\sum_{l=1}^{m}\pi_l)} = 0, \quad \forall i = 1, \ldots, m; \\ \frac{\partial l}{\partial p} &= \sum_{i=1}^{m}\frac{Y_i(1-\sum_{l=1}^{m}\pi_l)P^{(p)}(k_i|p)}{\{\pi_i + (1-\sum_{l=1}^{m}\pi_l)P(k_i|p)\}} + \frac{(n - Y.)}{p} - \frac{(n\bar{X} - \sum_{l=1}^{m} k_l Y_l)}{(1-p)} = 0 \end{aligned} \tag{2.10}$$

where, $P^{(p)}(k_i|p) = (\partial/\partial p)P(k_i|p)$.

At this point, we can not say which of these two estimation techniques (MME or MLE) provide overall better estimators.To the best of our knowledge, no comparative study has been reported in

literature. Further we do not have any closed form expressions for these estimators which makes it even more difficult to compare their performance. Therefore, we conduct simulation studies in the next chapter which can provide some guidance about their performances.

CHAPTER 3

Simulation Study

We have considered the following three cases for our simulation study:

(i) $m = 1, k_1 = 0$ (Zero Inflated Geometric (ZIG) distribution)

(ii) $m = 2, k_1 = 0, k_2 = 1$ (Zero-One Inflated Geometric (ZOIG) distribution)

(ii) $m = 3, k_1 = 0, k_2 = 1, k_3 = 2$ (Zero-One-Two Inflated Geometric (ZOTIG) distribution)

For each model mentioned above, we generate random data $X_1, ..., X_n$ from the distribution (with given parameter values) $N = 10,000$ times. Let us denote a parameter (either π_i or p) by the generic notation θ. The parameter θ is estimated by two possible estimators $\hat{\theta}_{MM}^{(c)}$ (the corrected MME) and $\hat{\theta}_{ML}$ (the MLE). At the lth replication, $1 \leq l \leq N$, the estimates of θ are $\hat{\theta}_{MM}^{(c)(l)}$ and $\hat{\theta}_{ML}^{(l)}$ respectively. Then the standardized bias (called 'SBias') and standardized mean squared error (called 'SMSE') are defined and approximated as below

$$\begin{aligned} \text{SBias}(\hat{\theta}) &= E(\hat{\theta} - \theta)/\theta \approx \{\sum_{l=1}^{N} (\hat{\theta}^{(l)} - \theta)/\theta\}/N \\ \text{SMSE}(\hat{\theta}) &= E(\hat{\theta} - \theta)^2/\theta^2 \approx \{\sum_{l=1}^{N} (\hat{\theta}^{(l)} - \theta)^2/\theta^2\}/N \end{aligned} \tag{3.1}$$

Note that $\hat{\theta}$ will be replaced by $\hat{\theta}_{MM}^{(c)}$ and $\hat{\theta}_{ML}$ in our simulation study. Further observe that we are using SBias and SMSE instead of the actual Bias and MSE, because the standardized versions provide more information. An error of magnitude 0.01 in estimating a parameter with true value 1.00 is more severe than a situation where the parameter's true value is 10.0. This fact is revealed through SBias and/or SMSE more than the actual bias and/or MSE.

3.1 The ZIG Distribution

In our simulation study for the Zero Inflated Geometric (ZIG) distribution, we fix $p = 0.2$ and vary π_1 from 0.1 to 0.8 with an increment of 0.1 for n = 25 and n = 50. The constrained optimization algorithm "L-BFGS-B" (Byrd et al. (1995))[1] is implemented in R programming language to obtain the maximum likelihood estimators (MLEs) of the parameters p and π_1, and

the corrected MMEs are obtained by solving a system of equations and imposing appropriate restrictions on the parameters. In order to compare the performances of the MLEs with that of the CMMEs, we plot the standardized biases (SBias) and standardized MSE (SMSE) of these estimators obtained over the allowable range of π_1. The SBias and SMSE plots are presented in Figure(3.1) and Figure(3.2) for sample sizes 25 and 50 respectively.

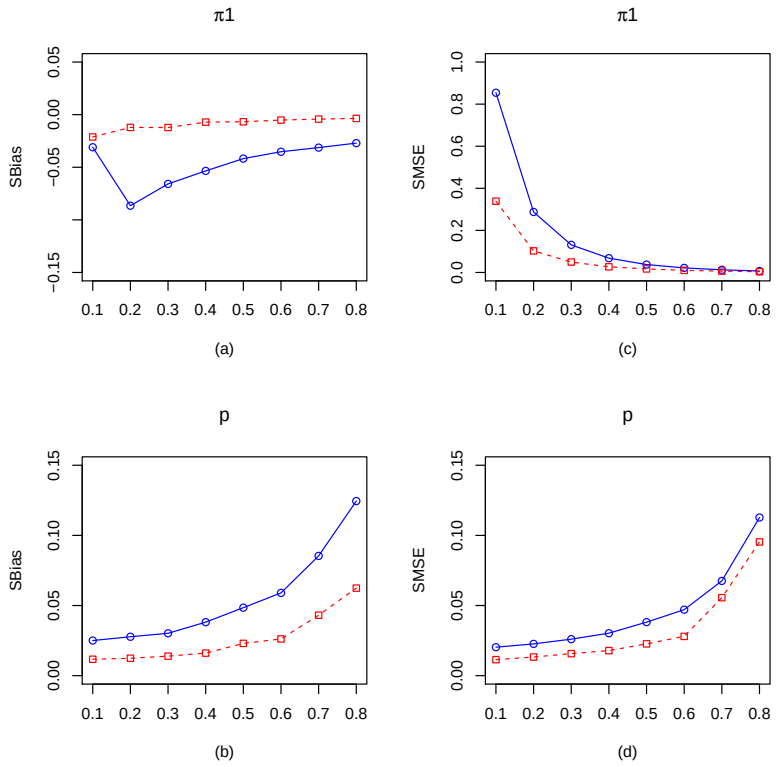

Figure 3.1: Plots of the absolute SBias and SMSE of the CMMEs and MLEs of π_1 and p (from ZIG distribution) plotted against π_1 for $p = 0.2$ and $n = 25$. The solid line represents the SBias or SMSE of the CMME. The dashed line represents the SBias or SMSE of the MLE. (a) Comparison of SBias of π_1 estimators. (b) Comparison of SBias of p estimators. (c) Comparison of SMSE of π_1 estimators. (d) Comparison of SMSE of p estimators.

For the ZIG distribution with $n = 25$, from Figure 3.1(a) we see that MLE outperforms CMME for all values of π_1 with respect to SBias. Here MLE is almost unbiased for all values of π_1 beyond 0.4. The SBias seems to be maximum for CMME at 0.2 and for MLE at around 0.1. In Figure 3.1(b), we again see that MLE uniformly outperforms the CMME and both are increasing with values of π_1. In Figure 3.1(c), MLE consistently outperforms CMME at all points until 0.65

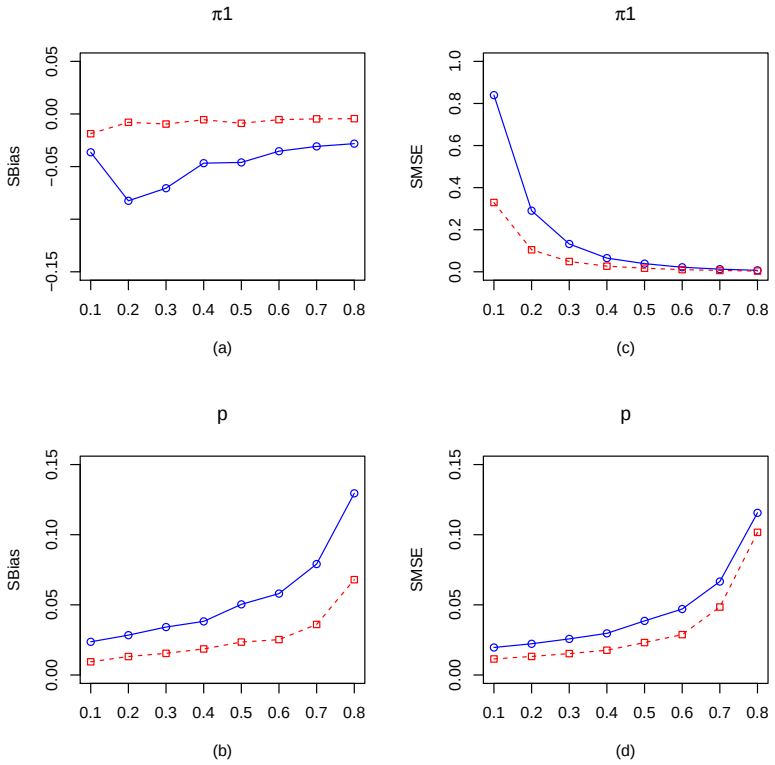

Figure 3.2: Plots of the absolute SBias and SMSE of the CMMEs and MLEs of π_1 and p (from ZIG distribution) plotted against π_1 for $p = 0.2$ and $n = 50$. The solid line represents the SBias or SMSE of the CMME. The dashed line represents the SBias or SMSE of the MLE. (a) Comparison of SBias of π_1 estimators. (b) Comparison of SBias of p estimators. (c) Comparison of SMSE of π_1 estimators. (d) Comparison of SMSE of p estimators.

after which they both seem to have nearly the same SMSE. For both MLE and CMME, the SMSE starts off at their highest values and then decreases rapidly until it reaches nearly zero. The SMSE of MLE consistently outperforms that of CMME in Figure 3.1(d). They both start off at their lowest values, and increase as values of π_1 get higher.

