Winter 2011

The Doubly Inflated Poisson and Related Regression Models

Manasi Sheth-Chandra
Old Dominion University

THE DOUBLY INFLATED POISSON AND RELATED REGRESSION MODELS

by

Manasi Sheth-Chandra
B.S. Mathematics and Biochemistry, May 2002, Loyola University, Chicago, IL
M.S. Mathematics, December 2003, Loyola University, Chicago, IL
M.S. Statistics, May 2010, Old Dominion University, Norfolk, VA

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
December 2011

Approved by:

N. Rao Chaganty (Director)

Norou Diawara (Member)

Larry Filer (Member)

Dayanand Naik (Member)

ABSTRACT

THE DOUBLY INFLATED POISSON AND RELATED REGRESSION MODELS

Manasi Sheth-Chandra
Old Dominion University, 2011
Director: Dr. N. Rao Chaganty

Most real life count data consists of some values that are more frequent than allowed by the common parametric families of distributions. For data consisting of only excess zeros, in a seminal paper Lambert (1992) introduced Zero-Inflated Poisson (ZIP) model, which is a mixture model that accounts for the inflated zeros. In this thesis, two Doubly Inflated Poisson (DIP) probability models, DIP (p, λ) and DIP (p_1, p_2, λ), are discussed for situations where there is another inflated value $k > 0$ besides the inflated zeros. The distributional properties such as identifiability, moments, and conditional probabilities are also discussed for both probability models. For the data consisting of raw counts as well as grouped frequencies, we have considered parameter estimation using maximum likelihood (ML) and method of moments techniques. Efficiencies show that the ML estimators perform far better than the moment estimators. An application to DIP models is illustrated using data on patients' length of stay in a hospital. Parameter estimation of DIP regression models using maximum likelihood approach is also discussed using data on dental cavities. Finally, we conclude with a brief introduction to two Doubly Inflated Negative Binomial (DINB) distributions and their related regression models.

DEDICATION

My parents deserve special mention for their relentless faith and support. It is them who put the fundamentals of my learning character, showing me the joy of intellectual pursuit ever since I was a child, and for raising me with care and love.

Words fail to express my appreciation to my husband Ramesh who has shown dedication, love and persistent confidence in me. Furthermore, to his family, with their thoughtful support, I thank you.

To them, I dedicate this thesis.

ACKNOWLEDGEMENTS

First and foremost I offer my sincerest gratitude to my mentor and advisor, Dr. N. Rao Chaganty, who has supported me throughout my thesis with his patience and knowledge while allowing me the room to learn and develop professional and personal skills. He provided encouragement, sound advice, excellent teaching, and good company. With his enthusiasm, inspiration, and his relentless efforts to explain things clearly and simply, he has made my graduate education one of the best life experiences.

I would also like to express my sincere and deep gratitude to my graduate professors Dr. Dayanand Naik, Dr. Norou Diawara, Dr. Nak-Kyeong Kim, and Dr. Richard Noren. Many thanks goes to many people who have taught me mathematics and statistics, especially my high school math teacher Mr. Shaw and my undergraduate professor and advisor Dr. Anne McDonald. I gratefully thank my colleagues and now friends from UW-Whitewater for their moral support. I would also like to thank Dr. Dayanand Naik, Dr. Norou Diawara, and Dr. Larry Filer for serving on my dissertation committee in midst of all their respective activities. Additionally, I would also like to thank Dr. Dorrepaal, and the Department of Mathematics and Statistics for providing me with kind financial assistance and a pleasant learning environment. Collective and individual acknowledgements are also owed to my many student colleagues for providing a stimulating and memorable environment in which to learn and grow.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

I.1 MOTIVATION

Usually, count data is analyzed using a standard framework of Poisson regression models. However, in practice, count data often show more variability compared to what one would expect if the response distribution truly was Poisson. That is, the observed sample variances are much larger than the sample means (i.e. over-dispersion), whereas for the Poisson family the variance is identical to the mean. The general problem of how to account for over-dispersion has been studied extensively in the statistical literature. An excellent summary of approaches to verify and analyze over-dispersed count data can be found in Cameron and Trivedi (1998). Usually over-dispersed count data consists of certain values occurring more frequently than allowed by the common parametric families of distributions. Lambert (1992) introduced the Zero-Inflated Poisson (ZIP) model in which the poisson mean λ_i is parameterized as a log link function of observable vector of covariates $\boldsymbol{B}_i$ and the mixture probability p_i is parameterized as a logistic function of the observable vector of covariates $\boldsymbol{z}_i$ for $i = 1, \ldots, n$. That is, for n independent responses $y_1, \ldots, y_n$,

$$\begin{cases} y_i = 0, & \text{with probability } p_i \\ y_i \sim \text{Poisson}(\lambda_i), & \text{with probability } 1 - p_i, \end{cases} \tag{1}$$

where

$$\begin{aligned} \log(\lambda_i) &= \boldsymbol{B}_i\boldsymbol{\beta} \\ \text{and} \qquad p_i &= \frac{\exp(\boldsymbol{z}_i\boldsymbol{\gamma})}{1 + \exp(\boldsymbol{z}_i\boldsymbol{\gamma})}. \end{aligned} \tag{2}$$

Here, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are regression parameters. This model (1) accommodates the count data consisting of excess zeros, and has received much attention in the literature recently. However, there may be real-life instances (see Section I.2) where observations may include higher incidences of count zero as well as another count value, say

$k > 0$. In such cases, one can use doubly inflated probability models. The doubly inflated models allow researchers to analyze whether or not the behavior of two inflation occurs, and if the behavior does occur, what influences the frequency of occurrence of the two peaks. Additionally, doubly inflated models have a statistical advantage to standard Poisson and Zero-Inflated Poisson models in that they model the preponderance of zeros and k counts as well as the distribution of positive counts simultaneously. In fact, ordinary poisson and zero-inflated poisson probability models are special cases of Doubly Inflated Poisson probability models. Unfortunately, there is not a specific frequency or proportions of zero and k counts or ratio of zero to nonzero counts and vice versa that can be used to determine if a particular distribution is doubly inflated or not. First, let's consider some real-life examples where double inflation of counts may occur.

I.2 EXAMPLES

I.2.1 SCHOOL DISCIPLINARY ACTIONS

In a growing number of schools, student misbehavior exacts a heavy toll on their academic achievement and interferes with the education of their classmates. In one intermediate/high school with 880 students, there were more than 5,100 office referrals in a single academic year, and two-thirds of the students had at least one referral. The cost associated with the disciplinary process, assuming each referral takes approximately 10 minutes to complete and submit, translates into 51,000 administrative minutes or 850 hours or 140 school days. Due to the "Zero-tolerance" policies, schools have become increasingly willing to suspend or expel students even for one minor offense in certain misbehavioral categories. Considering the frequency of disciplinary actions per student, we would observe an inflated number of zero's for the well-behaved students and an inflated numbers of 1's for those receiving immediate suspension or expulsion. Various factors such as school level, socio-economic status, gender, ethnicity, location, and student type (i.e., special education or not) can also play a significant role in determining students' misbehavior. In such cases, one can apply as an alternative to Poisson regression the Doubly Inflated Poisson (DIP) regression model.

