A GENERALIZED INFLATED POISSON 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
Patrick Stewart
Approved by
Dr. Avishek Mallick, Committee Chairperson
Dr. Laura Adkins
Dr. Alfred Akinsete

Marshall University
May 2014

ACKNOWLEDGEMENTS

I would like thank my thesis advisor, Dr. Avishek Mallick. He has helped me in numerous ways by giving me his advice, support, and guidance throughout the entire project. He has also assisted me by writing numerous recommendation letters and by challenging me in his classes.

I would also like to thank the other members of my thesis committee: Dr. Laura Adkins and Dr. Alfred Akinsete. Dr. Akinsete was always available to help me and to offer advice and is the most-wonderful chair of the math department. Dr. Laura Adkins is a wonderful professor who is always willing to help and answer questions that I may have.

Two other professors that I would like to thank are Dr. Gerald Rubin and Dr. Scott Sarra. Dr. Rubin is the professor who has challenged me the most and pushed me the hardest. I have learned so much from his classes, and I feel that he has thoroughly prepared me for a doctoral program. Also, his constant help, support, and advice on various situations are greatly appreciated. He is the professor who has made the largest impression on me throughout my time at Marshall University. Dr. Sarra is a professor who also has challenged me with his classes. Also, his wonderful help and support as my teaching mentor helped me through my teaching experience. I also appreciate both of these professors for their help in writing recommendation letters for me.

Dr. Carl Mummert also deserves my gratitude for being a constant source of help. He has helped me in numerous ways, and without his help I would not have made it through these past two years.

I would also like to thank my family for their support and help throughout all stages of my life. I would not have made it this far without them. Above all, glory to God for all things!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Count data with excess number of zeros, ones or twos are commonly encountered in experimental situations. In this thesis we have examined one such fertility data from Sweden. The standard Poisson distribution, which is widely used to model such count data, may not provide a good fit to model women's fertility (defined as the number of children per woman in her lifetime) in a specific population due to various cultural and sociological reasons. Therefore, the usual Poisson distribution is inflated at specific values suitably, as dictated by the societal norms, to fit the available data. The data set is examined using various tests and techniques to determine the validity of using a multi-point inflated Poisson distribution as compared to the standard Poisson distribution.

The various tests and techniques used include comparing the method of moment estimator of various multi-point inflated Poisson distributions along with the standard Poisson distribution. The maximum-likelihood estimators for Poisson distributions are also found and compared. Using simulation study, the maximum-likelihood and method of moment estimators were compared, and the maximum-likelihood estimator was found to have an overall better performance.

Validation for the results found involves using the Chi-square goodness of fit test on the various Poisson distributions. Another validation test involves comparing the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) of the various Poisson distributions. The results of the various tests and techniques demonstrate that a multi-point inflated Poisson distribution provides a better fit and model as compared to the standard Poisson distribution.

Keywords and Phrases: Maximum-likelihood estimation, Method of moment estimator, Chi-square Goodness of fit test.

CHAPTER 1
INTRODUCTION

The widely used Poisson distribution of a discrete random variable that stands for the number or count of statistically independent events occurring within a unit time or space has the probability mass function (pmf) given as

$$p(k|\lambda) = P(X = k) = \lambda^k exp(-\lambda)/k! \tag{1.1}$$

where $k = 0, 1, 2, ...$; and $\lambda > 0$. Apart from its property as a limiting distribution of a binomial distribution, one can find many other characterizations in Feller (1968, 1971) [3] [4]. The list of applications of the Poisson distribution is quite varied and long as indicated by some of the references below:

- The number of soldiers of the Prussian army killed accidentally by horse-kick per year (von Bortkiewicz(1898) [9]);
- The number of bankruptcies that are filed in a month (Jaggia and Kelly (2012)[5]);
- The number of arrivals at a car wash in one hour (Anderson et al. (2012)[1]);
- The number of network failures per day (Levine et al. (2011) [6]);
- The number of blemishes per sheet of white bond paper (Doane and Seward (2010)[2]);
- The number of a particular type of insect that can be found in a 1-square-foot farmland (Pelosi and Sandifer (2003)[8]);
- The number of births, deaths, marriages, divorces, suicides and homicides over a given period of time (Weiers (2008)[10]).

Note that the Poisson model in (1.1), henceforth known as "Poisson(λ)," has both mean and variance = λ, which can pose problems in some applications where variation may differ from the mean. It has been observed that in many applications the dispersion of Poisson(λ) underestimates

or overestimates the observed dispersion. This happens because the single parameter λ, over which the Poisson distribution is dependent, is often insufficient to describe the true observed distribution. In fact, in many cases, it is suspected that the overdispersion in the observed data is caused by population heterogeneity which goes unnoticed. This population heterogeneity is unobserved, in other words, the population consists of several subpopulations, but the subpopulation membership is not observed in the sample. A special form of heterogeneity is described by a 'two-mass distribution' giving mass π to count 0, and mass $(1-\pi)$ to the second class which follows Poisson(λ). The result of this 'two-mass distribution' is the so called 'Zero-Inflated Poisson distribution' or ZIP distribution with the probability mass function

$$p(k|\lambda,\pi) = \begin{cases} \pi + (1-\pi)e^{-\lambda} & \text{if } k = 0 \\ (1-\pi)p(k|\lambda) & \text{if } k = 0,1,2,.... \end{cases} \tag{1.2}$$

where $\lambda > 0$, $0 < \pi < 1$ and $p(k|\lambda)$ is given in (1.1).

A further generalization of (1.2) can be obtained by inflating the Poisson 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:

$$p(k|\lambda,\pi_i, 1 \le i \le m) = \begin{cases} \pi_i + (1-\sum_{i=1}^{m}\pi_i)p(k|\lambda) & \text{if } k = k_1, k_2...., k_m \\ (1-\sum_{i=1}^{m}\pi_i)p(k|\lambda) & \text{if } k \neq k_i, 1 \le i \le m \end{cases} \tag{1.3}$$

where $k = 0,1,2,....$; $\lambda > 0$, and $\pi_i \in (0,1)$, $1 \le i \le m$, $0 < \sum_{i=1}^{m}\pi_i < 1$. For the remaining part of this work, we will refer to (1.3) as the General Inflated Poisson (GIP) distribution which is the main focus of this work.

A special case of the GIP is the Zero-Two Inflated Poisson (ZTIP) obtained when using $k = 2$, with $k_1 = 0$ and $k_2 = 2$, which has been justified to model the Swedish women's fertility dataset by Melkersson and Rooth (2000) [7]. The fertility dataset they considered, which represents a sample of 1170 Swedish women of the age group 45-76 (as of 1991), is given in Table 1.1. It presents the number of child(ren) per woman who, in 1991, crossed the child-bearing age.

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

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

It has been suggested that the fertility of Swedish women tends to have higher counts of zeroes and twos. Zero children may be due to medical reasons or because some women might not have found the "right" man. On the other hand, a relative excess of twos may be explained by social processes, traditions of two-child family, and national institutional arrangements.

Whether a GIP with focus on $(0, 2)$ i.e., ZTIP (as argued by Melkersson and Rooth (2000) [7]) or a GIP with focus on $(0, 1, 2)$ (called 'Zero-One-Two Inflated Poisson' or ZOTIP) or some other set $\{k_1, k_2...., k_m\}$ is appropriate for the above data will be eventually decided by a proper goodness of fit test.

The rth raw moment of X having a GIP (i.e., GIP$(\lambda, \pi_i, 1 \leq i \leq m; k_1, ..., k_m)$) can be obtained from the following expression:

$$\begin{aligned} E(X^r) &= \sum_{i=1}^{m} k_i^r \pi_i + (1 - \sum_{i=1}^{m} \pi_i) \sum_{k=0}^{\infty} k^r p(k|\lambda) \\ &= \sum_{i=1}^{m} k_i^r \pi_i + (1 - \sum_{i=1}^{m} \pi_i) \mu_r^{'}. \end{aligned} \tag{1.4}$$

where $\mu_r^{'}$ is the rth raw moment of Poisson(λ) which can easily be found from its moment generating function (MGF) $exp\{\lambda(exp(t) - 1)\}$. Closed form expressions for the expectation $E(X)$

and variance $Var(X)$ of X for the ZTIP model can be obtained as

$$E(X) = 2\pi_2 + \lambda(1 - \pi_1 - \pi_2) \tag{1.5}$$

$$Var(X) = 4\pi_2(1 - \pi_2) + \lambda(1 - \pi_1 - \pi_2)\{1 + \lambda(\pi_1 + \pi_2) - 4\pi_2\} \tag{1.6}$$

In the next chapter, we first write the equations to obtain the method of moments estimators (MMEs) and then the maximum likelihood estimators (MLEs) of the parameters. In Chapter 3, we compare the performances of MMEs and MLEs for different GIP models using simulation studies. In Chapter 4, we revisit the dataset given in Table 1.1 and find the proper GIP model to fit the dataset.

CHAPTER 2

ESTIMATION OF GIP MODEL PARAMETERS

Given a random sample $X_1, ..., X_n$, i.e. independent and identically distributed (iid) observations from the GIP in (1.3) with parameters $\pi_1, ..., \pi_m$ and λ, we first discuss the point estimation of the parameters.