Again MLEs outperform CMMEs with respect to both SBias and SMSE for sample size 50, as exhibited by Figure 3.2. So we see for the ZIG distribution, the MLEs of the parameters π_1 and p perform better than the CMMEs for all values of π_1 we have considered.

3.2 The ZOIG Distribution

In the case of the Zero-One Inflated Geometric (ZOIG) distribution we have three parameters to consider, namely π_1, π_2 and p. For fixed $p = 0.3$ we vary π_1 and π_2 one at a time for sample sizes 25 and 50. Figure 3.3 presents the six comparisons for $\hat{\pi}^{(c)}_{1(MM)}$, $\hat{\pi}^{(c)}_{2(MM)}$ and $\hat{p}^{(c)}_{(MM)}$ with $\hat{\pi}_{1(ML)}$, $\hat{\pi}_{2(ML)}$ and $\hat{p}_{(ML)}$ in terms of standardized bias and standardize MSE for $n = 25$, varying π_1 from 0.1 to 0.5 and keeping π_2 and λ fixed at 0.15 and 0.3 respectively. Figure 3.4 shows the same for $n = 50$.

In Figure 3.3(a), MLE outperforms CMME at all points with respect to SBias. SBias of MLE starts above zero and quickly becomes negative, whereas that of CMME is throughout negative. In Figure 3.3(b), we see that the MLE is almost unbiased for all values of π_1, and SBias of CMME is always negative. Again in Figure 3.3(c), we see that MLE is essentially unbiased but SBias of CMME is increasing with values of pi_1. In Figure 3.3(d), SMSE of CMME and MLE are both decreasing, but SMSE of MLE stays below that of CMME. In Figures 3.3(e) and 3.3(f), SMSE of MLE stays constant at 0.55 and 0.05 respectively for all permissible values of π_1. Also CMME performs way worse for both the cases.

In Figure 3.4, we observe similar performance as in Figure 3.3, i.e, MLEs of all the parameters is outperforming CMMEs with respect to both SBias and SMSE for all considered values of pi_1.

In our second scenario which is presented in Figures 3.5 and 3.6 for sample sizes 25 and 50 respectively, we vary π_2 keeping π_1 and λ fixed at 0.15 and 3 respectively. In Figures 3.5(a), we see that both MLE and CMME are negatively biased, but MLE is always performing better that the CMME. However in Figure 3.5(b), MLE starts of with a positive bias and then becomes

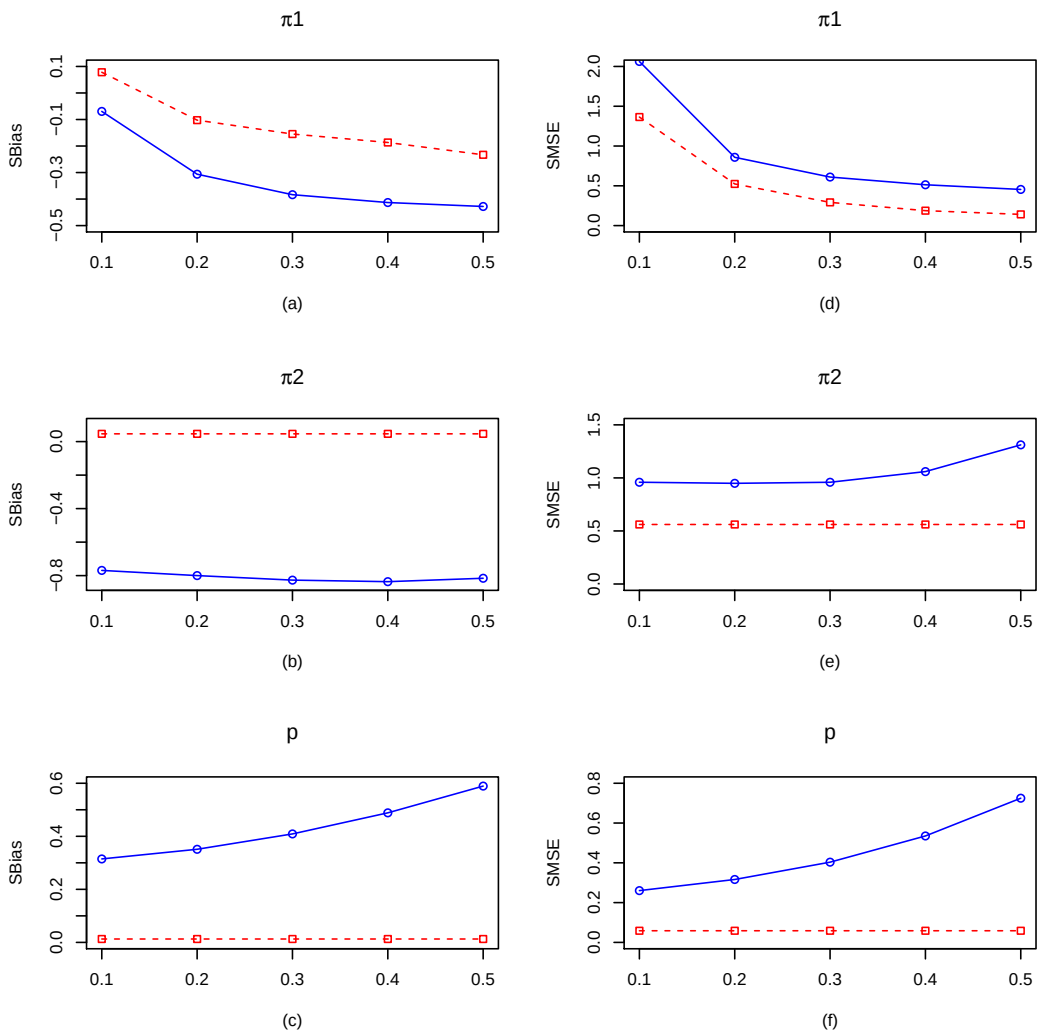

Figure 3.3: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2 and p (from ZOIG distribution) by varying π_1 for fixed $\pi_2 = 0.15$, $p = 0.3$ and $n = 25$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(c) Comparisons of SBiases of π_1, π_2 and p estimators respectively. (d)-(f) Comparisons of SMSEs of π_1, π_2 and p estimators respectively.

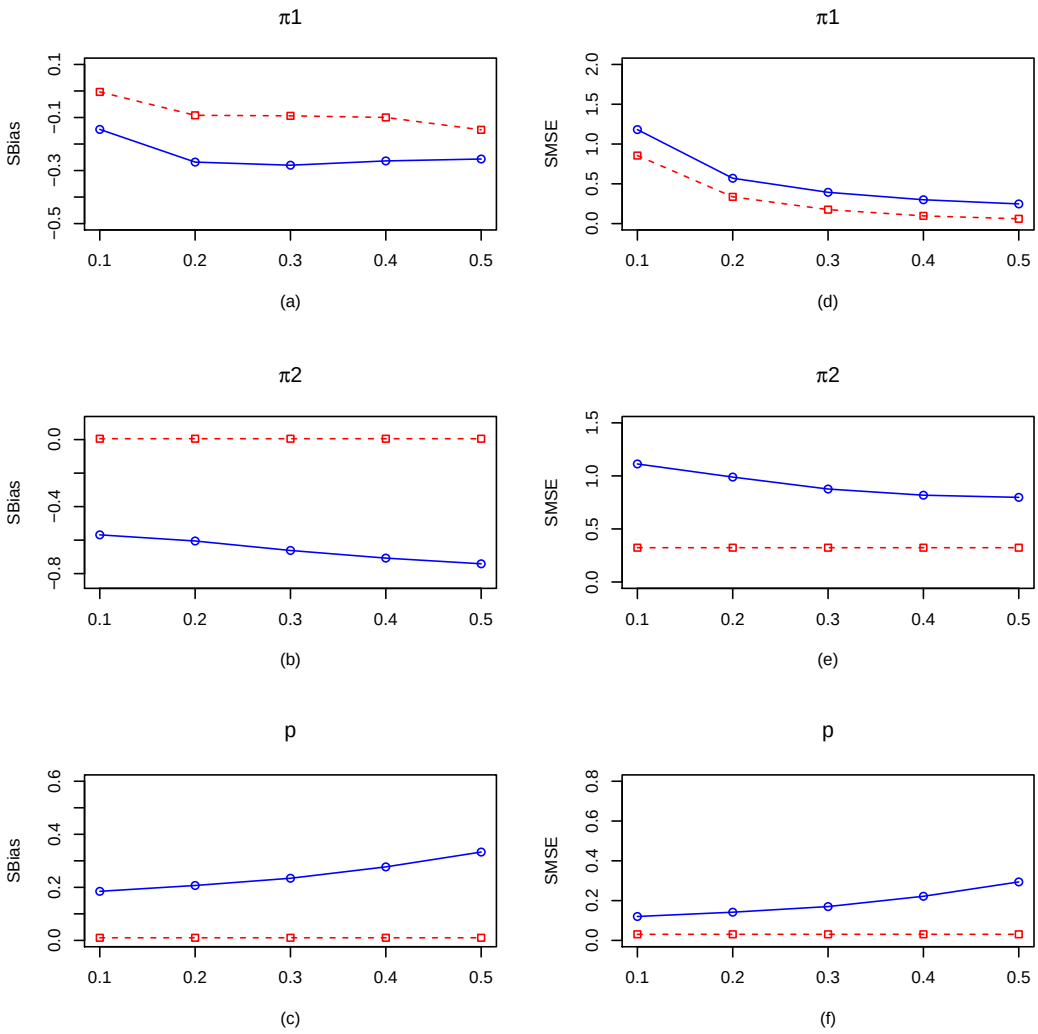

Figure 3.4: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2 and p (from ZOIG distribution) by varying π_1 for fixed $\pi_2 = 0.15$, $p = 0.3$ and $n = 50$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(c) Comparisons of SBiases of π_1, π_2 and p estimators respectively. (d)-(f) Comparisons of SMSEs of π_1, π_2 and p estimators respectively.

unbiased for π_2 between 0.2 and 0.3 and eventually ends with a negative bias. SBias of CMME stays throughout between -0.6 and -0.8. In Figure 3.5(c), we see that MLE outperforms MME throughout with respect to SBias. From Figures 3.5(d), 3.5(e) and 3.5(f), it is clear that MLE outperforms MME with respect to SMSE for all permissible values of π_2. Thus we observe that the MLEs of the all three parameters perform better than the MMEs in terms of the both absolute SBias and SMSE.