I.2.2 MORTGAGE PAYMENTS

Mortgage lending is the primary mechanism used in many countries to finance private ownership of residential and commercial property. Mortgage loans are generally structured as long-term loans with periodic payments that are similar to an annuity and calculated according to the time value of money formulae. The most basic arrangement would require a fixed monthly payment over a period of ten to thirty years. Mortgage payments are usually scheduled on the first day of each month with 14 day grace period. It is a common knowledge to link one's mortgage payments to one's pay day, which is either monthly or bi-weekly. Also, sometimes, paying half your mortgage every two weeks instead of a full payment once a month can be done with any type of loan but is most common with a 30-year fixed-rate loan. Thus, one would notice high numbers of home-owners making their mortgage payments either on the 1st of the month or the 15th. This would result in inflated counts for 0's for those making payments on time and inflated counts of 14's for those making payments on the last day of the grace period. Having inflated frequencies for these two counts one can apply the estimation methods using Doubly Inflated Poisson (DIP) models.

I.2.3 PATIENTS' LENGTH OF STAY

The length of stay (LOS) patterns of patients admitted to hospitals has been a longstanding concern of clinicians and administrators. It is often used as an indicator of how efficient a hospital is at any location, and also used for health planning purposes. Inpatient length of stay days are calculated by subtracting day of admission from day of discharge. Due to exponential growth and development of clinical and pharmaceutical practices, health services provided for patients have led to improved outcomes for patients. This leads to a decline in LOS and an increase in prevalence of same-day separations. For the patients with longer LOS, there is current belief that the type of reimbursement system or health insurance plan now plays a more significant role in the patient LOS in hospitals. Hence, there might be another increase in prevalence in LOS for patients with acute medical care. We will be discussing this application in depth using a sample data set in Chapters II and III.

I.2.4 DENTAL EPIDEMIOLOGY

In dental epidemiology, the DMFT-index is an important indicator and measure for the dental status of a person. It is a count number of DECAYED, MISSING, and FILLED Teeth. Consider an example data presented in Table 1 on 1013 school children of age 7. Only the eight deciduous molars are considered. Thus, the smallest possible value of the DMFT-index is 0 and the largest is 8. DMFT was measured at the beginning of the study and after a year. Thus, we obtain a count for the change in DMFT, δDMFT= DMFT1 − DMFT2. The zero count corresponds to those children showing no improvement and/or consistent dental care. The one count corresponds to those children showing improvement in one cavity. In this data, inflated counts of 0's and 1's are observed. The aim of the study is to compare methods of six schools of treatments: oral health education, enrichment of the school diet with rice bran, mouthwash with 0.2% of NaF solution, oral hygiene, all of the four treatments, and control. Other covariates such as gender and ethnicity groups (White, Black, Others including predominantly Hispanic) were also considered. See Table 1 for a subset of the observations. In Chapter IV, Doubly Inflated Poisson regression models will be applied to this data.

Table 1: Dental Epidemiology Data

ID	Treatment	Gender	Ethnicity	δDMFT
1	1	m	w	0
2	2	f	w	1
3	4	f	b	2
4	5	f	o	0
⋮	⋮	⋮	⋮	⋮
1012	3	m	b	3
1013	2	m	o	0

NOTE: The treatment groups are: (1) for oral health education, (2) for rice bran diet, (3) for mouthwash rinse, (4) for oral hygiene, (5) for all above, and (6) for control. Gender is abbreviated as (m) for males and (f) for females. Ethnicity groups are: (w) for white, (b) for black, and (o) for others who are predominantly hispanic.

I.3 ORGANIZATION OF THESIS

The rest of this dissertation is organized as follows. In the next two chapters, Chapters II and III, we introduce two Doubly Inflated Poisson (DIP) models, DIP (p, λ) and DIP (p_1, p_2, λ), along with a motivating example of hospital's length of stay data. For both probability models, we discuss the distributional properties such as identifiability, moments, conditional properties, and Fisher information. We also discuss the two parameter estimation techniques, maximum likelihood (ML) and method of moments (MOM), for data consisting of raw counts as well as grouped frequencies of counts. Asymptotic as well as small sample comparisons show that ML estimators are more efficient than the moment estimators.

In Chapter IV, regression models that mix counts of zeros, counts of k's, and Poisson counts are described in detail using logistic and log-linear link functions. Maximum likelihood estimates (MLE)'s for parameters are discussed in length for data consisting of raw counts as well as grouped frequencies. Applications to DIP regression models are illustrated using sample data on dental cavities. In Chapter V, we briefly discuss construction of Doubly Inflated Negative Binomial distributions as well as their related regression models.

In Chapter VI, we summarize the results obtained in this dissertation. Finally, the Appendix contains a subset of SAS programs that were used to obtain the results presented in this dissertation.

CHAPTER II

THE FIRST DOUBLY INFLATED POISSON MODEL

Count data with many zeros are common in a wide variety of disciplines including econometrics, medicines, road safety, and manufacturing. A fair amount of statistical methodology has been developed to deal with such data. One of the most popular statistical approaches is due to Lambert (1992) who introduced zero-inflated Poisson (ZIP) regression models to accommodate for inflated number of zeros. However, there are real life instances as discussed in Section I.2 where data consists of inflated number of zeros as well as inflated number of another count $k > 0$. In this chapter, we introduce a Doubly Inflated Poisson model to deal with count responses with two inflated values.

In Section II.1, we first begin with a motivating example on hospital stay data to illustrate a need for applied and theoretical development of the doubly inflated models. In Section II.2, we describe in detail the development of Doubly Inflated Poisson model with parameters p and λ using a latent variable. We also discuss the distributional properties of the probability model including identifiability of the parameters, moments, expectation, variance, and Fisher information. Section II.3 describes estimation techniques for data that consists of raw counts using maximum likelihood approach and method of moments. Section II.4 describes parameter estimation methods such as maximum likelihood estimation and moment estimation for data consisting of grouped frequencies of count responses. We also compare the two estimation approaches using asymptotic relative efficiencies as well as small sample efficiency calculations. In Section II.5, we illustrate the use of DIP (p, λ) probability model on the hospital stay data and compare the results to those using ZIP model.

II.1 MOTIVATING EXAMPLE

To motivate both applied and theoretical development of the doubly inflated models, the patients' length of stay data discussed below will be used. The patients' length of stay (LOS) is usually used to determine the efficiency of a hospital since it is considered to be a surrogate measure of costs and the quality of health-care provided to the patients. The example data consists of observations from a group of 261 patients and the study is done to analyze their length of stay. The observed frequencies for LOS, displayed in Table 2), shows that there are inflated frequencies for LOS $= 0$ as well as for LOS $= 3$. We observe inflated number of zeros (i.e. LOS $= 0$) for patients receiving outpatient care, i.e. they receive treatment/care without having to stay at the hospital. We also observe inflated counts of 3 as a consequence of many patients staying three nights at the hospital to receive short-term inpatient care. Generally speaking, due to insurance coverage, necessary acute treatment, type of treatment(s), and improvements in medical fields, most inpatient care requires patients to stay 3 days in order for them to get necessary treatment and/or recovery. Lots of public health literature have been published to indicate that a 3-day stay is quite common for most acute inpatient care. Hence, this data is indicative of the situation-driven application for doubly.inflated count models. In the next section, we consider how to construct the first Doubly Inflated Poisson Model with parameters p and λ.