2.1 Method of Moments Estimation (MME)

The easiest way to obtain estimators of the parameters is through the method of moments estimation (MME). Assuming that the sample is a cross section of the population, we equate the first (m + 1) sample moments with their population moments, i.e., we obtain a system of (m + 1) equations of the form

$$m_r^{'} = \sum_{i=1}^{m} k_i^r \pi_i + (1 - \sum_{i=1}^{m} \pi_i)\mu_r^{'}(\lambda), r = 1, 2, ..., (m+1); \tag{2.1}$$

where $m_r^{'} = \sum_{j=1}^{n} X_j^r/n$ is the rth sample raw moments, and $\mu_r^{'}(\lambda) = (d^r/dt^r)exp\{\lambda(exp(t)-1)\}|_{t=0}$ is the rth raw moment of Poisson(λ). The values of $\pi_i, i = 1, 2, ..., m$, and λ obtained by solving the system of equations (2.1) are denoted by $\hat{\pi}_{i(MM)}$ and $\hat{\lambda}_{MM}$ respectively. The subscript "(MM)" indicates the MME approach. Note that all parameters are nonnegative, and hence all estimates also ought to be so. However, there is no guarantee that the corresponding MMEs of the parameters would obey this restriction. Hence, we propose 'corrected MMEs' as

$$\hat{\pi}_{i(MM)}^{(c)} = \hat{\pi}_{i(MM)} \text{ truncated at 0 and 1 and } \hat{\lambda}_{MM}^{(c)} = \hat{\lambda}^* \tag{2.2}$$

where $\hat{\lambda}^*$ is the solution of λ in (2.1) after substituting $\hat{\pi}_{i(MM)}^{(c)}$s. Later in Chapter 3, we will see in our simulation studies how to ensure that each $\hat{\pi}_{i(MM)}^{(c)}$ is between 0 and 1 as well as $\hat{\lambda}_{MM}^{(c)} > 0$.

In the special case of ZIP distribution, i.e., $m = 1, k_1 = 0$, we have only two parameters: π_1 and λ. The population mean and variance are, respectively,

$$E(X) = (1 - \pi_1)\lambda \text{ and } V(X) = \lambda(1 - \pi_1)(1 + \pi_1\lambda) \tag{2.3}$$

By equating the above expressions with sample mean ($\bar{X}$) and sample variance $s^2 = \sum_{j=1}^{n}(X_j - \bar{X})^2/(n-1)$ (which is an alternative approach instead of dealing with $m_1^{'}$ and $m_2^{'}$), we get the MMEs of π_1 and λ as $\hat{\pi}_{1(MM)} = (s^2 - \bar{X})/\{\bar{X}^2 + (s^2 - \bar{X})\}$ and $\hat{\lambda}_{MM} = \bar{X} + (s^2/\bar{X}) - 1$. Note that $\hat{\pi}_{1(MM)}$ becomes negative if $\bar{X} > s^2$. Hence, our corrected MMEs are

$$\hat{\pi}_{1(MM)}^{(c)} = max\{0, \hat{\pi}_{1(MM)}\} = \begin{cases} 0 & \text{if } \bar{X} > s^2 \\ \hat{\pi}_{1(MM)} & \text{if } \bar{X} \leq s^2 \end{cases} \tag{2.4}$$

$$\hat{\lambda}_{(MM)}^{(c)} = \begin{cases} \bar{X} & \text{if } \bar{X} > s^2 \\ \hat{\lambda}_{(MM)} & \text{if } \bar{X} \leq s^2 \end{cases} \tag{2.5}$$

In the above, $\hat{\lambda}_{(MM)}^{(c)}$ becomes $\hat{\lambda}^* = \bar{X}$ when $\bar{X} > s^2$, i.e., $\hat{\pi}_{1(MM)}^{(c)} = 0$. This is the estimated value of λ one obtains from (2.1) (for the special case of ZIP) after substituting $\hat{\pi}_{1(MM)} = 0$.

In another special case of GIP, the Zero-Two Inflated Poisson (ZTIP) distribution, i.e., $m = 2, k_1 = 0, k_2 = 2$, we have three parameters: π_1, π_2 and λ. To obtain the MMEs of π_1, π_2 and λ we equate the first three raw sample moments with their population counterparts. We obtain a system of three equations in three unknowns as follows:

$$\begin{aligned} 2\pi_2 + \lambda(1 - \pi_1 - \pi_2) &= m_1^{'} \\ 4\pi_2 + \lambda(1 + \lambda)(1 - \pi_1 - \pi_2) &= m_2^{'} \\ 8\pi_2 + \lambda(1 + 3\lambda + \lambda^2)(1 - \pi_1 - \pi_2) &= m_3^{'} \end{aligned} \tag{2.6}$$

In another special case of GIP, the Zero-One-Two Inflated Poisson (ZOTIP) distribution, i.e., $m = 3, k_1 = 0, k_2 = 1, k_3 = 2$, we have four parameters: π_1, π_2, π_3 and λ. To obtain the MMEs of π_1, π_2, π_3 and λ, we equate the first four raw sample moments with their population counterparts.

Thus we obtain a system of four equations in four unknowns as follows:

$$\begin{aligned}
\pi_2 + 2\pi_3 + \lambda(1 - \pi_1 - \pi_2 - \pi_3) &= m_1^{'} \\
\pi_2 + 4\pi_3 + \lambda(1 + \lambda)(1 - \pi_1 - \pi_2 - \pi_3) &= m_2^{'} \\
\pi_2 + 8\pi_3 + \lambda(1 + 3\lambda + \lambda^2)(1 - \pi_1 - \pi_2 - \pi_3) &= m_3^{'} \\
\pi_2 + 16\pi_3 + \lambda(1 + 7\lambda + 6\lambda^2 + \lambda^3)(1 - \pi_1 - \pi_2 - \pi_3) &= m_4^{'}
\end{aligned} \tag{2.7}$$

Algebraic solutions to these systems of equations, (i.e. the algebraic expressions for the MMEs of the parameters of interest) in (2.6) and (2.7) are obtained using Mathematica and are given in Appendix (B). We note that these solutions may not fall in the feasible regions of the parameter space, so we put restrictions to these solutions as discussed for the ZIP distribution to obtain the corrected MMEs.

2.2 Maximum Likelihood Estimation (MLE)

Another other approach of estimating parameters is the maximum likelihood estimation (MLE) method. Based on the data $\mathbf{X} = (X_1, ..., X_n)$, the likelihood function $L = L(\lambda, \pi_i, 1 \leq i \leq m; \mathbf{X})$ is defined as follows. Let Y_i = number of observations at k_i with inflated probability, i.e., $Y_i = \sum_{j=1}^{n} I(X_j = k_i), 1 \leq i \leq m$, where I is an indicator variable. Also, let $Y\bullet = \sum_{i=1}^{m} Y_i$ = total number of observations with inflated probabilities, n = total number of observations, and $(n - Y\bullet)$ = total number of non-inflated observations. Then,

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

Thus, the loglikelihood function $l^* = \ln L$ is

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

Since

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

where $c =$ (term free from the parameters), the loglikelihood function becomes

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

The MLEs, $\hat{\pi}_{iML}$,$1 \leq i \leq m$, and $\hat{\lambda}_{ML}$, are the values of π_i, $1 \leq i \leq m$, and λ which maximize l^* in (2.9) over the parameter space $\Theta = \{(\lambda, \pi_1, ..., \pi_m)|0 \leq \pi_i \leq 1, 1 \leq i \leq m; 0 \leq \sum_{i=1}^{m} \pi_i \leq 1, \lambda \geq 0\}$. There are user-friendly softwares available which allow direct optimization of a multivariate function. But if maximization of l^* is to be done by solving the system of equations, one can use the following traditional steps.

Taking partial derivatives of l^* w.r.t. the parameters and setting them equal to zero yields

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

where $p^{(\lambda)}(k_i|\lambda) = (\partial/\partial\lambda) p(k_i|\lambda) = p(k_i - 1|\lambda) - p(k_i|\lambda)$, and $p(-1|\lambda) \equiv 0$.

It is not clear whether the MME or the MLE provides overall better estimators. To the best of our knowledge, no comparative study has been reported in literature. Since the estimators do not have any general closed form expressions, simulation studies can provide some guidance about the performance of these two types of estimators. For this reason, we consider some special cases of the GIP with $m = 1$, 2 and 3 in the next chapter.

CHAPTER 3

SIMULATION STUDY

The following three cases are considered for our simulation study:

(i) $m = 1, k_1 = 0$ (Zero Inflated Poisson (ZIP) distribution)

(ii) $m = 2, k_1 = 0, k_2 = 2$ (Zero-Two Inflated Poisson (ZTIP) distribution)

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

For each special model mentioned above, we generate random data $X_1, ..., X_n$ from the distribution (with given parameter values) $N = 10000$ times. Let us denote a parameter (either π_i or λ) 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

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

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 are more informative. 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 than the actual bias and/or MSE.

3.1 The ZIP Distribution

In order to set the stage for the simulation study for the Zero Inflated Poisson (ZIP) distribution, we fix $\lambda = 3$ and vary π_1 from 0.1 to 0.8 with an increment of 0.1 for n = 25. The constrained optimization algorithm "L-BFGS-B" is implemented to obtain the maximum likelihood estimators (MLEs) of the parameters λ and π_1, and the 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 MMEs, we plot the absolute 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.