In Figure 3.6, we observe similar performance as in Figure 3.5, i.e, MLEs of all the parameters is outperforming CMMEs with respect to both SBias and SMSE for all values of π_2 from 0.1 to 0.5.

3.3 The ZOTIG Distribution

For the Zero-One-Two Inflated Geometric (ZOTIG) distribution we have four parameters to consider, namely π_1, π_2, π_3 and p. For fixed $p = 3$ we vary π_1, π_2, and π_3 one at a time from 0.1 to 0.5 for sample sizes $n = 25$ and $n = 50$. Thus we have eight comparisons for $\hat{\pi}^{(c)}_{1(MM)}$, $\hat{\pi}^{(c)}_{2(MM)}$, $\hat{\pi}^{(c)}_{3(MM)}$ and $\hat{p}^{(c)}_{(MM)}$ with $\hat{\pi}_{1(ML)}$, $\hat{\pi}_{2(ML)}$, $\hat{\pi}_{3(ML)}$ and $\hat{p}_{(ML)}$. These comparisons in terms of absolute standardized bias and standardize MSE are presented in Figures 3.7-3.12.

In the first scenario of ZOTIG distribution, which is presented in Figures 3.7 and 3.8, we vary π_1 keeping π_2, π_3 and p fixed at 0.2, 0.2 and 0.3 respectively. From Figure 3.7(a, b, c, d), we see that the CMMEs of all the four parameters perform consistently worse than the MLEs with respect to SBias. However SBias of CMMEs and MLEs are same at $\pi_1 = 0.5$. Also from parts (e, f, g, h) in Figure 3.7 concerning the SMSE, we notice that the CMMEs of all the parameters perform consistently worse than the MLEs. But as in the case of SBias, we see that SMSE of CMMEs and MLEs are same at $\pi_1 = 0.5$. In Figure 3.8, for sample size 50 we again observe that MLE of all the four parameters are outperforming CMME with respect to SBias and SMSE.

In our second scenario which is presented in Figures 3.9 and 3.10, we vary π_2 keeping π_1, π_3 and p fixed at 0.2, 0.2 and 0.3 respectively. We observe some interesting things in these plots. From Figure 3.9 we see that MLE of both π_1 and p is performing better upto $\pi_2 = 0.3$ with respect to SBias but after that CMME is performing slightly better than MLE. Also MLE of π_2 is better till 0.3 but after that both MLE and CMME have the same SBias. However we see that the MLEs of all four parameters perform better than their CMME counterparts with respect to

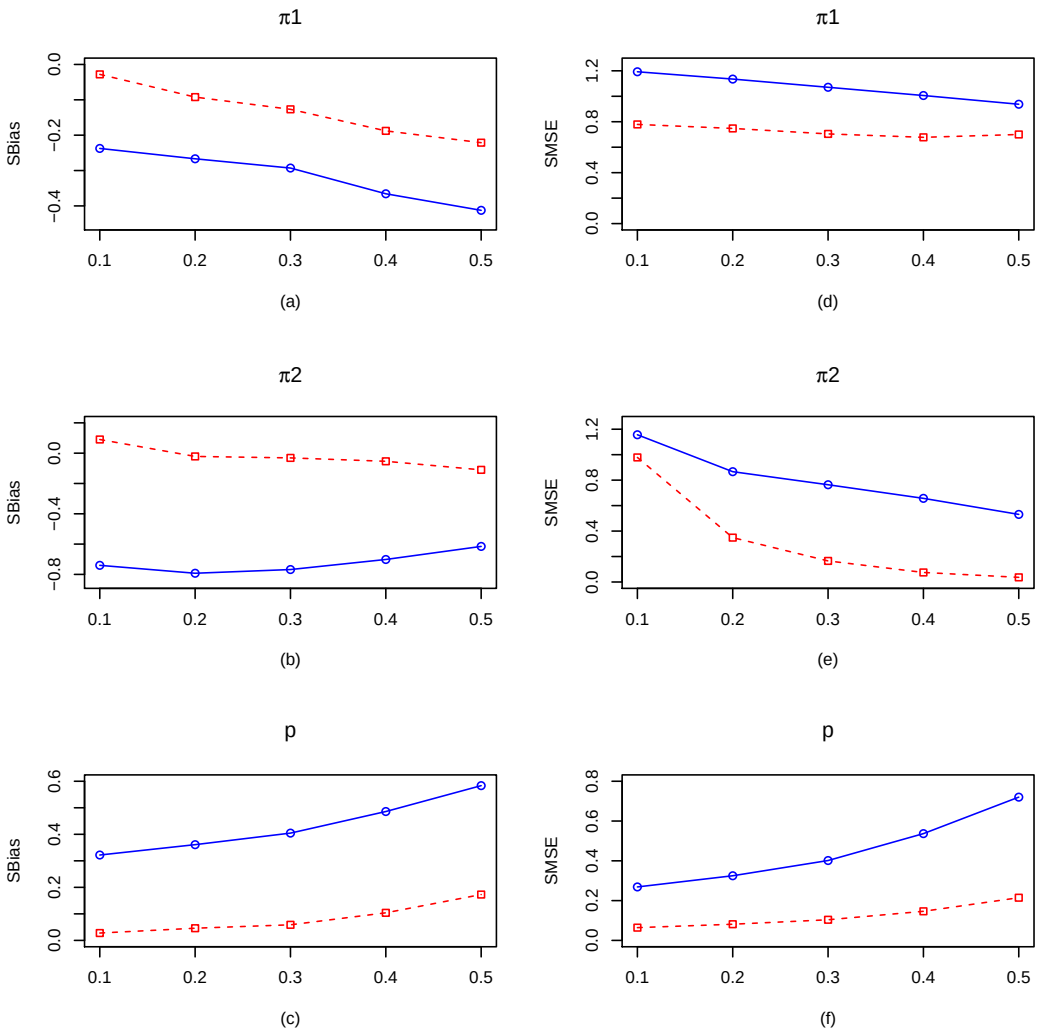

Figure 3.5: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2 and p (from ZOIG distribution) by varying π_2 for fixed $\pi_1 = 0.15$, $p = 0.3$ and $n = 25$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(c) Comparisons of SBiases of π_1, π_2 and p estimators respectively. (d)-(f) Comparisons of SMSEs of π_1, π_2 and p estimators respectively.