Table 2: Observed Data for Patient's Length of Stay

LOS	Frequency
0	55
1	35
2	35
3	75
4	40
5	20
6	13
7	8
8	4
9	5
10	3
11	1
12	4
13	0
14	1

II.2 DOUBLY INFLATED POISSON (P, λ) MODEL

When large frequencies for count zero and count k are observed, a simple way to model the count behavior is by introducing a latent random variable Z which is distributed as Binomial (2, p), $0 \leq p \leq 1$. The probability mass function for Z is

$$P(Z = z) = \begin{cases} p^2, & \text{if } z = 2; \\ 2pq, & \text{if } z = 1; \\ q^2, & \text{if } z = 0. \end{cases}$$

Let Y be a random variable with the following characteristic: Y given $Z = 2$ is degenerate at 0; Y given $Z = 1$ is degenerate at k; and Y given $Z = 0$ is a Poisson distribution with parameter λ, where $\lambda > 0$. That is, the conditional probability for Y given Z is

$$P(Y = y|Z = z) = \begin{cases} 1, & \text{for } z = 2, y = 0; \\ 1, & \text{for } z = 1, y = k; \\ \dfrac{\exp(-\lambda)\lambda^y}{y!}, & \text{for } z = 0, y = 0, 1, 2, \ldots \end{cases} \tag{3}$$

Then, the joint distribution of Y and Z is given by

$$P(Z = z, Y = y) = \begin{cases} p^2, & \text{for } z = 2, y = 0; \\ 2pq, & \text{for } z = 1, y = k; \\ q^2\left(\dfrac{\exp(-\lambda)\lambda^y}{y!}\right), & \text{for } z = 0, y = 0, 1, 2, \ldots \end{cases}$$

The marginal probability mass function of Y is therefore:

$$P(Y = y) = \begin{cases} p^2 + q^2 \exp(-\lambda), & \text{for } y = 0; \\ 2pq + q^2\left(\dfrac{\exp(-\lambda)\lambda^k}{k!}\right), & \text{for } y = k; \\ q^2\left(\dfrac{\exp(-\lambda)\lambda^y}{y!}\right), & \text{for } y = 1, 2, \ldots \neq k. \end{cases} \tag{4}$$

Thus, an additional proportion of p^2 explains the inflated zero counts and an additional proportion of $2pq$ explains the inflated k counts. We call this distribution (4) as *Doubly Inflated Poisson* (p, λ), abbreviated as DIP (p, λ). This distribution can also be interpreted as a mixture of Poisson distribution and degenerate distributions with mass concentrated at count zero and at count k. It is also interesting to note that as $p \to 0$, this model reduces to an ordinary Poisson distribution with mean parameter λ. We generated the probability mass functions for a known k using arbitrary values of the parameters p and λ. The plots of these mass functions are as shown in Figure 1. Each of the graphs indicate presence of two peaks one at count 0 and the other at count k. The conditional distribution of Z given Y is as shown in Table 3.

Table 3: $P(Z = z|Y = y)$ of DIP (p, λ)

	Y		
Z	0	k	$1, 2, ... \neq k$
0	$\dfrac{q^2 \exp(-\lambda)}{p^2 + q^2 \exp(-\lambda)}$	$\dfrac{q^2 \exp(-\lambda)\lambda^k}{2pq(k!) + q^2 \exp(-\lambda)\lambda^k}$	1
1	0	$\dfrac{2pq(k!)}{2pq(k!) + q^2 \exp(-\lambda)\lambda^k}$	0
2	$\dfrac{p^2}{p^2 + q^2 \exp(-\lambda)}$	0	0

NOTE: The sum of the entries in any column is 1.

II.2.1 IDENTIFIABILITY

The DIP(p, λ) model is identifiable. That is, each pair of distinct values of the parameters p and λ lead to a unique probability mass function. To see that this is true, suppose (p_1, λ_1) and (p_2, λ_2) are such that the DIP probability models corresponding to these values are the same. Therefore, $P_{(p_1,\lambda_1)}[Y = y] = P_{(p_2,\lambda_2)}[Y = y]$ for all