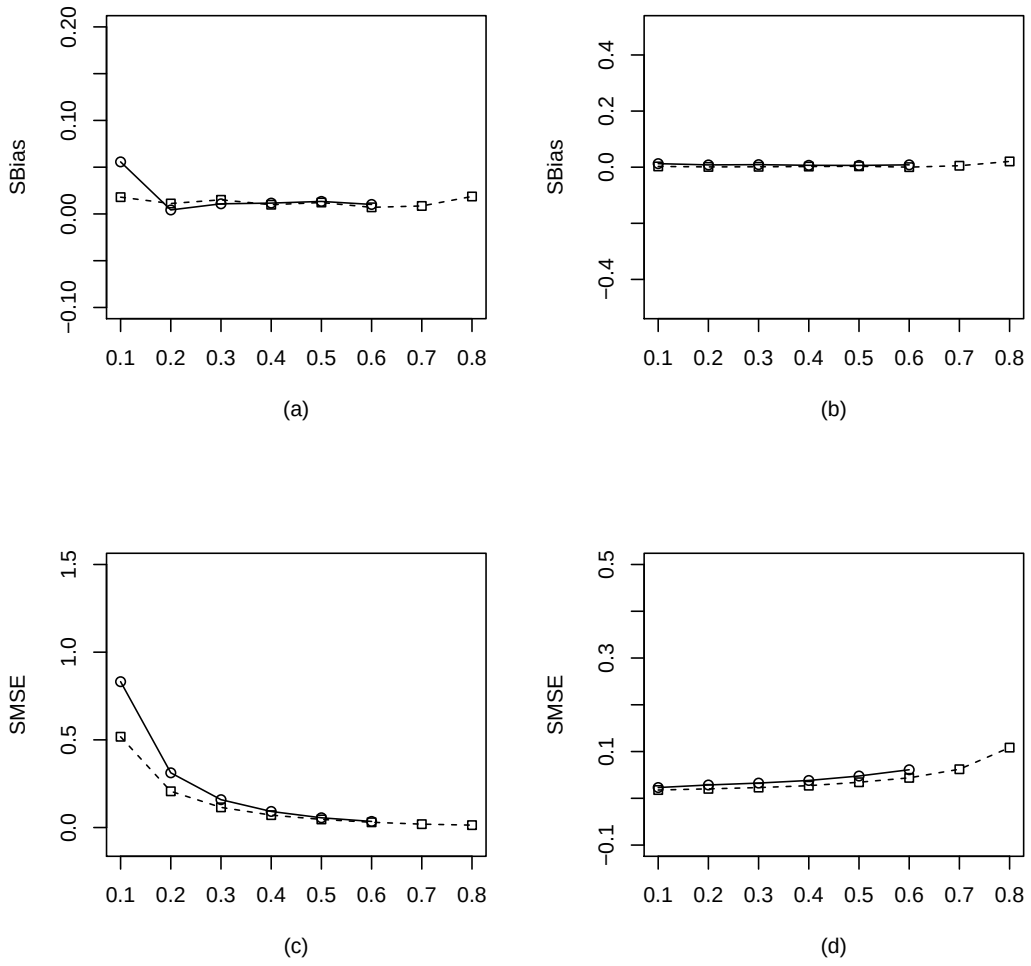

Figure 3.1: Plots of the absolute SBias and SMSE of the MMEs and MLEs of π_1 and λ (from ZIP distribution) plotted against π_1 for $\lambda = 3$ and $n = 25$. The solid line represents the absolute SBias or SMSE of the corrected MME. The dashed line represents the absolute SBias or SMSE of the MLE. (a) Comparison of absolute SBias of π_1 estimators. (b) Comparison of absolute SBias of λ estimators. (c) Comparison of SMSE of π_1 estimators. (d) Comparison of SMSE of λ estimators.

In Figure 3.1(a), we see that for the values of π_1 from 0.1 until about 0.18, MLE outperforms

MME with respect to SBias. However, MME outperforms the MLE from about 0.18 until around 0.35. From this point until about $\pi_1 = 0.6$, MLE slightly outperforms the MME. After this point, SBias of MME can no longer be calculated. The Sbias seems to be the smallest for MME at 0.2 and for MLE at around 0.4. In Figure 3.1(b), we see that MLE uniformly outperforms the MME until 0.6, after which again SBias of MME can no longer be calculated. They are both essentially unbiased since the SBias seems to be basically zero for MLE and around .01 or less for MME. In Figure 3.1(c), MLE consistently outperforms MME at all points until 0.6 where they both seem to have nearly the same SMSE. After 0.6, SMSE of MME cannot be calculated anymore. For both MLE and MME, the SMSE starts off at their highest values and then decreases rapidly until it reaches nearly zero. SMSE of MLE consistently outperforms that of MME in Figure 3.1(d). They both start off at their lowest values, and at this point, both MME and MLE has nearly the same SMSE. Again, SMSE of MME cannot be calculated after 0.6. It appears that SMSEs for both MME and MLE increase as values of π_1 get higher.

So we see for the ZIP distribution, the MLEs of the parameters π_1 and λ perform better than the MMEs almost everywhere over a certain range of π_1, namely 0.1 - 0.6, when the sample size is 25. We note that MLEs of both parameters have smaller absolute SBias and SMSE as compared to those of MMEs.

3.2 The ZTIP Distribution

In the case of the Zero-Two Inflated Poisson (ZTIP) distribution we have three parameters to consider, namely π_1, π_2 and λ. For fixed $\lambda = 3$ we vary π_1 and π_2 one at a time for sample size n = 25. Figure 3.2 presents the six comparisons for $\hat{\pi}^{(c)}_{1(MM)}$, $\hat{\pi}^{(c)}_{2(MM)}$ and $\hat{\lambda}^{(c)}_{(MM)}$ with $\hat{\pi}_{1(ML)}$, $\hat{\pi}_{2(ML)}$ and $\hat{\lambda}_{(ML)}$ in terms of absolute standardized bias and standardize MSE for $n = 25$, varying π_1 from 0.1 to 0.4 and keeping π_2 and λ fixed at 0.15 and 3 respectively.

In Figure 3.2(a), MLE outperforms MME at all points with respect to absolute SBias. Both start above zero and decrease slightly until 0.2. Absolute SBias of MLE increases linearly until 0.3, but absolute SBias of MME stays same until this point. However, after 0.3 absolute SBias of both MME and MLE is increasing until the end. In Figure 3.2(b), we see that the MLE is essentially unbiased for all values of π_1, and absolute SBias of MME vary a lot and is always more

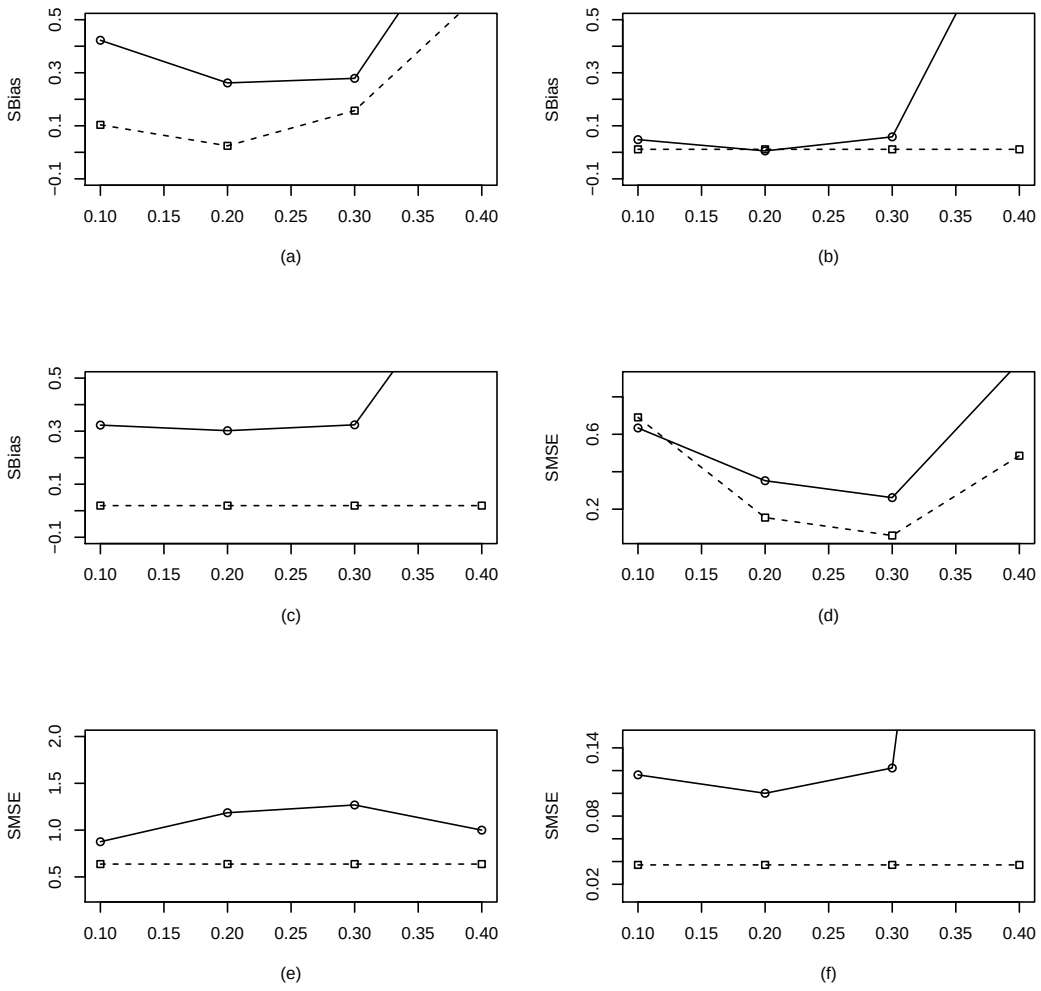

Figure 3.2: Plots of the absolute SBias and SMSE of the MMEs and MLEs of π_1, π_2 and λ (from ZTIP distribution) by varying π_1 for fixed $\pi_2 = 0.15$, $\lambda = 3$ and $n = 25$. The solid line represents the absolute SBias or SMSE of the corrected MME. The dashed line represents the absolute SBias or SMSE of the MLE. (a)-(c) Comparisons of absolute SBiases of π_1, π_2 and λ estimators respectively. (d)-(f) Comparisons of SMSEs of π_1, π_2 and λ estimators respectively.