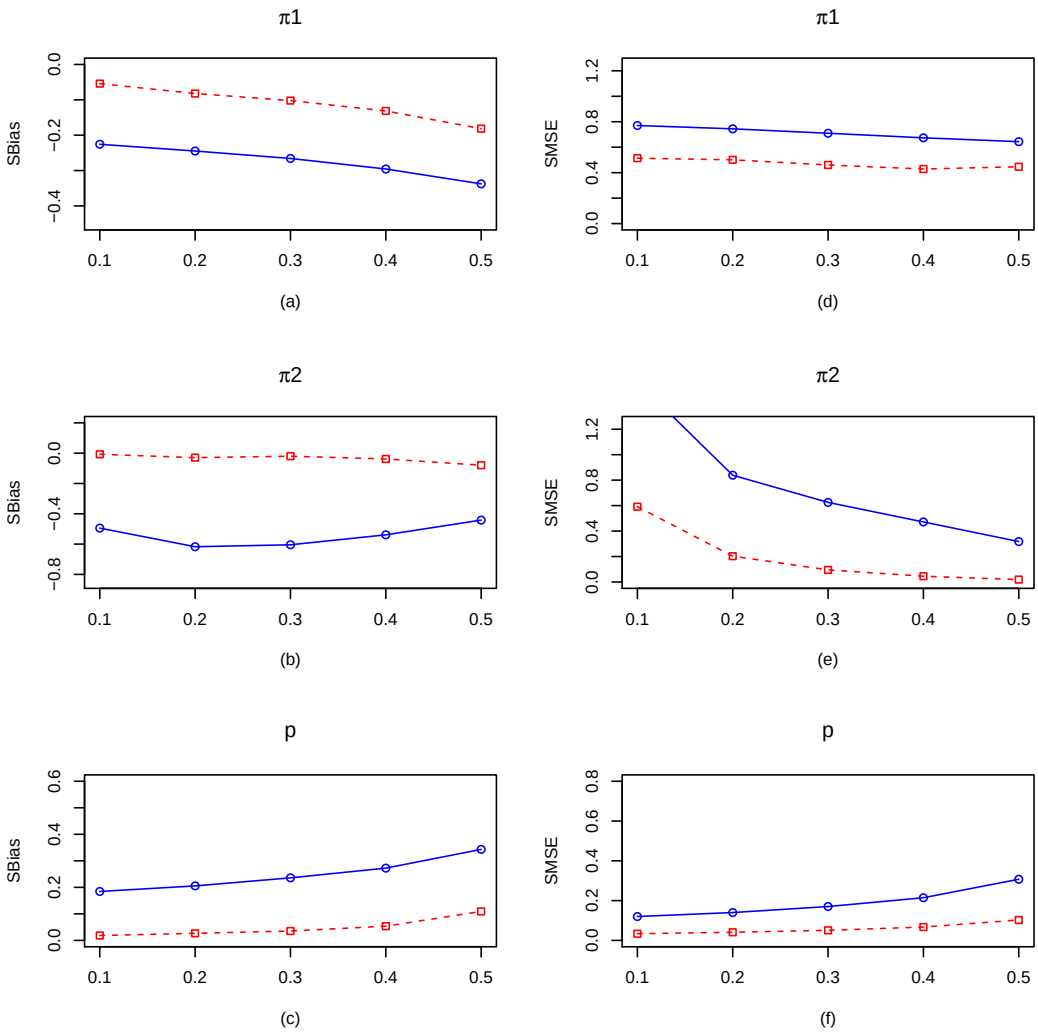

Figure 3.6: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2 and p (from ZOIG distribution) by varying π_2 for fixed $\pi_1 = 0.15$, $p = 0.3$ and $n = 50$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(c) Comparisons of SBiases of π_1, π_2 and p estimators respectively. (d)-(f) Comparisons of SMSEs of π_1, π_2 and p estimators respectively.

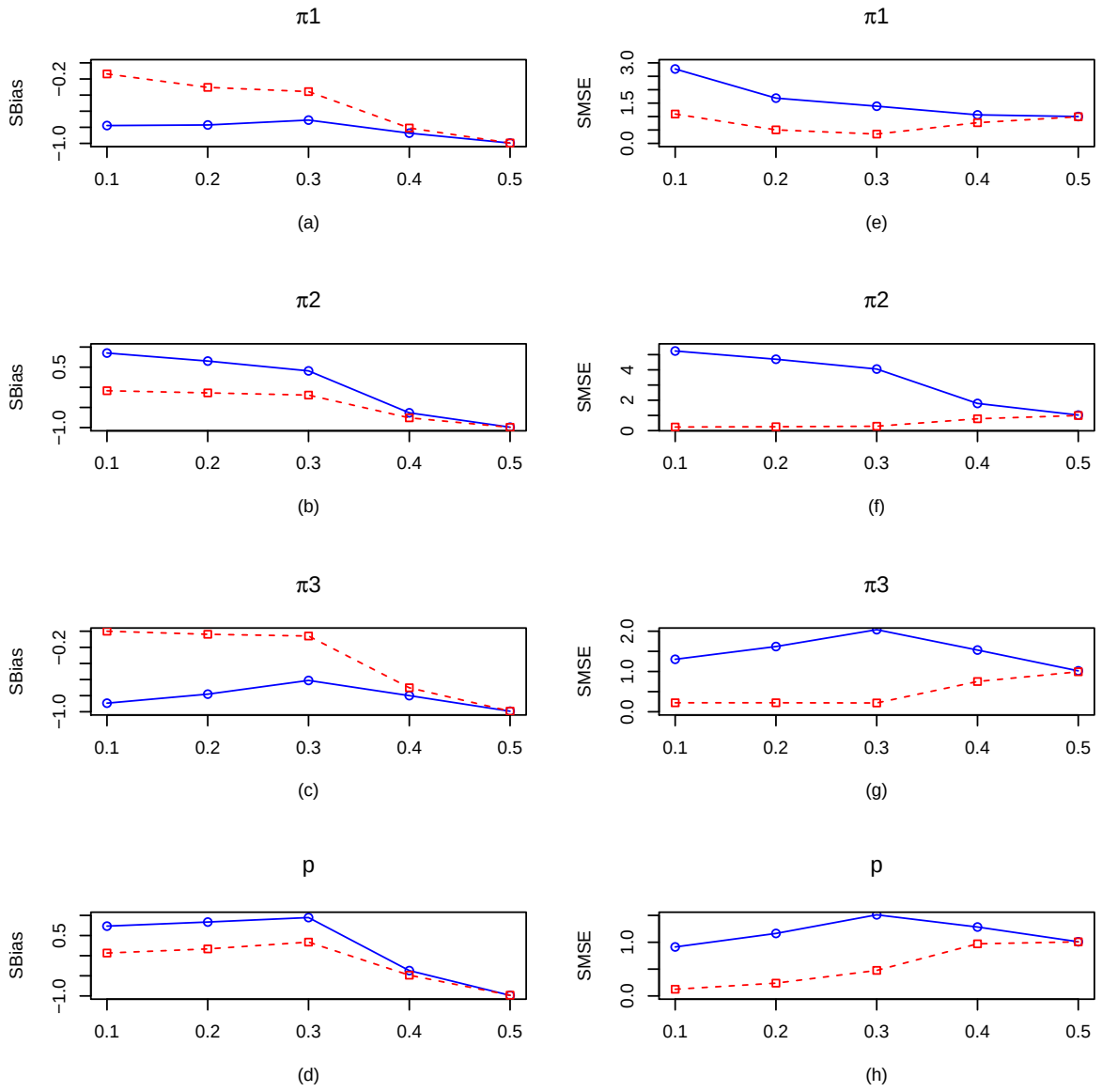

Figure 3.7: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2, π_3 and p (from ZOTIG distribution) by varying π_1 for fixed $\pi_2 = \pi_3 = 0.2$ and $p = 0.3$ and $n = 25$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(d) Comparisons of SBiases of π_1, π_2, π_3 and p estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and p estimators respectively.

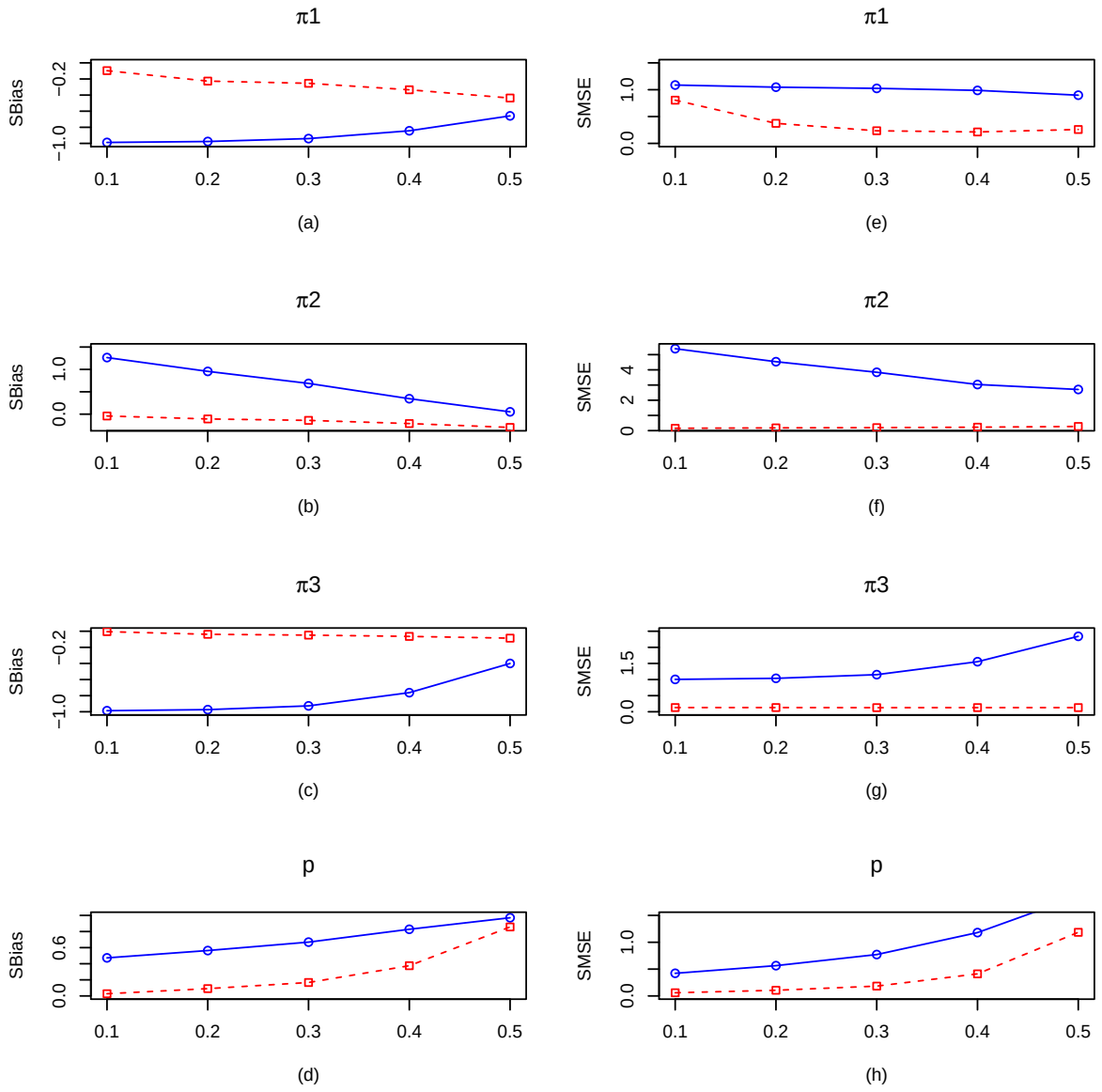

Figure 3.8: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2, π_3 and p (from ZOTIG distribution) by varying π_1 for fixed $\pi_2 = \pi_3 = 0.2$ and $p = 0.3$ and $n = 50$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(d) Comparisons of SBiases of π_1, π_2, π_3 and p estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and p estimators respectively.