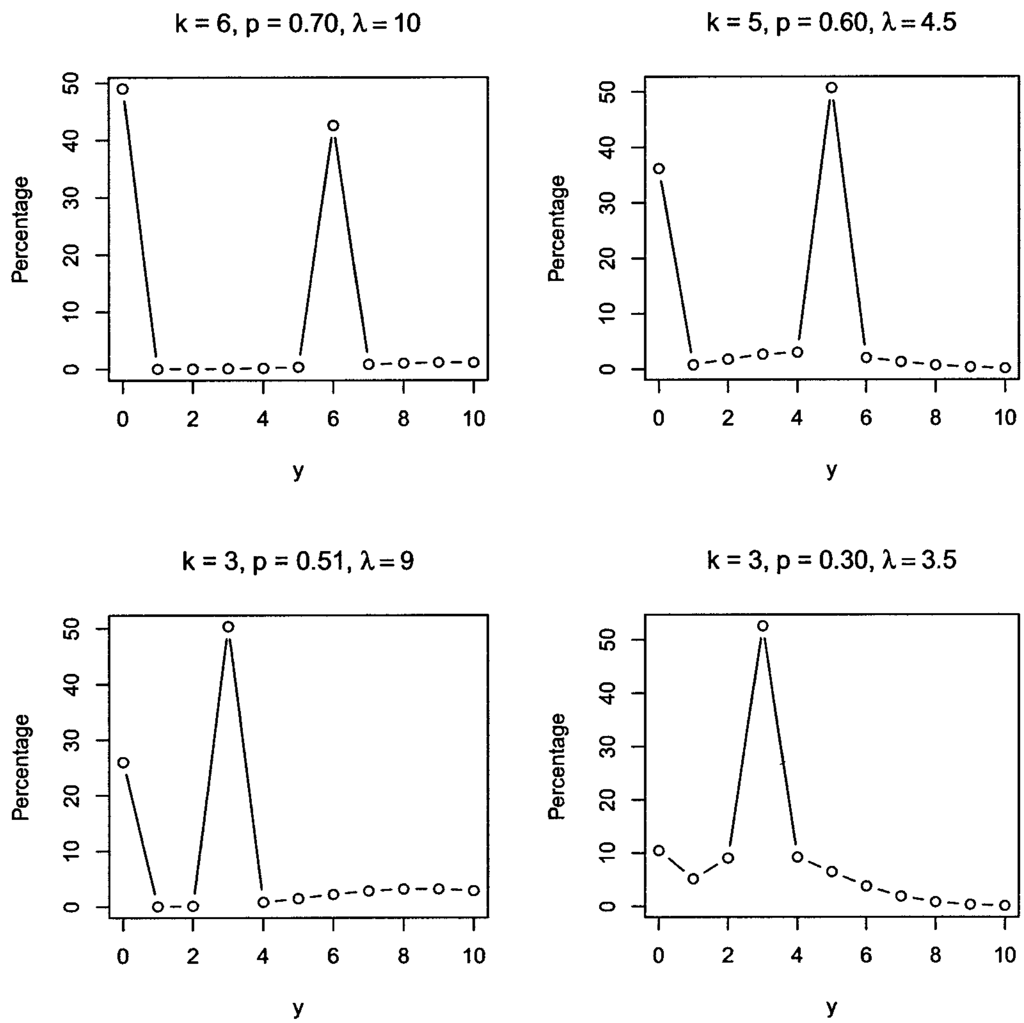

Figure 1: A Few DIP (p, λ) Distributions for known k

values of y. Then, for $y \neq k$,

$$q_1^2 \frac{\exp(-\lambda_1)\lambda_1^y}{y!} = q_2^2 \frac{\exp(-\lambda_2)\lambda_2^y}{y!}$$

or equivalently

$$\left(\frac{\lambda_1}{\lambda_2}\right)^y = \frac{q_2^2 \exp(-\lambda_2)}{q_1^2 \exp(-\lambda_1)}.$$

Since the left hand side depends on y and the right hand side does not depend on y, the equality holds for all $y \neq k$ if and only if $\lambda_1/\lambda_2 = 1$ or $\lambda_1 = \lambda_2$ and hence $p_1 = p_2$ also. Thus, $P_{(p_1,\lambda_1)}(Y = y) = P_{(p_2,\lambda_2)}(Y = y)$ for all y, implies $(p_1, \lambda_1) = (p_2, \lambda_2)$. Hence, DIP$(p,\ \lambda)$ model is identifiable.

II.2.2 MOMENTS, EXPECTATION, AND VARIANCE

Using the conditional distribution of Y given Z as given in equation (3), we can check that the conditional expectation of Y given Z is

$$E(Y|Z = z) = \begin{cases} 0, & \text{if } z = 2; \\ k, & \text{if } z = 1; \\ \lambda, & \text{if } z = 0 \end{cases}$$

and

$$E(Y^2|Z = z) = \begin{cases} 0, & \text{if } z = 2; \\ k^2, & \text{if } z = 1; \\ \lambda^2 + \lambda, & \text{if } z = 0. \end{cases}$$

Further if $E(Y^r) = \mu_r$ is the marginal rth moment of Y, we can easily verify that the first four moments are

$$\begin{aligned} \mu_1 &= E(Y) = 2pqk + q^2\lambda, \\ \mu_2 &= E(Y^2) = 2pqk^2 + q^2(\lambda^2 + \lambda), \\ \mu_3 &= E(Y^3) = 2pqk^3 + q^2(\lambda + 3\lambda^2 + \lambda^3), \text{ and} \\ \mu_4 &= E(Y^4) = 2pqk^4 + q^2(\lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4). \end{aligned} \tag{5}$$

Thus, the expected value $\mu = E(Y)$ and variance $\sigma^2 = Var(Y)$ are

$$\begin{aligned} \mu &= 2pqk + q^2\lambda, \\ \text{and} \quad \sigma^2 &= 2pqk^2 + q^2(\lambda^2 + \lambda) - \mu^2 \\ &= \mu - \mu^2 + q^2\lambda^2 - 2p\,q\,k\,(k-1). \end{aligned} \tag{6}$$

Unlike the Poisson family of distributions where the mean and variance are equal, the Doubly Inflated Poisson $(p,\, \lambda)$ can model data where mean and variance have different values.

II.2.3 FISHER INFORMATION

Let Y be a random variable with probability mass function $p(y) = P(Y = y)$ given in equation (4). The Fisher information matrix for the distribution is given by

$$\boldsymbol{I} = \begin{pmatrix} I(p) & I(p,\lambda) \\ I(p,\lambda) & I(\lambda) \end{pmatrix}, \tag{7}$$

where the elements of the matrix are

$$\begin{aligned} I(p) &= -E\left(\frac{\partial^2 \log(p(Y))}{\partial p^2}\right) \\ &= \frac{-2\Big(1+\exp(-\lambda)\Big)\Big(p^2+q^2\exp(-\lambda)\Big) - \Big(2p - 2q\exp(-\lambda)\Big)^2}{\Big(p^2+q^2\exp(-\lambda)\Big)} \\ &\quad - \left[\frac{\Big(2pq + q^2\frac{\exp(-\lambda)\lambda^k}{k!}\Big)\Big(-4 + 2\frac{\exp(-\lambda)\lambda^k}{k!}\Big) - \Big(2 - 4p - 2q\frac{\exp(-\lambda)\lambda^k}{k!}\Big)^2}{\Big(2pq + q^2\frac{\exp(-\lambda)\lambda^k}{k!}\Big)}\right] \\ &\quad + 2\left(1 - \exp(-\lambda) - \frac{\exp(-\lambda)\lambda^k}{k!}\right), \end{aligned} \tag{8a}$$

$$\begin{aligned} I(p,\lambda) &= -E\left(\frac{\partial^2 \log(p(Y))}{\partial\lambda\partial p}\right) \\ &= \frac{-2pq\exp(-\lambda)}{\Big(p^2+q^2\exp(-\lambda)\Big)} - \frac{2q^2\exp(-\lambda)\lambda^k}{\Big(2pq(k!) + q^2\exp(-\lambda)\lambda^k\Big)}\left(1 - \frac{k}{\lambda}\right), \end{aligned} \tag{8b}$$

and

$$
\begin{aligned}
I(\lambda) &= -E\left(\frac{\partial^2 \log(p(Y))}{\partial\lambda^2}\right) \\
&= \frac{-p^2q^2\exp(-\lambda)}{\left(p^2 + q^2\exp(-\lambda)\right)} + \frac{q^2}{\lambda}\left(1 - \frac{\exp(-\lambda)\lambda^{k-1}}{(k-1)!}\right) \\
&+ \frac{q^2\exp(-\lambda)\lambda^k}{\left(2pq(k!) + q^2\exp(-\lambda)\lambda^k\right)} \times \left(\left(\frac{k}{\lambda} - 1\right)^2 q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right) \\
&- q^2\frac{\exp(-\lambda)\lambda^k}{k!}\left(\left(\frac{k}{\lambda} - 1\right)^2 - \frac{k}{\lambda^2}\right).
\end{aligned}
\tag{8c}
$$

The Cramér-Rao lower bounds for the variances of unbiased estimators of p and λ can be obtained from the diagonal elements of I^{-1}.

II.3 METHODS OF ESTIMATION FOR RAW COUNTS

In this section, we will explore parameter estimation methods including maximum likelihood estimation and moment estimation for data that consists of independent doubly inflated raw counts.