than that of MLE. Thus for all permissible values of π_1, MME performs more poorly than MLE, except at $\pi_1 = 0.2$, where both are unbiased. Again in Figure 3.2(c), we see the same trend. MLE is unbiased throughout and MME is performing very poorly. In Figure 3.2(d), MME starts of with a lower SMSE than MLE. Both intersect at about $\pi_1 = 0.15$. After this point, MLE consistently outperforms the MME with respect to MSE. Both decrease until about 0.3 before going up, but SMSE of MLE stays below that of MME. In Figures 3.2(e) and 3.2(f), SMSE of MLE stays constant at 0.6 and 0.04 respectively for all permissible values of π_1. Also MME performs way worse for both the cases.

In our second scenario which is presented in Figure 3.3, we vary π_2 keeping π_1 and λ fixed at 0.15 and 3 respectively. In Figures 3.3(a) and 3.3(c), we see that MLE outperforms MME throughout with respect to absolute SBias. Moreover MLE is unbiased at $\pi_2 = 0.2$ in Figure 3.3(a) and almost so at all values of π_2 in Figure 3.3(c). However in Figure 3.3(b), absolute SBias of MME starts off quite high, then it sharply decreases until $\pi_2 = 0.2$. After that MME performs nearly as well as the MLE. From Figures 3.3(d), 3.3(e) and 3.3(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.

3.3 The ZOTIP Distribution

For the Zero-One-Two Inflated Poisson (ZOTIP) distribution we have four parameters to consider, namely π_1, π_2, π_3 and λ. For fixed $\lambda = 3$ we vary π_1, π_2, and π_3 one at a time for sample size $n = 25$. Thus we have eight comparisons for $\hat{\pi}^{(c)}_{1(MM)}$, $\hat{\pi}^{(c)}_{2(MM)}$, $\hat{\pi}^{(c)}_{3(MM)}$ and $\hat{\lambda}^{(c)}_{(MM)}$ with $\hat{\pi}_{1(ML)}$, $\hat{\pi}_{2(ML)}$, $\hat{\pi}_{3(ML)}$ and $\hat{\lambda}_{(ML)}$. These comparisons in terms of absolute standardized bias and standardize MSE are presented in Figures 3.4-3.6.

In the first scenario of ZOTIP distribution, which is presented in Figure 3.4, we vary π_1 keeping π_2, π_3 and λ fixed at 0.2, 0.2 and 3 respectively. From Figure 3.4(a, b, c, d), we see that the MMEs of all the four parameters perform consistently worse than the MLEs. Also, the MLEs seem to be unbiased for all permissible values of π_1. Moreover absolute SBias as well as SMSE of MME of λ become infinite (or cannot be calculated) after $\pi_1 = 0.2$, which is evident from boxes (d) and (h). Also from the boxes in Figure 3.4 concerning the SMSE, we notice that the MMEs of all the

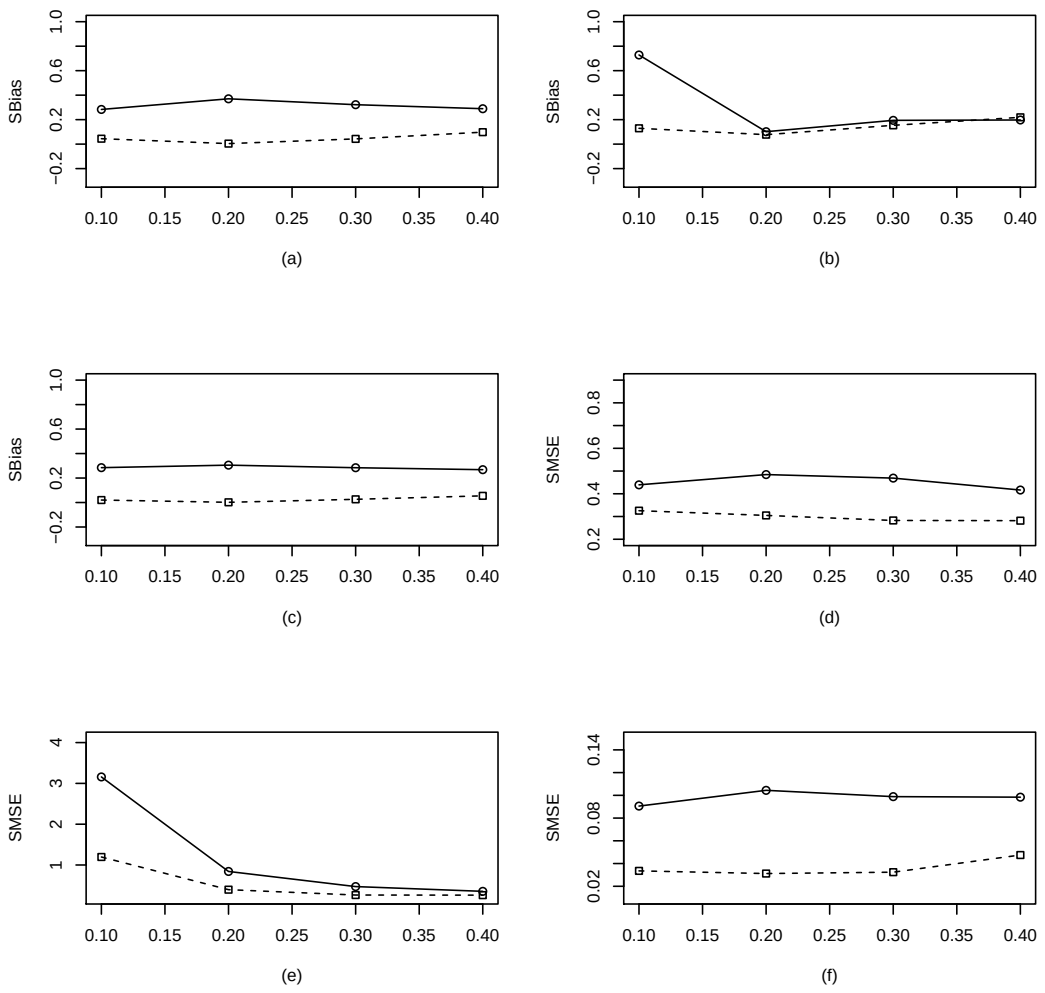

Figure 3.3: Plots of the absolute SBias and SMSE of the MMEs and MLEs of π_1, π_2 and λ (from ZTIP distribution) by varying π_2 for fixed $\pi_1 = 0.15$, $\lambda = 3$ and $n = 25$. The solid line represents the absolute SBias or SMSE of the corrected MME. The dashed line represents the absolute SBias or SMSE of the MLE. (a)-(c) Comparisons of absolute SBiases of π_1, π_2 and λ estimators respectively. (d)-(f) Comparisons of SMSEs of π_1, π_2 and λ estimators respectively.