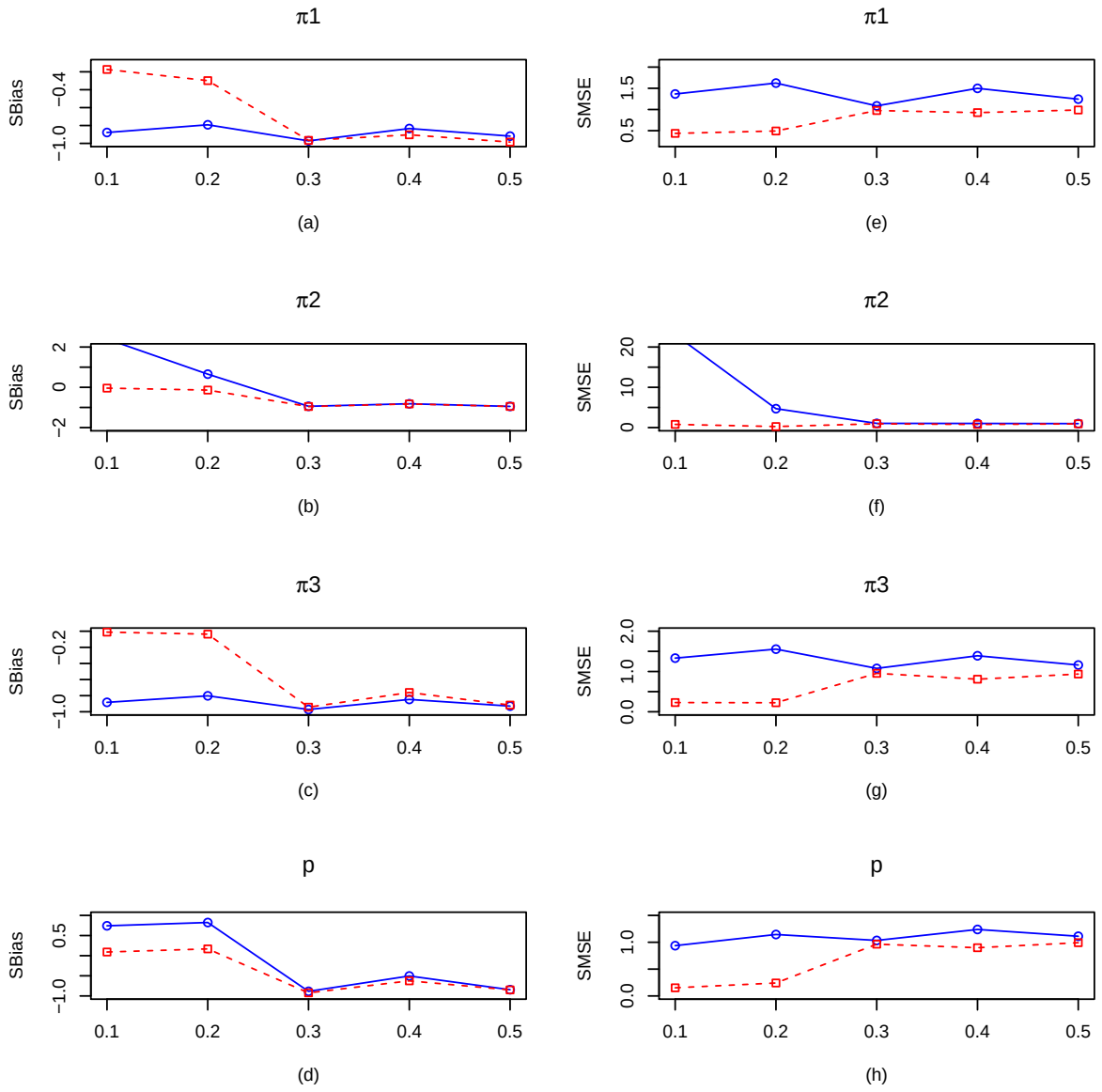

Figure 3.9: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2, π_3 and p (from ZOTIG distribution) by varying π_2 for fixed $\pi_1 = \pi_3 = 0.2$ and $p = 0.3$ and $n = 25$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(d) Comparisons of SBiases of π_1, π_2, π_3 and p estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and p estimators respectively.

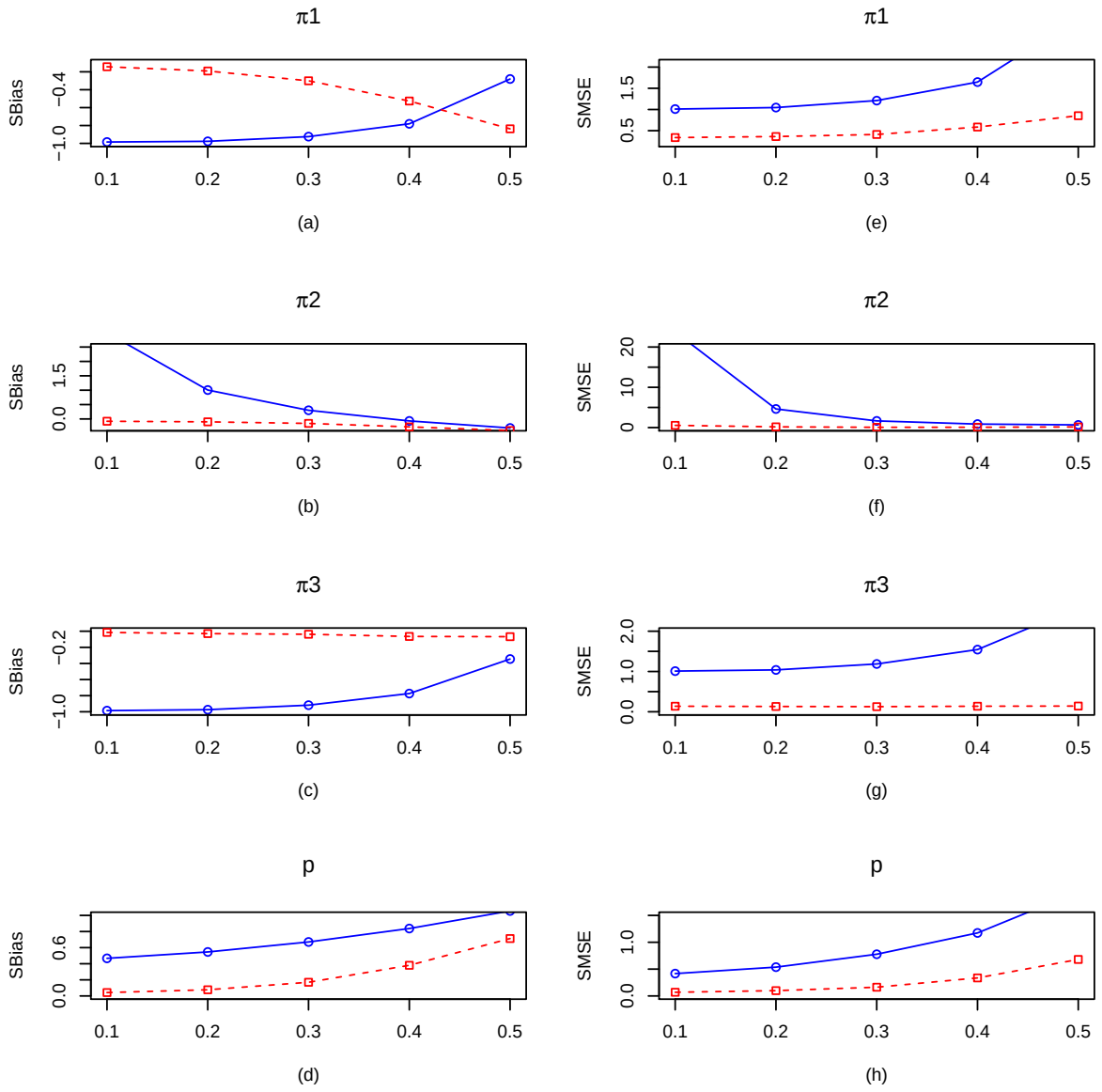

Figure 3.10: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2, π_3 and p (from ZOTIG distribution) by varying π_2 for fixed $\pi_1 = \pi_3 = 0.2$ and $p = 0.3$ and $n = 50$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(d) Comparisons of SBiases of π_1, π_2, π_3 and p estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and p estimators respectively.

SMSE. For sample size 50 from Figure 3.10, we see that MLEs of all the four parameters are performing better than the CMMEs with respect to both SBias and SMSE except in the case of π_1, where Sbias of MLE of π_1 is larger than that of CMME beyond 0.43.