II.3.1 MAXIMUM LIKELIHOOD ESTIMATION

Suppose our data consists of independent count responses $y_i, i = 1, \ldots, n$, distributed as DIP(p, λ) for a known k. With $q = 1 - p$, the likelihood function can be written as

$$
\begin{aligned}
L(p, \lambda|\boldsymbol{y}) = \prod_{\{i:y_i=0\}} \left(p^2 + q^2\exp(-\lambda)\lambda^{y_i}\right) \prod_{\{i:y_i=k\}} \left(2pq + q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right) \\
\prod_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} \left(q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right).
\end{aligned}
$$

And the log-likelihood function is

$$\ell(p,\lambda|\boldsymbol{y}) = \sum_{\{i:y_i=0\}} \log\left(p^2 + q^2\exp(-\lambda)\lambda^{y_i}\right) + \sum_{\{i:y_i=k\}} \log\left(2pq + q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)$$
$$+ \sum_{\{i: {y_i=y \atop y\neq 0,k}\}} \log\left(q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right). \tag{9}$$

We differentiate the log-likelihood function (9) with respect to the parameters p and λ to obtain the following score functions:

$$\frac{\partial\ell(p,\lambda)}{\partial p} = \sum_{\{i:y_i=0\}} \frac{2p - 2q\exp(-\lambda)\lambda^{y_i}}{p^2 + q^2\exp(-\lambda)\lambda^{y_i}} + \sum_{\{i:y_i=k\}} \frac{(2-4p)(k!) - 2q\exp(-\lambda)\lambda^{y_i}}{2pq(y_i!) + q^2\exp(-\lambda)\lambda^{y_i}}$$
$$- \frac{2}{q}\sum_{\{i: {y_i=y \atop y\neq 0,k}\}} 1,$$

and

$$\frac{\partial\ell(p,\lambda)}{\partial \lambda} = \sum_{\{i:y_i=0\}} \frac{-q^2\exp(-\lambda)\lambda^{y_i}}{p^2 + q^2\exp(-\lambda)\lambda^{y_i}} + \sum_{\{i:y_i=k\}} \frac{q^2\exp(-\lambda)\lambda^{y_i}}{2pq(y_i!) + q^2\exp(-\lambda)\lambda^{y_i}}\left(\frac{y_i}{\lambda} - 1\right)$$
$$+ \sum_{\{i: {y_i=y \atop y\neq 0,k}\}} \left(\frac{y_i}{\lambda} - 1\right). \tag{10}$$

To find the maximum likelihood estimates $(\widehat{p}, \widehat{\lambda})$, we solve for the roots of the above score functions. The parameter constraints are $0 \leq p \leq 1$ and $\lambda > 0$. Since the solutions are not in a closed form, we can use the Newton-Rhapson Algorithm to find numerical solutions for the parameter estimates. The Hessian Matrix can be obtained by taking the second-order partial derivatives with respect to the parameters p and λ.

$$\boldsymbol{H} = \begin{pmatrix} \frac{\partial^2\ell(p,\lambda)}{\partial p^2} & \frac{\partial^2\ell(p,\lambda)}{\partial\lambda\partial p} \\ \frac{\partial^2\ell(p,\lambda)}{\partial p\partial\lambda} & \frac{\partial^2\ell(p,\lambda)}{\partial\lambda^2}, \end{pmatrix} \tag{11}$$

where

$$\frac{\partial^2 \ell(p,\lambda)}{\partial p^2} = \sum_{\{i:y_i=0\}} \frac{\left(2+2\exp(-\lambda)\lambda^{y_i}\right)\left(p^2+q^2\exp(-\lambda)\lambda^{y_i}\right)}{\left(p^2+q^2\exp(-\lambda)\lambda^{y_i}\right)^2}$$
$$- \sum_{\{i:y_i=0\}} \frac{\left(2p-2q\exp(-\lambda)\lambda^{y_i}\right)^2}{\left(p^2+q^2\exp(-\lambda)\lambda^{y_i}\right)^2} + \sum_{\{i:y_i=k\}} \frac{\left(-4(y_i!)+2\exp(-\lambda)\lambda^{y_i}\right)}{\left(2pq(y_i!)+q^2\exp(-\lambda)\lambda^{y_i}\right)}$$
$$- \sum_{\{i:y_i=k\}} \frac{\left((2-4p)(y_i!)-2q\exp(-\lambda)\lambda^{y_i}\right)^2}{\left(2pq(y_i!)+q^2\exp(-\lambda)\lambda^{y_i}\right)^2} - \frac{2}{q^2}\sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} 1,$$

$$\frac{\partial^2 \ell(p,\lambda)}{\partial \lambda \partial p} = \frac{\partial^2 \ell(p,\lambda)}{\partial p \partial \lambda}$$
$$= \sum_{\{i:y_i=0\}} \frac{2pq\exp(-\lambda)\lambda^{y_i}}{\left(p^2+q^2\exp(-\lambda)\lambda^{y_i}\right)^2} + \sum_{\{i:y_i=k\}} \frac{2q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\left(1-\frac{y_i}{\lambda}\right)}{\left(2pq+q^2\frac{\exp(-\lambda)\lambda^{y_i!}}{y_i!}\right)^2},$$

and

$$\frac{\partial^2 \ell(p,\lambda)}{\partial \lambda^2} = \sum_{\{i:y_i=0\}} \frac{p^2q^2\exp(-\lambda)\lambda^{y_i}}{\left(p^2+q^2\exp(-\lambda)\lambda^{y_i}\right)^2}$$
$$+ \sum_{\{i:y_i=k\}} \frac{q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}}{\left(2pq+q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)^2} \times \Bigg[\left(2pq+q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)$$
$$\left((\frac{y_i}{\lambda}-1)^2-\frac{y_i}{\lambda^2}\right) - q^2\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\left(\frac{y_i}{\lambda}-1\right)^2\Bigg] - \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} \frac{y_i}{\lambda^2}.$$

Thus, for data that consists of raw counts $y_i, i = 1, ..., n$, the observed Hessian matrix is

$$\widehat{\boldsymbol{H}} = \begin{pmatrix} \frac{\partial^2 \ell(p,\lambda)}{\partial p^2} & \frac{\partial^2 \ell(p,\lambda)}{\partial \lambda \partial p} \\ \frac{\partial^2 \ell(p,\lambda)}{\partial p \partial \lambda} & \frac{\partial^2 \ell(p,\lambda)}{\partial \lambda^2} \end{pmatrix} \tag{13}$$

evaluated at the maximum likelihood estimates $(\widehat{p}, \widehat{\lambda})$. The observed covariance matrix is the inverse of the negative of the observed Hessian matrix given in equation (13). The diagonal elements of the covariance matrix are the observed variances $\sigma^2(\widehat{p})$ and $\sigma^2(\widehat{\lambda})$ of $\widehat{p}$ and $\widehat{\lambda}$, respectively. Also, the only off-diagonal element is the observed covariance between the maximum likelihood estimates $\widehat{p}$ and $\widehat{\lambda}$.