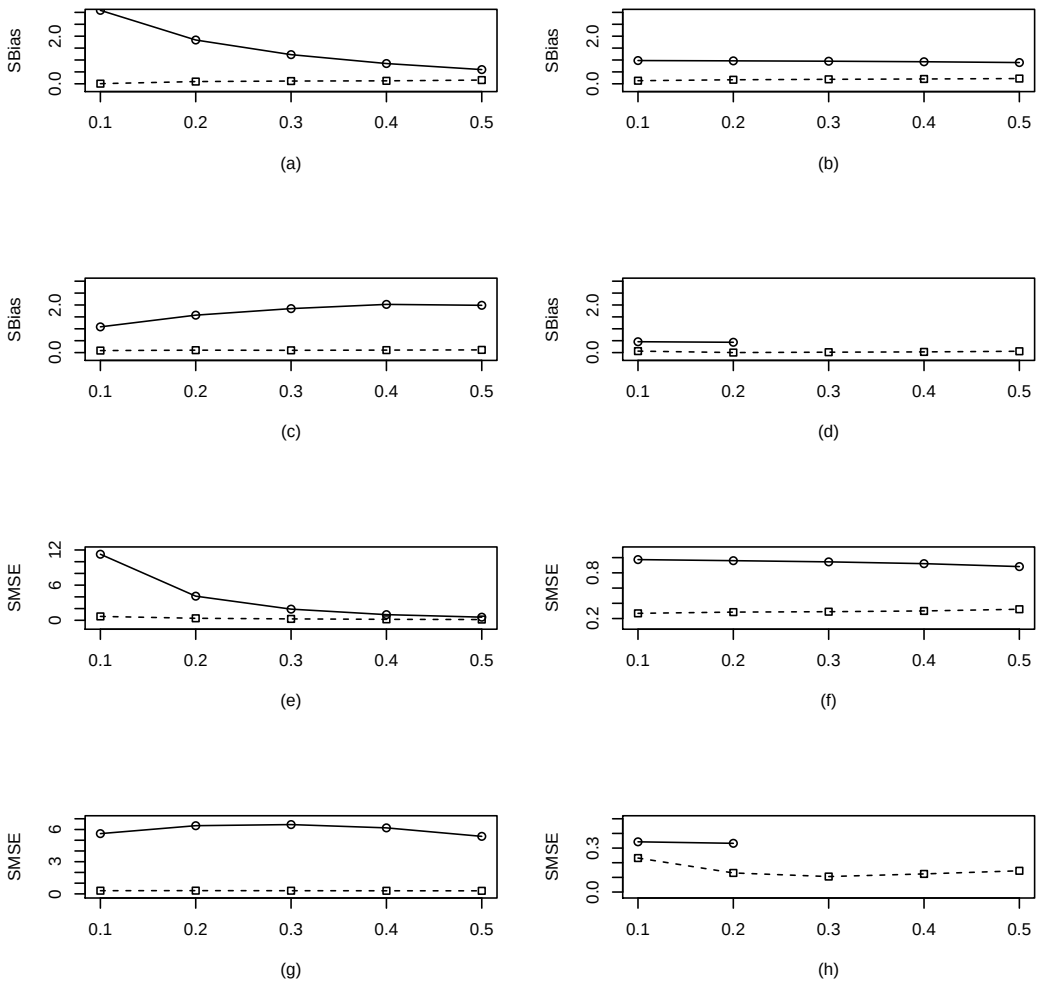

Figure 3.4: Plots of the absolute SBias and SMSE of the MMEs and MLEs of π_1, π_2, π_3 and λ (from ZOTIP distribution) by varying π_1 for fixed $\pi_2 = \pi_3 = 0.2$ and $\lambda = 3$ and $n = 25$. The solid line represents the absolute SBias or SMSE of the corrected MME. The dashed line represents the absolute SBias or SMSE of the MLE. (a)-(d) Comparisons of absolute SBiases of π_1, π_2, π_3 and λ estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and λ estimators respectively.

parameters perform consistently worse than the MLEs.

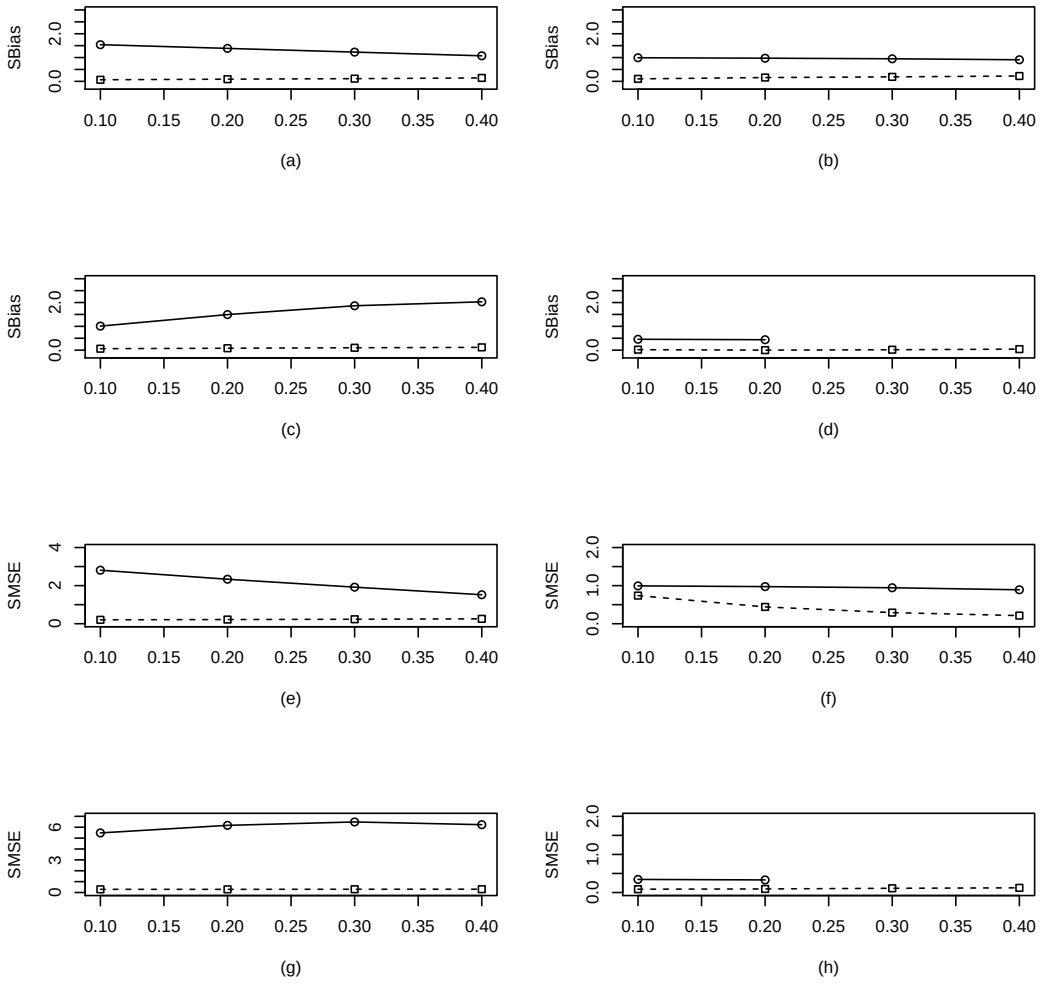

Figure 3.5: Plots of the absolute SBias and SMSE of the MMEs and MLEs of π_1, π_2, π_3 and λ (from ZOTIP distribution) by varying π_2 for fixed $\pi_1 = \pi_3 = 0.2$ and $\lambda = 3$ and $n = 25$. The solid line represents the absolute SBias or SMSE of the corrected MME. The dashed line represents the absolute SBias or SMSE of the MLE. (a)-(d) Comparisons of absolute SBiases of π_1, π_2, π_3 and λ estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and λ estimators respectively.

In our second scenario which is presented in Figure 3.5, we vary π_2 keeping π_1, π_3 and λ fixed at 0.2, 0.2 and 3 respectively. As before we see that the MLEs of all four parameters perform better than their MME counterparts with respect to both absolute SBias and SMSE. In particular the MME of λ performs the worst as its SBias and SMSE become infinite after $\pi_1 = 0.2$