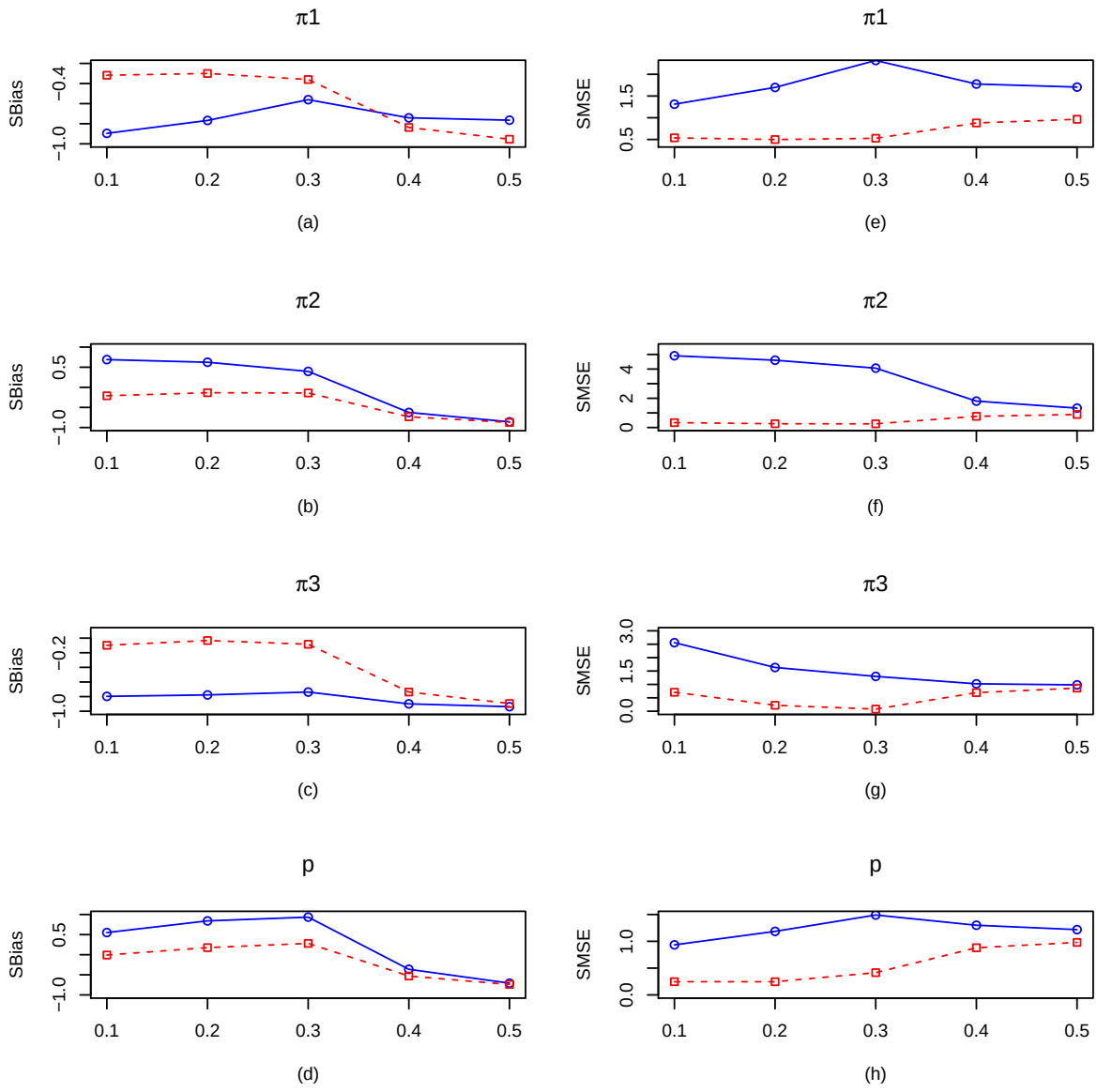

Figure 3.11: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2, π_3 and p (from ZOTIG distribution) by varying π_3 for fixed $\pi_1 = \pi_2 = 0.2$ and $p = 0.3$ and $n = 25$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(d) Comparisons of absolute SBiases of π_1, π_2, π_3 and p estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and p estimators respectively.

In the third scenario which is presented in Figures 3.11 and 3.12, we vary π_3 keeping π_1, π_2 and p fixed at 0.2, 0.2 and 0.3 respectively. Here also we observe similar results as the second case of ZOTIG distribution, MLEs for all the four parameters are uniformly outperforming CMMEs with respect to SBias except for π_1. Also as before MLEs uniformly outperform CMMEs of all the four parameters with respect to SMSE.

Thus from our simulation study it is evident that MLE has an overall better performance than CMME for all the Generalized Inflated Geometric models that we have considered. So in the next chapter, we consider an example where we fit an appropriate GIG model to a real life data set.

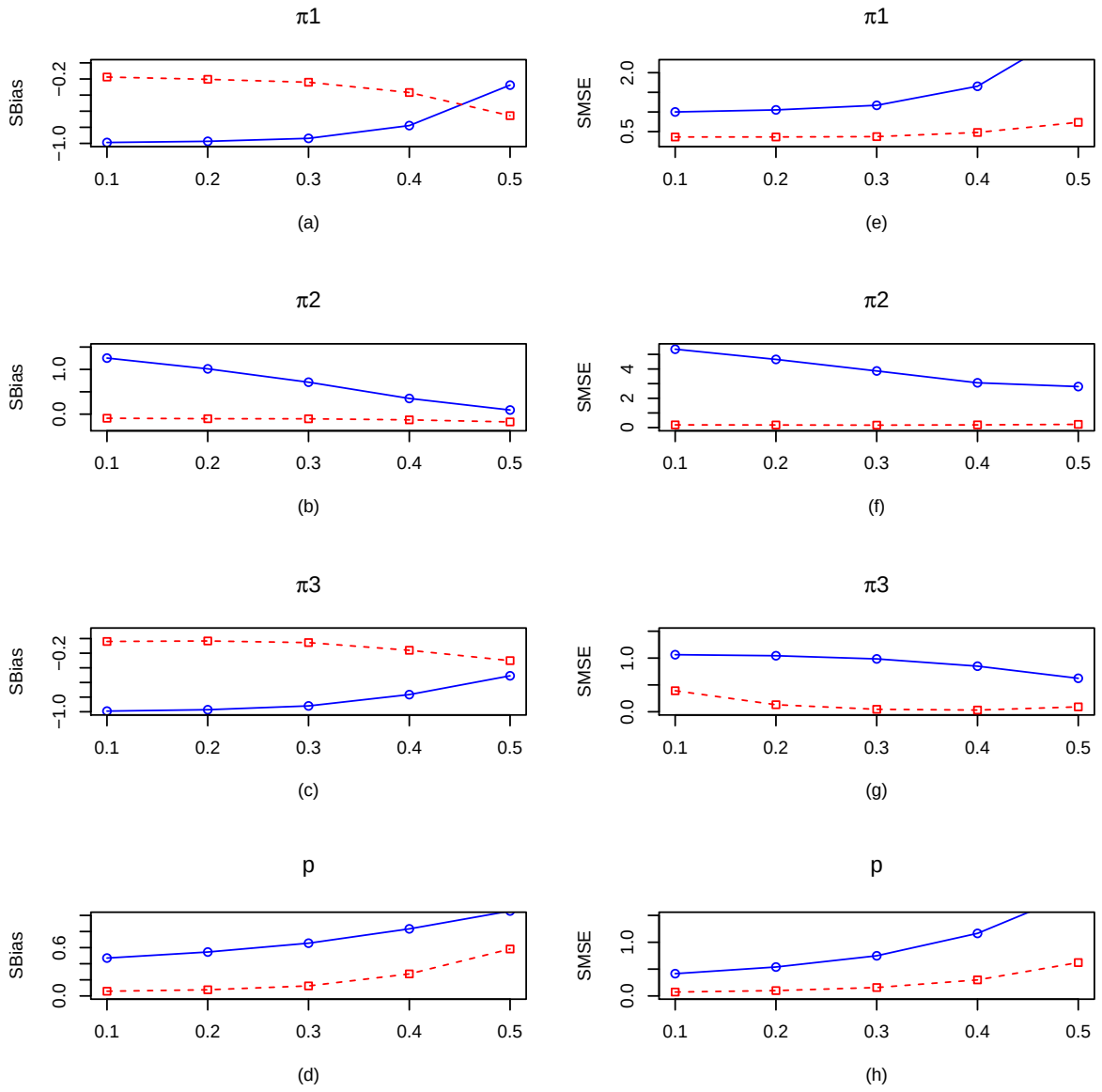

Figure 3.12: Plots of the SBias and SMSE of the CMMEs and MLEs of π_1, π_2, π_3 and p (from ZOTIG distribution) by varying π_3 for fixed $\pi_1 = \pi_2 = 0.2$ and $p = 0.3$ and $n = 50$. The solid line represents the SBias or SMSE of the corrected MME. The dashed line represents the SBias or SMSE of the MLE. (a)-(d) Comparisons of absolute SBiases of π_1, π_2, π_3 and p estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and p estimators respectively.

CHAPTER 4

APPLICATION OF GIG DISTRIBUTION

4.1 An Example

In this section, we consider the Swedish fertility data presented in Table 1.1 and try to fit a suitable GIG model. Since our simulation study in the previous chapter suggests that the MLE has an overall better performance, all of our estimations of model parameters are carried out using the maximum likelihood estimation approach. While fitting the Inflated Geometric models, parameter estimates of some mixing proportions (π_i) came out negative. So we need to make sure that all the estimated probabilities according to the fitted models are non-negative. We tried all possible combinations of GIG models and then we compared each of these Inflated Geometric models using the Chi-square goodness of fit test, the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). While performing the Chi-square goodness of fit test, the last three categories of Table 1.1 are collapsed into one group due to small frequencies. More details of our model fitting is presented below.

First, we try with single-point inflation at each of the four values (0, 1, 2 and 3). In this first phase, an inflation at 2 seems most plausible as it gives the smallest AIC and BIC values 4188.794 and 4198.924 respectively. However the p-value of the Chi-square test is very close to 0, suggesting that this is not a good model. Next, we try two-point inflations at {0, 1}, {0, 2}, {0, 3}, {1, 2}, etc. At this stage, {2, 3} inflation seems most appropriate going by the values of AIC (3947.064) and BIC (3962.258). But p-value of the Chi-square test close to 0 again makes it an inefficient model.