II.3.2 MOMENT ESTIMATION

Suppose our observations include counts $y_1, \ldots, y_n$ that are independently distributed as DIP(p, λ). Here, we assume that both p and λ are unknown with a pre-determined k and we desire moment estimators for both parameters. The first two population moments as established in Section II.2.2 are $\mu_1 = E(Y) = 2pqk + q^2\lambda$ and $\mu_2 = E(Y^2) = 2pqk^2 + q^2(\lambda + \lambda^2)$. The first two sample moments are $\overline{y_1} = (\sum_{i=1}^{n} y_i)/n$ and $\overline{y_2} = (\sum_{i=1}^{n} y_i^2)/n$. Equating the first two sample moments to those of the population moments yields the two nonlinear equations

$$\begin{cases} \overline{y_1} &= 2pqk + q^2\lambda \\ \overline{y_2} &= 2pqk^2 + q^2(\lambda + \lambda^2), \end{cases} \tag{14}$$

which now must be solved for both p and λ. One may use numerical algorithm such as Newton-Rhapson Method to find the solution. Thus, we can obtain the moment estimators $(\widetilde{p}, \widetilde{\lambda})$ for p and λ, respectively, where $0 \leq p \leq 1$ and $\lambda > 0$.

Consider $\boldsymbol{D}$ the matrix of the first order partial derivatives of the first two population moments with respect to p and λ. That is,

$$\begin{aligned} \boldsymbol{D} &= \begin{pmatrix} \dfrac{\partial \mu_1}{\partial p} & \dfrac{\partial \mu_1}{\partial \lambda} \\ \dfrac{\partial \mu_2}{\partial p} & \dfrac{\partial \mu_2}{\partial \lambda} \end{pmatrix} \\ &= \begin{pmatrix} 2k(1-2p) - 2q\lambda & q^2 \\ \\ 2k^2(1-2p) - 2q(\lambda + \lambda^2) & q^2(2\lambda + 1) \end{pmatrix}. \end{aligned} \tag{15}$$

Let $\boldsymbol{\Sigma}$ be the covariance matrix of Y and Y^2 given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} Var(Y) & Cov(Y, Y^2) \\ Cov(Y^2, Y) & Var(Y^2) \end{pmatrix} \tag{16}$$

where,

$$\begin{aligned} Var(Y) &= 2pqk^2 + q^2(\lambda^2 + \lambda) - \left(2pqk + q^2\lambda\right)^2, \\ Var(Y^2) &= 2pqk^4 + q^2\left(\lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda\right) - \Big(2pqk^2 + q^2\left(\lambda + \lambda^2\right)\Big)^2, \text{ and} \\ Cov(Y^2, Y) &= Cov(Y, Y^2) \\ &= \Big(2pqk^3 + q^2\left(\lambda^3 + 3\lambda^2 + \lambda\right)\Big) - \Big((2pqk + q^2\lambda)\left(2pqk^2 + q^2\left(\lambda + \lambda^2\right)\right)\Big). \end{aligned}$$

Then the asymptotic covariance matrix for the moment estimators $(\widetilde{p}, \widetilde{\lambda})$ is given by the matrix $\boldsymbol{A}$ (see Chaganty and Shi 2004).

$$\begin{aligned} \boldsymbol{A} &= \frac{1}{n}\left(\boldsymbol{D}\right)^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{D}^{\intercal}\right)^{-1} \\ &= \frac{1}{n(\det(D))^2}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \end{aligned} \tag{17}$$

where,

$$\begin{aligned} a_{11} &= 2q^6\lambda^2 + q^6\lambda^4 - q^8\lambda^4 + 8pq^5k^2\lambda^2 - 8q^5k^3p\lambda \\ &+ 8q^5pk^2\lambda - 16p^2k^2q^6\lambda^2 - 16k^2p^2q^6\lambda - 8pkq^7\lambda^3 - 4pkq^7\lambda^2 \\ &+ 16p^2k^3q^6\lambda + 4pk^2q^7\lambda^2 + 2pq^5k^2 - 4p^2k^2q^6 - 4k^3pq^5 \\ &+ 8k^3p^2q^6 + 2q^5pk^4 - 4p^2q^6k^4, \end{aligned}$$

$$\begin{aligned} a_{12} = a_{21} &= -4q^4pk\lambda^3 + 8q^4pk\lambda^2 + 12q^4pk^2\lambda^2 - 12q^4pk^3\lambda^2 \\ &+ 12q^4pk^2\lambda^3 + 2q^4k\lambda^3 - 4q^4k\lambda^2 + 4q^5\lambda^3 - 4pkq^4\lambda^4 \\ &- 2k^2q^4\lambda^3 + 2kq^4\lambda^4 + 2k^2q^6\lambda^3 - 2kq^6\lambda^4 - 12pk^2q^5\lambda^2 \\ &- 2kq^6\lambda^3 - 8pk^3q^4\lambda - 8pk^2q^5\lambda^3 + 8pk^3q^5\lambda + 4pk^2q^4\lambda \\ &- 4pk^2q^5\lambda + 12pk^3q^5\lambda^2 + 4pk^4q^4\lambda - 4pk^4q^5\lambda, \end{aligned}$$

$$
\begin{aligned}
a_{22} = {} & -112pk^2q^2\lambda^2 - 16pk^2q^2\lambda + 8pk^2q^3\lambda^2 + 32pk^3q^2\lambda \\
& + 8pk^2q^3\lambda^4 - 16pk^4q^2\lambda^2 + 16p^2k^4q^2\lambda^2 - 16pk^4q^2\lambda \\
& + 16q^2p^2k^4\lambda + 28k^2q^2\lambda^2 + 4k^2q^2\lambda - 4k^2q^2\lambda^2 + 8k^3q^4\lambda^3 \\
& + 4q^5\lambda^5 + 8q64\lambda^4 + 8pk^4q^3\lambda^2 + 32pkq^3\lambda^4 + 32pk^3q^2\lambda^3 \\
& - 32p^2q^2k^3\lambda^3 + 48pkq^3\lambda^3 + 96pq^2k^3\lambda^2 - 96p^2q^2k^3\lambda^2 \\
& - 32p^2q^2k^3\lambda - 16pk^3q^3\lambda^3 - 16pk^3q^3\lambda^2 - 96pq^2k^2\lambda^3 \\
& + 96p^2q^2k^2\lambda^3 + 112p^2q^2k^2\lambda^2 + 16p^2q^2k^2\lambda - 16k^2pq^2\lambda^4 \\
& + 16p^2q^2k^2\lambda^4 + 8k^2q^3\lambda^3 + 4k^4q^2\lambda^2 + 4k^4q^2\lambda - 4k^4q^4\lambda^2 \\
& - 16kq^3\lambda^4 - 8k^3q^2\lambda^3 - 24kq^3\lambda^3 - 24q^2k^3\lambda^2 - 8k^3q^2\lambda \\
& + 8k^3q^4\lambda^2 + 24k^2q^2\lambda^3 - 8k^2q^4\lambda^3 + 4k^2q^2\lambda^4 - 4k^2q^4\lambda^4,
\end{aligned}
$$

and

$$\det(D) = 4kq^2\lambda - 8pkq^2\lambda - 2q^3\lambda^2 + 2kq^2 - 4q^2pk - 2p^2k^2 + 4pk^2q^2.$$

The diagonal elements of covariance matrix $\boldsymbol{A}$ are the asymptotic variances $\sigma^2(\widetilde{p})$ and $\sigma^2(\widetilde{\lambda})$ of moment estimators $\widetilde{p}$ and $\widetilde{\lambda}$, respectively.