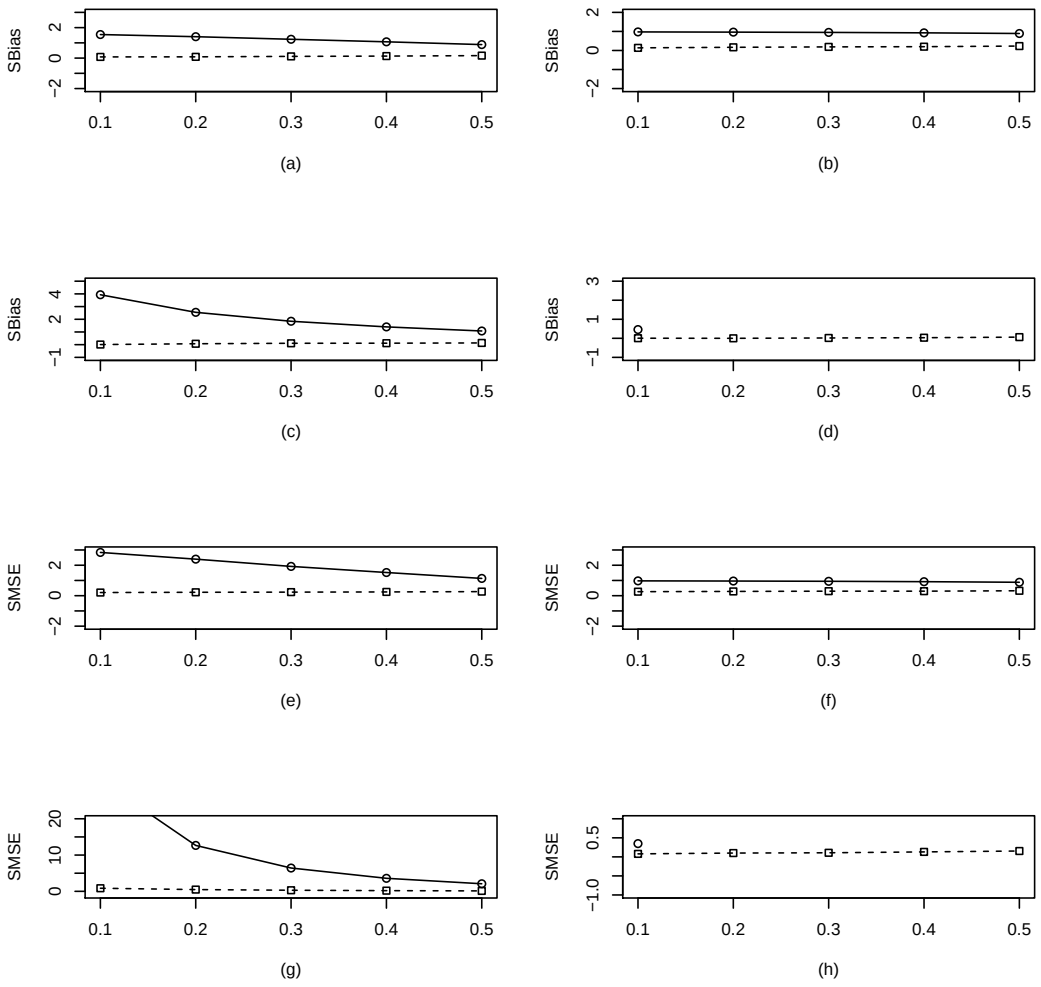

Figure 3.6: Plots of the absolute SBias and SMSE of the MMEs and MLEs of π_1, π_2, π_3 and λ (from ZOTIP distribution) by varying π_3 for fixed $\pi_1 = \pi_2 = 0.2$ and $\lambda = 3$ and $n = 25$. The solid line represents the absolute SBias or SMSE of the corrected MME. The dashed line represents the absolute SBias or SMSE of the MLE. (a)-(d) Comparisons of absolute SBiases of π_1, π_2, π_3 and λ estimators respectively. (e)-(h) Comparisons of SMSEs of π_1, π_2, π_3 and λ estimators respectively.

In the third scenario which is presented in Figure 3.6, we vary π_3 keeping π_1, π_2 and λ fixed at 0.2, 0.2 and 3 respectively. Here also we observe similar results as the first two cases of ZOTIP distribution, MLEs being unbiased for all the four parameters and uniformly outperforming MMEs. Also as before MLEs uniformly outperform MMEs of all the four parameters with respect to SMSE. SBias and SMSE of MME of λ become infinite just after $\pi_1 = 0.1$ as such performs even worse than both the previous cases.

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

CHAPTER 4

AN APPLICATION OF GIP DISTRIBUTION

In this chapter, we revisit the Swedish fertility data presented in Table 1.1. The objective here is to fit a suitable GIP. Melkersson and Rooth [7] proposed a ZTIP model for the dataset. But, our analysis shows that perhaps a ZTTIP ('Zero-Two-Three Inflated Poisson') is more suitable. Since our simulation study points out that the MLE has an overall better performance, all of our estimations of model parameters are carried out using this approach. For the sake of completeness, we have also included the MMEs. The details of our model fitting is presented below.

For various values of π_1, π_2, and π_3, we plotted the log-likelihood function of the ZTTIP for λ, as presented in Figure 4.1. Figure 4.1 demonstrates that for each value of π_i there exists one global maximum for λ. Therefore, we can conclude that using the MLE is a justified approach for estimating for λ. The same procedures were also carried out for each π_i, and in each case, there exists only one global maximum for each parameter.

Table 1.1 shows significantly high frequencies at the values 0, 1, 2 and 3. Therefore, we tried all possible combinations of GIP models. First, we try with single-point inflation at each of these four values (i.e., 0, 1, 2 and 3). In this first phase, an inflation at 2 seems most plausible as it gives the highest p-value. Next, we try two-point inflations at {0, 1}, {0, 2}, {0, 3}, {1, 2}, etc. At this stage, {2, 3} inflation seems the most appropriate going by both the p-value as well as AIC and BIC. This disproves the claim made by Melkersson and Rooth [7] that ZTIP, {0, 2}, is the best among the two-point inflated models. Table 4.1 gives the details from our model fitting. Table 4.1 includes all possible inflated Poisson models, chi-square goodness of fit test statistics, degrees of freedom (= number of categories in Table 1.1 - number of parameters in GIP model), p-values and AIC and BIC values. Note that the last three categories of Table 1.1 are collapsed into one due to small frequencies.

In the next stage, we try three-point inflation models, and here we note that a GIP with inflation set {0, 2, 3} significantly improves over the earlier {2, 3} inflation model (i.e., TTIP). This ZTTIP significantly improves the p-value while maintaining a low AIC and BIC. We fitted the full {0, 1,

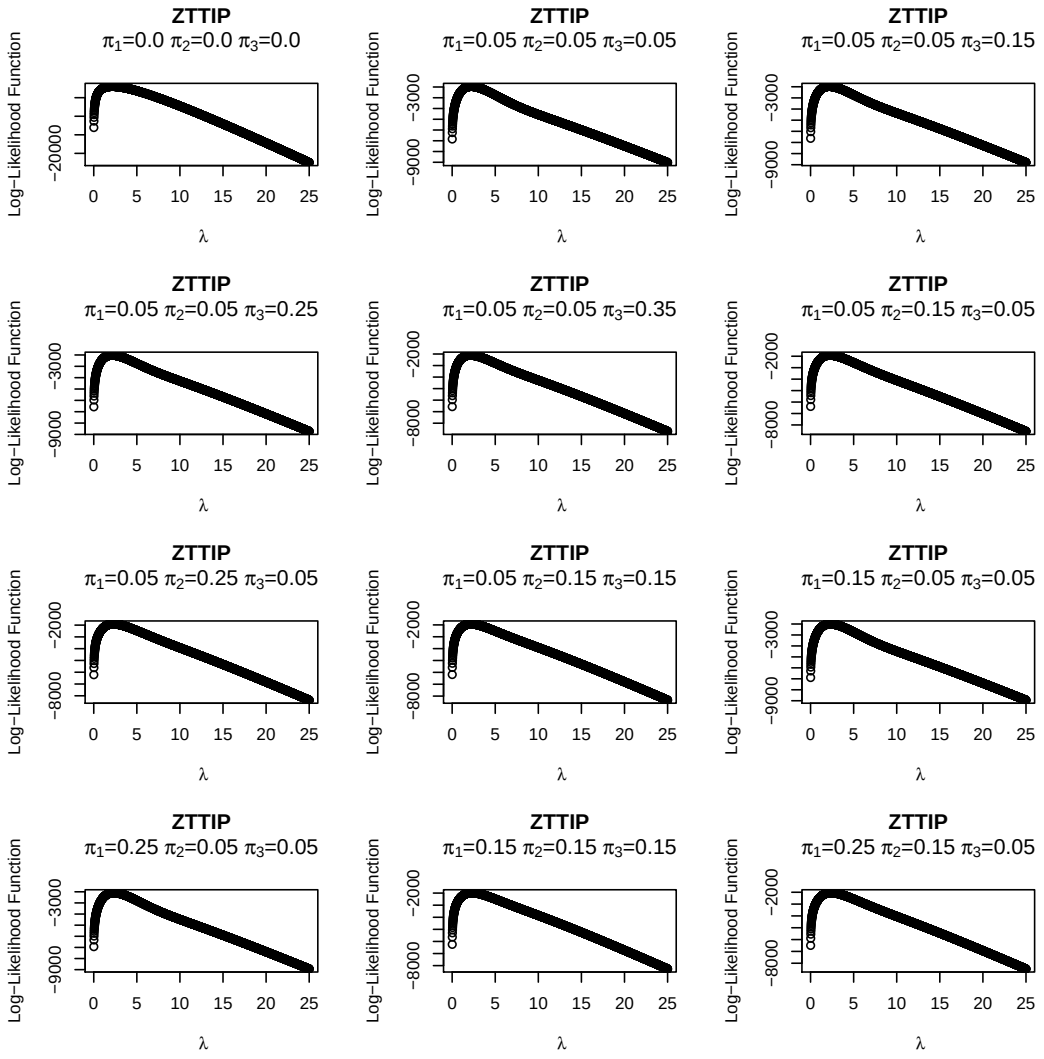

Figure 4.1: The graphs of the Log-Likelihood Function of a Zero-Two-Three Inflated Poisson with varying values of π_1, π_2, and π_3.

Inflation Points	$\hat{\lambda}$	$\hat{\pi}$	AIC	BIC	DF	Test Statistics	P-Value
Standard	2.161		3915.311	3863.858	7	121.566	3.615×10^{-23}
0	2.106	0	3912.301	3865.912	6	122.917	3.973×10^{-24}
1	2.036	0	3872.871	3826.482	6	122.917	3.973×10^{-24}
2	2.199	.178	3824.946	3778.557	6	25.989	.0002
3	2.149	.017	3915.994	3869.605	6	122.245	5.502×10^{-24}
0, 1	1.686	0, 0	3812.511	3777.187	5	264.135	5.085×10^{-55}
0, 2	2.229	.011, .018	3826.086	3784.762	5	105.619	3.449×10^{-21}
0, 3	2.092	0, .016	3913.121	3871.797	5	123.343	6.144×10^{-25}
1, 2	2.125	0, .157	3819.297	3777.973	5	29.169	2.148×10^{-05}
1, 3	2.036	0, .001	3874.868	3833.544	5	129.049	3.787×10^{-26}
2, 3	2.133	.201, .066	3806.158	3764.834	5	4.750	.447
0, 1, 2	1.881	0, 0, .094	3815.135	3778.876	4	91.531	6.224×10^{-19}
0, 1, 3	1.694	0, 0, 0	3820.068	3783.809	4	258.986	7.543×10^{-55}
0, 2, 3	2.188	.019, .209, .069	3805.485	3769.226	4	2.085	.720
1, 2, 3	2.104	0, .189, .060	3806.705	3770.446	4	5.535	.237
0, 1, 2, 3	2.439	.051, .062, .260, .095	3805.991	3774.797	3	1.376	.711

Table 4.1: Results of several Inflated-Poisson Models after running MLE and χ^2 Goodness of Fit Tests

2, 3} inflated model too, but since it does not enhance the p-value, AIC and BIC, we fall back on ZTTIP. This model seems quite reasonable given the Swedish fertility dataset. What it says is that a high percentage of Swedish women were found to be childless (for social and/or medical reasons). Those who have had children settled for mostly with two or three children; maybe the one-child pattern was not too attractive to the Swedish women.