In the next stage, we try three-point inflation models, and here we note that a GIG with inflation set {0, 2, 3} significantly improves over the earlier {2, 3} inflation model (i.e., TTIG). This ZTTIG model significantly improves the p-value (but still close to 0) while maintaining a low AIC and BIC of 3841.169 and 3861.428 respectively. The main reason for low p-value is that this model is unable to capture the tail behavior. The estimated value of the parameters are (with $k_1 = 0, k_2 = 2, k_3 = 3$): $\hat{\pi}_1$ = -0.2340517, $\hat{\pi}_2 = 0.2816816$, $\hat{\pi}_3 = 0.1376353$, and $\hat{p} =$

0.4068477 using the Maximum Likelihood Estimation approach. Interpretation of the negative $\hat{\pi}_1$ is very difficult in this case, perhaps it can be thought of as a deflation point.

Finally, we fitted the full {0, 1, 2, 3} inflated model. We obtained the maximum likelihood estimates of the model parameters as $\hat{\pi}_1$ = -2.60853103, $\hat{\pi}_2$ = -0.92017233, $\hat{\pi}_3$ = -0.04498528, $\hat{\pi}_3$ = 0.02737037, and $\hat{p}$ = 0.59520603. The AIC and the BIC values for this model are 3800.688 and 3826.012 respectively. Which is significantly lower than all the previous models. Also p-value of the Chi-square test is 0.810944, thus rendering this Zero-One-Two-Three inflated Geometric (ZOTTIG) model to be a very good fit. For the sake of completeness, we have included a plot of regular geometric model and the ZOTTIG model in Figure 4.1. It is evident from this plot that the ZOTTIG model is performing way better than the regular geometric model for the Swedish fertility data.

4.2 Conclusion

This work deals with a general inflated geometric distribution (GIG) which can be thought of as a generalization of the regular Geometric distribution. This type of distribution can effectively model datasets with elevated counts. We have outlined the parameter estimation procedure for this distribution using the method of moments estimation and the maximum likelihood estimation techniques. Simulation studies were also performed and we found that MLEs performed better than the corrected MMEs in estimating the model parameters with respect to the standardized bias (SBias) and standardized mean squared errors (SMSE). While performing the simulation, we observed that for certain ranges of the inflated proportions in the GIG models, the computation algorithm for calculating MLEs did not converge. Nonetheless, we selected all permissible values and compared the overall performance of the MLEs and CMMEs for three special cases of GIG. Different GIG models were obtained by analyzing the fertility data of Swedish women, it is found that the Zero-One-Two-Three inflated Geometric (ZOTTIG) model is a good fit. Because of the extra parameter(s), the GIG distribution seems to be much more flexible in model fitting than the regular geometric distribution.

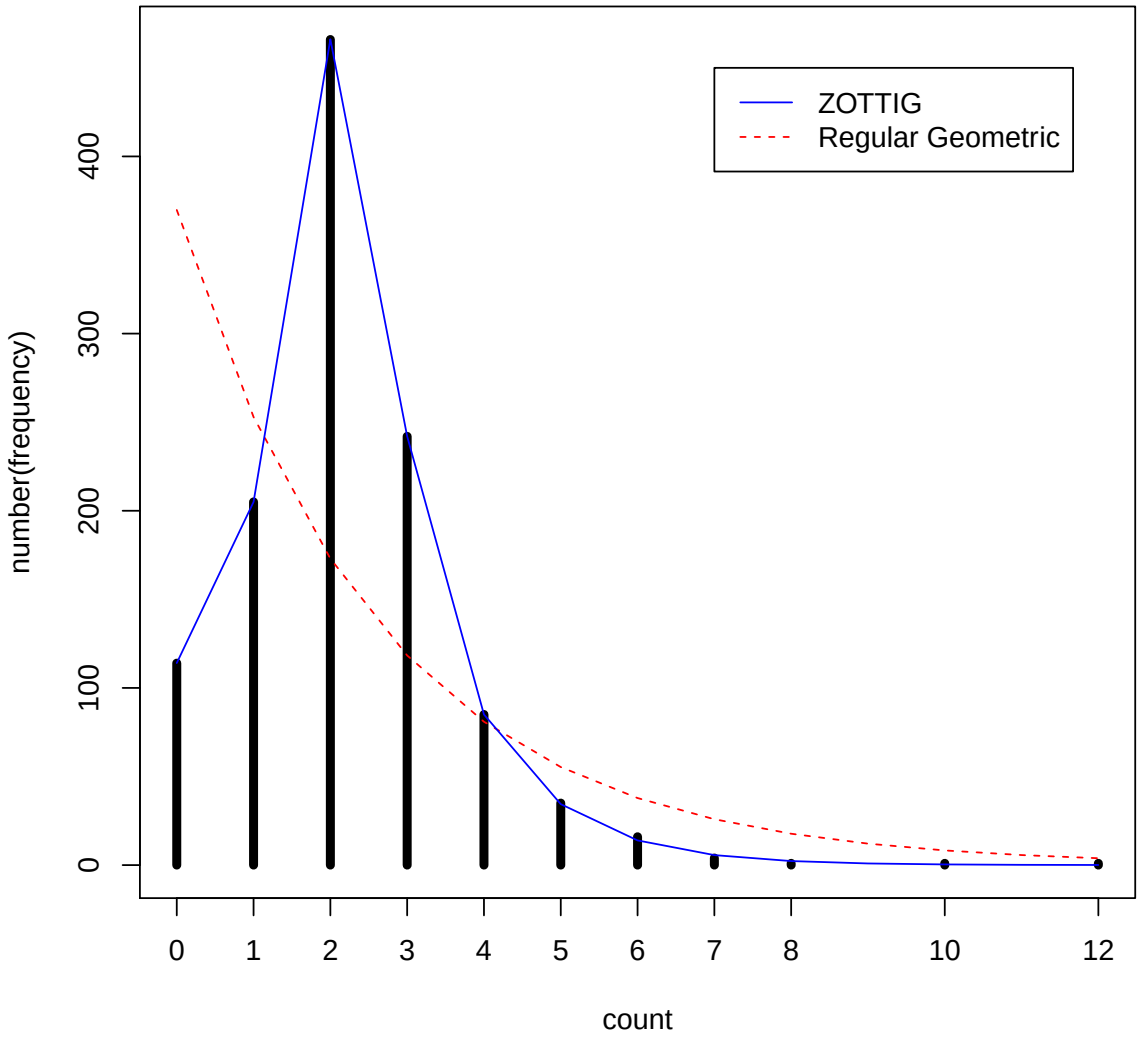

Figure 4.1: Plot of the observed frequencies compared to the estimated frequencies from the regular Geometric and the Zero-One-Two-Three Inflated Geometric model.

APPENDIX A

Algebraic Solutions for the Method of Moments Estimators

The general solutions for the system of equations (2.6) obtained using Mathematica is given below:

$$\begin{aligned}
\hat{\pi}_1 &= \frac{1}{2(2m_1' - 3m_2' + m_3')^2}(8{m_1'}^2 + 7{m_1'}^3 - 24m_1'm_2' - 24{m_1'}^2m_2' + 18{m_2'}^2 + 33m_1'{m_2'}^2 - 18{m_2'}^3 \\
&\quad + 8m_1'm_3' - 5{m_1'}^2m_3' - 12m_2'm_3' + 6m_1'm_2'm_3' + 3{m_2'}^2m_3' + 2{m_3'}^2 - 2m_1'{m_3'}^2) \\
\hat{\pi}_2 &= -1 + 2m_1' + \frac{3(m_1' - m_2')}{m_1' - m_3'} - \frac{3(m_1' - m_2')m_2'}{m_1' - m_3'} + \frac{1}{2(2m_1' - 3m_2' + m_3')^2}(8{m_1'}^2 + 7{m_1'}^3 - 24m_1'm_1'm_2' \\
&\quad - 24{m_1'}^2m_2' + 18{m_2'}^2 + 33m_1'{m_2'}^2 - 18{m_2'}^3 + 8m_1'm_3' - 5{m_1'}^2m_3' - 12m_2'm_3' + 6m_1'm_2'm_3' + 3{m_2'}^2m_3' \\
&\quad + 2{m_3'}^2 - 2m_1'{m_3'}^2) - \frac{3(m_1' - m_2')}{2(m_1' - m_3')(2m_1' - 3m_2' + m_3')^2}(8{m_1'}^2 + 7{m_1'}^3 - 24m_1'm_2' - 24{m_1'}^2m_2' + 18{m_2'}^2 \\
&\quad + 33m_1'{m_2'}^2 - 18{m_2'}^3 + 8m_1'm_3' - 5{m_1'}^2m_3' - 12m_2'm_3' + 6m_1'm_2'm_3' + 3{m_2'}^2m_3' + 2{m_3'}^2 - 2m_1'{m_3'}^2) \\
\hat{p} &= \frac{3(m_1' - m_2')}{(m_1' - m_3')},
\end{aligned}$$

where m_r' is the r^{th} raw sample moment of the ZOIG distribution.