In some cases where p is already known, then λ can be estimated by $\widehat{\lambda} = \dfrac{\overline{y_1} - 2pqk}{q^2}$. Here, $\overline{y_1}$ is calculated by $(\sum_{i=1}^{n} y_i)/n$. We know that $\sqrt{n}(y - \mu_1)$ asymptotically follows normal distribution with mean zero and variance $\sigma^2 = Var(Y) = 2pqk^2 + q^2(\lambda^2 + \lambda) - (2pqk + q^2\lambda)^2$. Then, by delta theorem, $\overline{y_1}$ is normally distributed with mean μ_1 and variance $(\sigma^2)/n$. Thus, $\widehat{\lambda}$ also follows a normal distribution with mean λ and variance $(\sigma^2)/n$.

II.4 METHODS OF ESTIMATION FOR GROUPED DATA

Sometimes, the data is observed as frequencies of counts. That is, instead of observing the raw counts, y_i, $i = 1, ..., n$, we observe frequency, n_j, of count $(Y_i = j)$, $j = 1, \ldots, m$. Thus, n_0 is the inflated numbers of 0's, n_k is the inflated numbers of $k's$, and n_j is the numbers of $j's$. In this section, we will explore both methods of estimation, the maximum likelihood and the method of moments, for count data that is given as grouped frequencies.

II.4.1 MAXIMUM LIKELIHOOD ESTIMATION

Under the assumption that the observations are independently distributed as DIP(p, λ), the log-likelihood function is

$$\begin{aligned}\ell(p,\lambda|n_j) &= n_0 \log\left(p^2 + q^2\exp(-\lambda)\right) \\ &\quad + n_k \log\left(2pq + q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right) \\ &\quad + \sum_{\substack{j=1 \\ j\neq k}}^{\infty} n_j\left(2\log(q) - \lambda + j\log(\lambda) - \log(j!)\right).\end{aligned} \tag{19}$$

The score functions are

$$\begin{aligned}\frac{\partial \ell(p,\lambda)}{\partial p} &= n_0\frac{(2p - 2q\exp(-\lambda))}{(p^2+q^2\exp(-\lambda))} + n_k\frac{\left(2 - 4p - 2q\frac{\exp(-\lambda)\lambda^k}{k!}\right)}{\left(2pq + q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right)} \\ &\quad + \sum_{\substack{j=1 \\ j\neq k}}^{\infty} n_j\left(\frac{-2}{q}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \ell(p,\lambda)}{\partial \lambda} &= n_0\frac{-q^2\exp(-\lambda)}{(p^2+q^2\exp(-\lambda))} + n_k\frac{q^2\exp(-\lambda)\lambda^k}{(2pq(k!) + q^2\exp(-\lambda)\lambda^k)}\left(\frac{k}{\lambda} - 1\right) \\ &\quad + \sum_{\substack{j=1 \\ j\neq k}}^{\infty} n_j\frac{(j-\lambda)}{\lambda}.\end{aligned}$$

In order to solve for the maximum likelihood estimates of (p, λ), we solve for the roots of the above score functions. The second order partial derivatives of the log-likelihood function in matrix form are

$$\boldsymbol{H} = \begin{pmatrix} \frac{\partial^2 \ell(p,\lambda)}{\partial p^2} & \frac{\partial^2 \ell(p,\lambda)}{\partial p\partial\lambda} \\ \frac{\partial^2 \ell(p,\lambda)}{\partial\lambda\partial p} & \frac{\partial^2 \ell(p,\lambda)}{\partial \lambda^2}\end{pmatrix} \tag{20}$$

where,

$$\frac{\partial^2 \ell(p,\lambda)}{\partial p^2} = \frac{n_0}{\left(p^2+q^2\exp(-\lambda)\right)^2}\left[\left(2+2\exp(-\lambda)\right)\left(p^2+q^2\exp(-\lambda)\right) - \right.$$
$$\left. - \left(2p - 2q\exp(-\lambda)\right)^2\right] + \frac{n_k}{\left(2pq+q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right)^2} \times \left[\left(2pq + q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right)\right.$$
$$\left.\left(-4+2\frac{\exp(-\lambda)\lambda^k}{k!}\right) - \left(2-4p-2q\frac{\exp(-\lambda)\lambda^k}{k!}\right)^2\right] + \sum_{\substack{j=1\\ j\neq k}}^{\infty} n_j\left(\frac{-2}{q^2}\right),$$

$$\frac{\partial^2 \ell(p,\lambda)}{\partial\lambda\partial p} = \frac{\partial^2 \ell(p,\lambda)}{\partial p\partial\lambda}$$
$$= \frac{2n_0 pq\exp(-\lambda)}{\left(p^2+q^2\exp(-\lambda)\right)^2} + \frac{2n_k q^2\frac{\exp(-\lambda)\lambda^k}{k!}}{\left(2pq+q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right)^2} \times \left(1-\frac{k}{\lambda}\right),$$

and

$$\frac{\partial^2 \ell(p,\lambda)}{\partial\lambda^2} = \frac{n_0 p^2 q^2\exp(-\lambda)}{\left(p^2+q^2\exp(-\lambda)\right)^2}$$
$$+ \frac{n_k q^2\frac{\exp(-\lambda)\lambda^k}{k!}}{\left(2pq+q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right)^2} \times \left[\left(2pq+q^2\frac{\exp(-\lambda)\lambda^k}{k!}\right)\left(\left(\frac{k}{\lambda}-1\right)^2 - \frac{k}{\lambda^2}\right)\right.$$
$$\left. -q^2\left(\frac{k}{\lambda}-1\right)^2\left(\frac{\exp(-\lambda)\lambda^k}{k!}\right)\right] + \sum_{\substack{j=1\\ j\neq k}}^{\infty} n_j\left(\frac{-j}{\lambda^2}\right).$$

Thus, for data that consists of frequency n_j for count j, the observed Hessian matrix is given by equation (20) evaluated at the maximum likelihood estimates $(\widehat{p}, \widehat{\lambda})$. And, the observed covariance matrix is the inverse of negative of the observed Hessian matrix given above. The diagonal elements are the observed asymptotic variances $\sigma^2(\widehat{p})$ and $\sigma^2(\widehat{\lambda})$ of $\widehat{p}$ and $\widehat{\lambda}$, respectively.