The estimated value of the parameters are (with $k_1 = 0, k_2 = 2, k_3 = 3$): $\hat{\pi}_1 = 0.01872974$, $\hat{\pi}_2 = 0.20984665$, $\hat{\pi}_3 = 0.06938253$, and $\hat{\lambda} = 2.18828104$ using Maximum Likelihood Estimation approach. Using Method of Moments Estimation with $m = 3, k_1 = 0, k_2 = 2$, and $k_3 = 3$, we obtain the following system of four equations in four unknowns.

$$\begin{aligned} 2\pi_2 + 3\pi_3 + \lambda(1 - \pi_1 - \pi_2 - \pi_3) &= 2.164103 \\ 4\pi_2 + 9\pi_3 + \lambda(1 + \lambda)(1 - \pi_1 - \pi_2 - \pi_3) &= 6.463248 \\ 8\pi_2 + 27\pi_3 + \lambda(1 + 3\lambda + \lambda^2)(1 - \pi_1 - \pi_2 - \pi_3) &= 24.23077 \\ 16\pi_2 + 81\pi_3 + \lambda(1 + 7\lambda + 6\lambda^2 + \lambda^3)(1 - \pi_1 - \pi_2 - \pi_3) &= 116.2991 \end{aligned} \tag{4.1}$$

Solving these equations in Mathematica, we obtain estimated values as $\hat{\pi}_1 = .137431$, $\hat{\pi}_2 = .521204$, $\hat{\pi}_3 = .15034$, and $\hat{\lambda} = 3.51094$.

Based on the chosen ZTTIP model, we then compute the 'observed Fisher Information matrix' denoted by $\hat{I}$ as follows.

Let $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\pi_1, \pi_2, \pi_3, \lambda)$ for notational convenience. Then,

$$\begin{aligned} \hat{\mathbf{I}} &= -\sum_{l=1}^{n}((\frac{\partial^2}{\partial_i \partial_j} \ln f(X_l|\theta)))_{4\times 4}|_{\theta=\hat{\theta}_{ML}} \\ &= -((\frac{\partial^2}{\partial_i \partial_j} l^*))|_{\theta=\hat{\theta}_{ML}} \end{aligned} \tag{4.2}$$

where l^* is the log-likelihood function given in (2.9).

The asymptotic dispersion matrix of $\hat{\theta}_{ML}$ is the inverse of $\hat{\mathbf{I}}$, i.e.,

$$\hat{V}(\hat{\theta}_{ML}) \approx (\hat{\mathbf{I}})^{-1} = \begin{pmatrix} 0.0001331583 & 0.0000550534 & 0.0000189737 & 0.0003585604 \\ 0.0000550534 & 0.0000412525 & 0.0000566233 & 0.0001748525 \\ 0.0000189737 & 0.0000566233 & 0.0002306987 & -0.000121804 \\ 0.0003585604 & 0.0001748525 & -0.000121804 & 0.0039695545 \end{pmatrix}$$

Since our $n = 1170$ is sufficiently large, the above estimate of the dispersion matrix of $\hat{\theta}_{ML}$ should be valid for making inference. We note also that the ML estimates of the ZTTIP parameters $(\pi_1, \pi_2, \pi_3, \lambda)$ are asymptotically efficient with low variance. We compute asymptotic t-test statistics to test the null hypotheses that each parameter can be taken as zero. The four asymptotic t-test statistics are:

$t_{\pi 1} = 1.623109, t_{\pi 2} = 10.331817, t_{\pi 3} = 4.568016, t_{\lambda} = 34.732193.$

We note that according to the asymptotic t-test, we reject the null hypotheses that the inflated probabilities π_1, π_2, π_3 and the mean parameter λ are equal to zero.

The above t-statistic values are obtained by dividing the estimate of each parameter by its standard error (which is the square-root of the corresponding diagonal element of the asymptotic dispersion matrix). While all t-statistic values are substantially large, thereby implying that the corresponding parameter is nonzero, the t-statistic value for π_1 may look as a potential suspect. Note that π_1 is the extra (inflated) probability at $k_1 = 0$. Note that in GIP, we can only test a parameter to be zero against the one-sided alternative that the parameter be greater than zero (since no parameter under GIP can be negative). Using the normal curve as an approximation to t_{1166} (because the df = n - number of parameters in ZTTIP = 1170 - 4 = 1166), we get the p-value corresponding to the t-statistic value of 1.623109 as 0.052, which is not large, but rather a borderline case. Therefore, based on the Swedish fertility data, we conclude that all ZTTIP model parameters are significant, and the model is a good fit as evident from Figure 4.2.

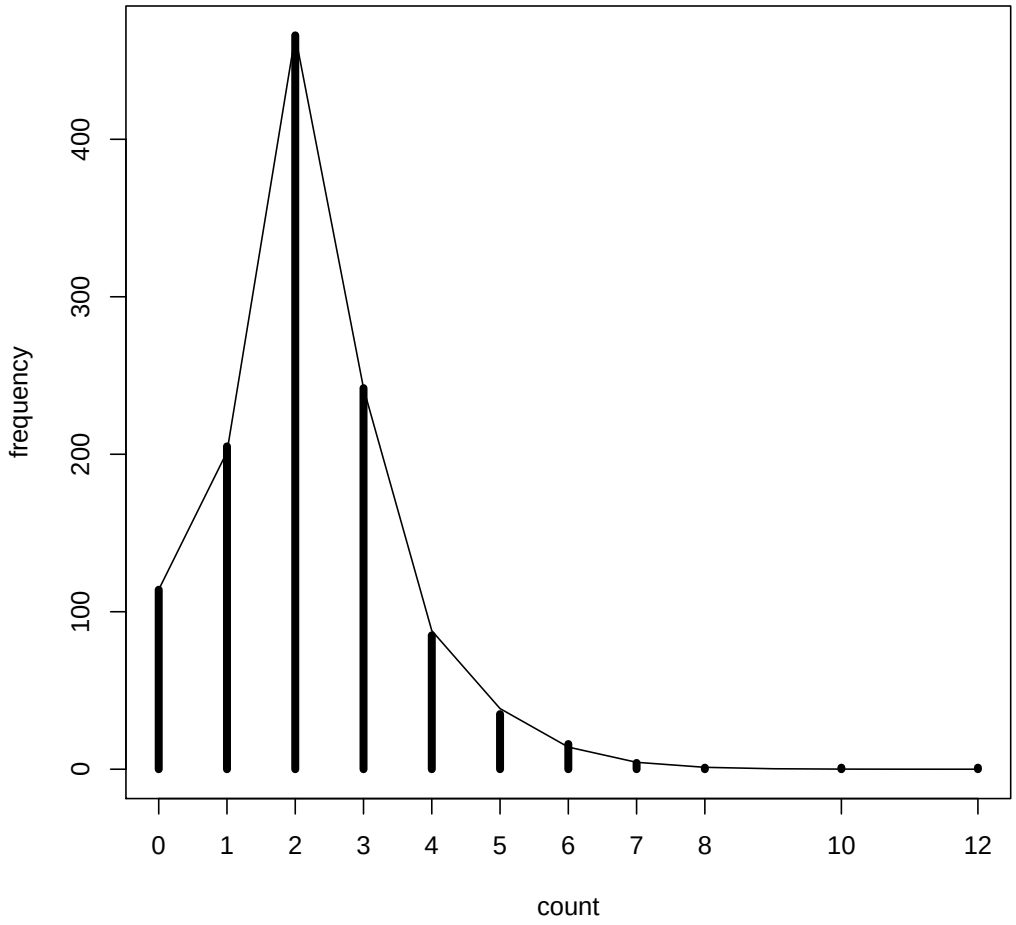

Figure 4.2: The graph of the observed frequencies compared to the estimated frequencies for the Zero-Two-Three Inflated Poisson.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This work deals with a general inflated Poisson distribution (GIP) which appears to be a very natural generalization of the regular Poisson distribution. This model can be effective in modeling a dataset where it seems plausible that certain count values may have higher probabilities due to natural reasons. We have used the GIP to model the fertility data of Swedish women, and found that the ZTTIP model appears to these data quite well. Because of the extra parameter(s), the GIP seems to be much more flexible in model fitting than the regular Poisson. Our simulation study indicates that MLEs are overall better than the corrected MMEs in estimating the model parameters. In performing the simulation, we note that for certain ranges of the inflated probabilities in all GIP models, the computation algorithm for calculating MLEs does not converge. Nonetheless, we selected all permissible values and compared the overall performance of the MLEs and CMMEs for three special cases of GIP. In the future, we would like to continue working on finding a computation algorithm for calculating MLEs that would converge. Towards this end we would further investigate the widely used parametric and non-parametric bootstrap algorithms.