For ZOTIG model the Mathematica solutions for the system of equations (2.7) are:

$$\begin{aligned}
\hat{\pi}_1 = \frac{1}{6(6m_1' - 11m_2' + 6m_3' - m_4')^3}(&1296{m_1'}^3 + 424{m_1'}^4 - 7128{m_1'}^2m_2' - 3508{m_1'}^3m_2' + 13068m_1'{m_2'}^2 \\
&+ 11746{m_1'}^2{m_2'}^2 - 7986{m_2'}^3 - 17967m_1'{m_2'}^3 + 10299{m_2'}^4 + 3888{m_1'}^2m_3' - 136{m_1'}^3m_3' - 14256m_1'm_2'm_3' \\
&- 3456{m_1'}^2m_2'm_3' + 13068{m_2'}^2m_3' + 15238m_1'{m_2'}^2m_3' - 15378{m_2'}^3m_3' + 3888m_1'{m_3'}^2 - 696{m_1'}^2{m_3'}^2 \\
&- 7128m_2'{m_3'}^2 - 3444m_1'm_2'{m_3'}^2 + 9016{m_2'}^2{m_3'}^2 + 1296{m_3'}^3 + 104m_1'{m_3'}^3 - 2552m_2'{m_3'}^3 + 304{m_3'}^4 \\
&- 648{m_1'}^2m_4' + 652{m_1'}^3m_4' + 2376m_1'm_2'm_4'' - 2384{m_1'}^2m_2'm_4' - 2178{m_2'}^2m_4' + 2103m_1'{m_2'}^2m_4' \\
&+ 135{m_2'}^3m_4' - 1296m_1'm_3'm_4' + 1368{m_1'}^2m_3'm_4' + 2376m_2'm_3'm_4' - 2420m_1'm_2'm_3'm_4' - 204{m_2'}^2m_3'm_4' \\
&- 648{m_3'}^2m_4' + 636m_1'{m_3'}^2m_4' + 196m_2'{m_3'}^2m_4' - 64{m_3'}^3m_4' + 108m_1'{m_4'}^2 - 146{m_1'}^2{m_4'}^2 - 198m_2'{m_4'}^2 \\
&+ 303m_1'm_2'{m_4'}^2 - 63{m_2'}^2{m_4'}^2 + 108m_3'{m_4'}^2 - 146m_1'm_3'{m_4'}^2 + 30m_2'm_3'{m_4'}^2 + 4{m_3'}^2{m_4'}^2 - 6{m_4'}^3 \\
&+ 9m_1'{m_4'}^3 - 3m_2'{m_4'}^3).
\end{aligned}$$

$$\begin{aligned}
\hat{\pi}_2 = \frac{1}{3(6m_1' - 11m_2' + 6m_3' - m_4')^2}(&56{m_1'}^3 - 148{m_1'}^2m_2' - 54m_1'{m_2'}^2 + 249{m_2'}^3 + 144{m_1'}^2m_3' \\
&- 112m_1'm_2'm_3' - 300{m_2'}^2m_3' + 48m_1'{m_3'}^2 + 152m_2'{m_3'}^2 - 32{m_3'}^3 - 56{m_1'}^2m_4' + 120m_1'm_2'm_4' \\
&- 30{m_2'}^2m_4' - 56m_1'm_3'm_4' + 12m_2'm_3'm_4' + 4{m_3'}^2m_4' + 6m_1'{m_4'}^2 - 3m_2'{m_4'}^2)
\end{aligned}$$

$$\hat{\pi}_3 = \frac{-2{m_1'}^2 + 3m_1'm_2' + 3{m_2'}^2 - 2m_1'm_3' - 6m_2'm_3' + 4{m_3'}^2 + 3m_1'm_4' - 3m_2'm_4'}{6(6m_1' - 11m_2' + 6m_3' - m_4')}$$

$$\hat{p} = \frac{4(2m_1' - 3m_2' + m_3')}{(2m_1' - m_2' - 2m_3' + m_4')}$$

where m_r' is the r^{th} raw sample moment of the ZOTIG distribution.

APPENDIX B

LETTER FROM INSTITUTIONAL RESEARCH BOARD

MARSHALL UNIVERSITY
www.marshall.edu

Office of Research Integrity
Institutional Review Board

February 6, 2015

Ram Datt Joshi
1412 7th Ave, Apt No.6
Huntington, WV, 25701

Dear Ram Datt:

This letter is in response to the submitted thesis abstract to conduct an analysis of a Swedish data using inflated geometric distribution. After assessing the abstract it has been deemed not to be human subject research and therefore exempt from oversight of the Marshall University Institutional Review Board (IRB). The Code of Federal Regulations (45CFR46) has set forth the criteria utilized in making this determination. Since the information in this study does not involve human subjects as defined in the above referenced instruction it is not considered human subject research. If there are any changes to the abstract you provided then you would need to resubmit that information to the Office of Research Integrity for review and a determination.

I appreciate your willingness to submit the abstract for determination. Please feel free to contact the Office of Research Integrity if you have any questions regarding future protocols that may require IRB review.

Sincerely,

Bruce F. Day, ThD, CIP
Director

REFERENCES

[1] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *A limited memory algorithm for bound constrained optimization*, SIAM Journal on Scientific Computing **16** (1995), no. 5, 1190–1208.

[2] R. Chen, *A survillance system for congenital malformations*, Journal of the American Statistical Association **73** (1978), no. 362, 323–327.

[3] F. Eggenberger and G. Pólya, *Über die statistik verketetter vorgäge*, Zeitschrift für Angewandte Mathematik und Mechanik **1** (1923), 279–289 (German).

[4] W. Feller, *An introduction to probability theory and its applications*, 3rd ed., vol. 1, John Wiley & Sons, Inc, New York, 1968.

[5] ———, *An introduction to probability theory and its applications*, 3rd ed., vol. 2, John Wiley & Sons, Inc, New York, 1971.

[6] P. Jagers, *How many people pay their tram fares?*, Journal of the American Statistical Association **68** (1973), no. 344, 801–804.

[7] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate discrete distributions*, 3rd ed., John Wiley & Sons, Inc, Hoboken, New Jersey, 2005.

[8] M. Melkersson and D. O. Rooth, *Modeling female fertility using inflated count data models*, Journal of Population Economics **13** (2000), no. 2, 189–203 (English).

[9] G. P. Patil, M. T. Boswell, S. W. Joshi, and M. V. Ratnaparkhi, *Dictionary and classified bibliography of statistical distributions in scientific work: Discrete models*, vol. 1, International Co-operative Publishing House, 1984.

[10] E. C. Pielou, *Runs of one species with respect to another in transects through plant populations*, Biometrics **18** (1962), no. 4, 579–593.

[11] ———, *Runs of healthy and diseased trees in transects through an infected forest*, Biometrics (1963), 603–614.

[12] P. Stewart, *A generalized inflated poisson distribution*, Master's thesis, Marshall University, 2014.

[13] D. D. Wackerly, W. Mendenhall, and R. L. Scheaffer, *Mathematical statistics with applications*, 7th ed., Cengage Learning, 2008.

Ram Datt Joshi

Department of Mathematics
Marshall University
Huntington, WV 25755

Phone: (304)638-4285
Email: joshi3@marshall.edu

Education

- **M.A** *Statistics*
  Marshall University, Huntington, WV, 2015
  Thesis Advisor: Dr. Avishek Mallick
- **M.S.** *Mathematics*
  Tribhuvan University, Kathmandu, Nepal, 1999
- **Bachelor of Science**
  Tribhuvan University, Kathmandu, Nepal,1997 .

Publications

- *United Mathematics ; ISBN 9937-582-17-2*
  A book for High School level students in Nepal for Grade 11 as a co-author on it.

Working Experience

- Worked as a lecturer of Mathematics at Geomatic Institute of Technology, Purbanchal University, Nepal from 2008 to 2013.
- Worked as a lecturer of mathematics at National School of Sciences, Kathmandu from 2008 to 2013.
- Worked as a Mathematics teacher at Saraswati Higher Secondary School, Kailali, Nepal from 2000 to 2008.
- Teaching the courses such as "Using and Understanding Mathematics- A quantitative Reasoning Approach" and "College-Algebra" at Marshall University as an instructor of record, Marshall University, West Virginia.

Trainings

- Participated in 15 days long conference on Nonlinear Systems and Summer School, Kathmandu, Nepal; jointly organized by Embry-Riddle Aeronautical University, USA and Tribhuvan University, Nepal, during the period of June 3-17, 2013. The courses attended were:

  Advanced Nonlinear Partial Differential Equations(30 hrs)

  Numerical methods for Partial Differential Equations with Mat lab (30 hrs)

  Nonlinear Analysis (15 hrs).
- Completed a Critical Thinking Workshop Conducted by the Center for Teaching and Learning, Marshall University, WV.

Contributed Talks and Seminars

- Presented a talk on the topic "Generalized Inflated Geometric Distribution" at 99^{th} MAA Ohio Section meeting held at Marshall University from March 27-28, 2015.

Membership and Affiliation

- Member of American Statistical Association.
- Member of American Mathematical Society.
- Member of America-Nepal Mathematical Society.
- Life member of Nepal Mathematical Society.

Language Proficiency.

- Hindi, Nepali, English and Urdu.

Courses Taken

- **Undergraduate Courses at TU, Nepal**.
  Physics, Chemistry, Mathematics.
- **Graduate Courses at TU, Nepal**.
  Algebra,Mathematical Analysis,Complex variables and Differential Equations, Topology and Differential Geometry,Mechanics, Functional Analysis and Integration, Integral Transforms, Dynamics of Viscous Fluids and Distribution Theory.
- **Graduate Courses taken at Marshall University, West Virginia.**
  Advanced Calculus, Game Theory, Probability and Statistics, Biostatistics, Regression Analysis, Statistical Computation Using R, Stochastic Processes, Advanced Mathematical Statistics, Multivariate Statistics, Design of experiments and, Survival Analysis.