REFERENCES

[1] D. R. Anderson, D. J. Sweeney, and T. A. Williams, *Essentials of modern buisness statistics with microsoft excel*, 5th ed., Cengage Learning, Mason, 2012.

[2] D. Doane and L. Seward, *Applied statistics in business and economics*, 3rd ed., McGraw-Hill, New York, 2011.

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

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

[5] S Jaggia and A Kelly, *Business statistics - communicating with numbers*, 1st ed., McGraw-Hill Irvin, New York, 2012.

[6] D. M. Levine, D. F. Stephan, T. C. Krehbiel, and M. L. Berenson, *Statistics for managers using microsoft excel*, 6th ed., Pearson, Boston, 2011.

[7] 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).

[8] M. K. Pelosi and T. M. Sandifer, *Elementary statistics: from discovery to decision*, 1st ed., John Wiley & Sons, Inc, New York, 2003.

[9] L. von Bortkiewicz, *Das gesetz der kleinen zahlen*, B.G. Teubner, Leipzig, 1898.

[10] R. M. Weiers, *Introduction to business statistics*, 6th ed., Cengage Learning, Mason, 2008.

APPENDIX A
LETTER FROM INSTITUTIONAL RESEARCH BOARD

MARSHALL UNIVERSITY
www.marshall.edu

Office of Research Integrity

February 10, 2014

Patrick Stewart
Holderby Hall Room 321
Huntington, WV 25755

Dear Mr. Stewart:

This letter is in response to the submitted thesis abstract entitled *"A Generalized Inflated Poisson Distribution with Application to Modeling Fertility Data."* 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 involves publicly available data and 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

APPENDIX B
Algebraic Solutions for the Method of Moments Estimators

Using Mathematica, the general solution to equation (2.6) is as follows:

$$\hat{\pi}_1 = \frac{1}{4(2m_1^{'}-3m_2^{'}+m_3^{'})}(-24(m_1^{'})^2 - 21(m_2^{'})^2 + m_1^{'}(8 + 50m_2^{'} - 14m_3^{'} - 6\sqrt{20(m_1^{'})^2 - 44m_1^{'}m_2^{'} + 21(m_2^{'})^2 + 12m_1^{'}m_3^{'} - 10m_2^{'}m_3^{'} + (m_3^{'})^2}) - m_3^{'}(-4 + m_3^{'} + \sqrt{20(m_1^{'})^2 - 44m_1^{'}m_2^{'} + 21(m_2^{'})^2 + 12m_1^{'}m_3^{'} - 10m_2^{'}m_3^{'} + (m_3^{'})^2}) + m_2^{'}(-12 + 10m_3^{'} + 5\sqrt{20(m_1^{'})^2 - 44m_1^{'}m_2^{'} + 21(m_2^{'})^2 + 12m_1^{'}m_3^{'} - 10m_2^{'}m_3^{'} + (m_3^{'})^2}))$$

$$\hat{\pi}_2 = \frac{1}{4}(-4m_1^{'} + 5m_2^{'} - m_3^{'} - \sqrt{20(m_1^{'})^2 - 44m_1^{'}m_2^{'} + 21(m_2^{'})^2 + 12m_1^{'}m_3^{'} - 10m_2^{'}m_3^{'} + (m_3^{'})^2})$$

$$\hat{\lambda} = \frac{-2m_1^{'}+3m_2^{'}-m_3^{'}+\sqrt{20(m_1^{'})^2-44m_1^{'}m_2^{'}+21(m_2^{'})^2+12m_1^{'}m_3^{'}-10m_2^{'}m_3^{'}+(m_3^{'})^2}}{4m_1^{'}-2m_2^{'}}$$

Likewise, the general solution to equation (2.7) is as follows:

$$\hat{\pi}_1 = (-376(m_1^{'})^4 + 514(m_2^{'})^4 + 432(m_3^{'})^3 + 50(m_3^{'})^4 - 216(m_3^{'})^2m_4^{'} - 14(m_3^{'})^3m_4^{'} + 36m_3^{'}(m_4^{'})^2 + (m_3^{'})^2(m_4^{'})^2 - 2(m_4^{'})^3 + 4(m_1^{'})^3(108 + 505m_2^{'} - 302m_3^{'} + 65m_4^{'}) - (m_2^{'})^3(2662 + 870m_3^{'} + 111m_4^{'}) - m_2^{'}(286(m_3^{'})^3 - 4(m_3^{'})^2(-594 + m_4^{'}) - 12m_3^{'}m_4^{'}(66 + m_4^{'}) + (m_4^{'})^2(66 + m_4^{'})) + (m_2^{'})^2(679(m_3^{'})^2 - 6m_4^{'}(121+4m_4^{'})+6m_3^{'}(726+17m_4^{'}))-2(m_1^{'})^2(1729(m_2^{'})^2+606(m_3^{'})^2-24m_3^{'}(27+11m_4^{'})+m_4^{'}(108+25m_4^{'})+m_2^{'}(1188-2112m_3^{'}+496m_4^{'}))+m_1^{'}(1587(m_2^{'})^3-332(m_3^{'})^3+3(m_4^{'})^2(12+m_4^{'})+36(m_3^{'})^2(36+7m_4^{'})-2m_3^{'}m_4^{'}(216+25m_4^{'})+(m_2^{'})^2(4356-3206m_3^{'}+1005m_4^{'})+m_2^{'}(1872(m_3^{'})^2+3m_4^{'}(264+35m_4^{'})-4m_3^{'}(1188 + 257m_4^{'}))))/(2(6m_1^{'} - 11m_2^{'} + 6m_3^{'} - m_4^{'})^3)$$

$$\hat{\pi}_2 = (40(m_1^{'})^3 + 5(m_2^{'})^3 + (m_3^{'})^2(-7m_3^{'} + m_4^{'}) - 4(m_1^{'})^2(37m_2^{'} - 21m_3^{'} + 5m_4^{'}) - (m_2^{'})^2(15m_3^{'} + 13m_4^{'})+m_2^{'}(20(m_3^{'})^2+6m_3^{'}m_4^{'}-(m_4^{'})^2)+2m_1^{'}(67(m_2^{'})^2-74m_2^{'}m_3^{'}+18(m_3^{'})^2+22m_2^{'}m_4^{'}-10m_3^{'}m_4^{'}+(m_4^{'})^2))/(-6m_1^{'} + 11m_2^{'} - 6m_3^{'} + m_4^{'})^2$$

$$\hat{\pi}_3 = \frac{-2(m_1^{'})^2-2(m_2^{'})^2+(m_3^{'})^2-m_2^{'}m_4^{'}+m_1^{'}(5m_2^{'}-2m_3^{'}+m_4^{'})}{2(6m_1^{'}-11m_2^{'}+6m_3^{'}-m_4^{'})}$$

$$\hat{\lambda} = \frac{-6m_1^{'}+11m_2^{'}-6m_3^{'}+m_4^{'}}{2m_1^{'}-3m_2^{'}+m_3^{'}}$$

Patrick Stewart

Marshall University
Department of Mathematics
Smith Hall
Huntington, WV 25755

Phone: (304) 893-7860
Email:stewart152@marshall.edu

Education

- Marshall University: Expected Graduation, May 2014
  Master of Arts in Mathematics with an emphasis in Statistics
  GPA (4.0 scale): (44 credit hours)

- Marshall University: August 2011 to May 2012
  Master of Arts in Teaching (Changed programs before receiving degree)
  GPA (4.0 Scale): 4.0 (15 credit hours)

- Marshall University: Graduated May 2011
  Bachelor of Science in Computer Science
  Minor in Mathematics
  GPA(4.0 scale): 3.97 (Graduated Summa Cum Laude)
  Capstone: *Calculating Cardiovascular Risk Factors Based on the Carotid Intima-Media Thickness*

- Concord University: January 2006 to June 2006
  Taken while in high school
  GPA(4.0 scale): 4.0

Graduate-Level Math Classes Taken

- Modern Algebra I and II
- Probability and Statistics I and II
- Number Theory
- Time Series Forecasting
- Time Scale Calculus
- Advanced Calculus I and II
- Numerical Analysis
- Game Theory
- Advanced Mathematical Statistics
- Multivariate Statistics
- Biostatistics

Programming Languages
I have knowledge in the following programming languages

- R
- Python
- Java
- C++
- Mathematica
- SAS

Teaching Experience

- Marshall University
  Teaching Assistant 2012-Present
  Primary instructor for College Algebra-Expanded (5 credit hour class) for 2 semesters and Mathematics Skills II (3 credit hour class) for 2 semesters and a mathematics tutor for 4 semesters.

Awards and Distinctions

- Pi Mu Epsilon, Honour Society
- Phi Kappa Phi, Honour Society
- Golden Key, Honour Society
- John Marshall Scholarship Recipient
- West Virginia Engineering Science and Technology Scholarship Recipient
- West Virginia Promise Scholar
- AP Scholar
- Marshall University Dean's list (All Semesters from Fall 2007 to Spring 2011)

Conferences Attended

- iPED: Inquiring Pedagogies Teaching Conference, August 2012