II.4.2 METHOD OF MOMENTS ESTIMATION

Suppose that our data consists of n independent observations of frequencies n_j of count $j, j = 0, \ldots, m$. We assume that the frequencies n_0 and n_k are inflated frequencies of counts 0's and k's. We can obtain the moment estimates for the parameter p and λ using the following technique. For the observations consisting of grouped frequencies, n_j for count j, the first two sample moments are $\overline{n_1} = (1/n) \sum_{j=0}^{m}(jn_j)$ and $\overline{n_2} = (1/n) \sum_{j=0}^{m}(j^2 n_j)$, which are computationally equivalent to the sample moments for raw counts (see Section II.3.2). Here n is sum of all frequencies, $n_j's$. We know that the population moments are $\mu_1 = E(Y)$ and $\mu_2 = E(Y^2)$ as in equation (5). Equating the population moments to those of sample moments, we obtain a similar system of of equations seen previously in Section II.3.2.

$$\begin{aligned} \overline{n_1} &= 2pqk + q^2\lambda \\ \overline{n_2} &= 2pqk^2 + q^2\left(\lambda + \lambda^2\right) \end{aligned} \tag{22}$$

We solve the above system of equations to obtain the moment estimates of p and λ. Note that k is known here, and the parameter constraints are $0 \leq p \leq 1$ and $\lambda > 0$.

Let $\boldsymbol{D}$ be the matrix of first order partial derivatives of the first two moments with respect to p and λ. That is,

$$\boldsymbol{D} = \begin{pmatrix} \dfrac{\partial \mu_1}{\partial p} & \dfrac{\partial \mu_1}{\partial \lambda} \\ \dfrac{\partial \mu_2}{\partial p} & \dfrac{\partial \mu_2}{\partial \lambda} \end{pmatrix} \tag{23}$$

where the elements are as given by equation (15). Similarly, let $\boldsymbol{\Sigma}$ be the covariance matrix as given in equation (16). Then, the asymptotic covariance matrix for the method of moments estimates of (p, λ) is given by

$$\boldsymbol{A} = \frac{1}{n}\left(\boldsymbol{D}\right)^{-1} \Sigma \left(\boldsymbol{D}^{\mathsf{T}}\right)^{-1}. \tag{24}$$

Matrix $\boldsymbol{A}$ is the inverse Godambe's information matrix. The diagonal elements of $\boldsymbol{A}$ are the asymptotic variances $\sigma^2(\widetilde{p})$ and $\sigma^2(\widetilde{\lambda})$ of moment estimators $\widetilde{p}$ and $\widetilde{\lambda}$, respectively.

II.4.3 ASYMPTOTIC RELATIVE EFFICIENCY COMPARISONS

To compare the performance of the two maximum likelihood and method of moment estimators, of the parameters p and λ, we use a large-sample measure of relative efficiency due to Pitman. Pitman's asymptotic relative efficiency (ARE) of moment estimators $(\widetilde{p},\ \widetilde{\lambda})$ with respect to maximum likelihood estimators $(\widehat{p}, \widehat{\lambda})$ are the ratios of their asymptotic variances $e(\widetilde{p},\ \widehat{p}) = \sigma^2(\widehat{p})/\sigma^2(\widetilde{p})$ and $e(\widetilde{\lambda},\ \widehat{\lambda}) = \sigma^2(\widehat{\lambda})/\sigma^2(\widetilde{\lambda})$, respectively. If ARE is less than 1, we conclude that the maximum likelihood estimators are better than the moment estimators and vice versa if ARE is greater than 1.

In Section II.2.3, we found that Fisher information $I(p)$ given by the equation

$$
\begin{aligned}
I(p) = {} & \frac{-2\,(1+\exp(-\lambda))\,(p^2+q^2\exp(-\lambda)) - (2p-2q\exp(-\lambda))^2}{(p^2+q^2\exp(-\lambda))} \\
& - \frac{\left[\left(2pq+q^2\dfrac{\exp(-\lambda)\lambda^k}{k!}\right)\left(-4+2\dfrac{\exp(-\lambda)\lambda^k}{k!}\right) - \left(2-4p-2q\dfrac{\exp(-\lambda)\lambda^k}{k!}\right)^2\right]}{\left(2pq+q^2\dfrac{\exp(-\lambda)\lambda^k}{k!}\right)} \\
& + 2\left(1-\exp(-\lambda) - \frac{\exp(-\lambda)\lambda^k}{k!}\right).
\end{aligned}
$$

The asymptotic variance $\sigma^2(\widehat{p})$ of maximum likelihood estimate $\widehat{p}$ has a Cramér-Rao lowerbound of $1/I(p)$ evaluated at $\widehat{p}$. In Sections II.3.2 and II.4.2, we obtained the inverse Godambe's information matrix $\boldsymbol{A}$ whose first diagonal element is the asymptotic variance of the moment estimator of p, $\sigma^2(\widetilde{p})$. Taking the ratio $\sigma^2(\widehat{p})/\sigma^2(\widetilde{p})$ we obtain the relative efficiency of moment estimator with respect to maximum likelihood estimator of p. Table 4 represents efficiencies $e(\widetilde{p},\ \widehat{p})$ calculated for various values of p and λ in the parameter space $0 \le p \le 1$ and $1 \le \lambda \le 10$ for DIP (p, λ) model with inflated 0's and 3's. Table 5 represents efficiencies $e(\widetilde{p},\ \widehat{p})$ calculated for various values of p and λ for the DIP (p, λ) model with inflated 0's and 6's. As we can see, the efficiencies $e(\widetilde{p},\ \widehat{p})$ are less than 1 indicating that the moment estimator are less efficient than the maximum likelihood estimators of the parameter p. Efficiencies were also calculated for models with various values of k, and they show the same trend.

Similarly, we found that Fisher information, $I(\widehat{\lambda})$, for λ which is given by the

Table 4: Relative Efficiencies for DIP (p, λ) Model for Inflated 0's and 3's

	$e(\widetilde{p}, \widehat{p})$				$e(\widetilde{\lambda}, \widehat{\lambda})$			
p	$\lambda = 3$	$\lambda = 5$	$\lambda = 7$	$\lambda = 9$	$\lambda = 3$	$\lambda = 5$	$\lambda = 7$	$\lambda = 9$
0.1	0.0766	0.0859	0.4343	0.6629	0.7709	0.2219	0.5807	0.6988
0.2	0.2926	0.2592	0.6384	0.8174	0.2930	0.3895	0.6172	0.6973
0.3	0.0819	0.4404	0.7524	0.8854	0.2784	0.4869	0.6215	0.6712
0.4	0.2594	0.5853	0.8205	0.9223	0.4875	0.5419	0.6089	0.6302
0.5	0.4475	0.6953	0.8659	0.9454	0.5946	0.5746	0.5892	0.5829
0.6	0.6106	0.7808	0.9007	0.9614	0.6555	0.5951	0.5678	0.5352
0.7	0.7427	0.8499	0.9297	0.9735	0.6938	0.6086	0.5469	0.4907
0.8	0.8478	0.9075	0.9551	0.9835	0.7197	0.6180	0.5276	0.4495
0.9	0.9319	0.9569	0.9784	0.9921	0.7384	0.6248	0.5101	0.4132

NOTE: The efficiencies are calculated for the model with inflated zero's and 3's. $e(\widetilde{p}, \widehat{p})$ is the asymptotic relative efficiency of moment estimator of p relative to its maximum likelihood estimator. $e(\widetilde{\lambda}, \widehat{\lambda})$ is the asymptotic relative efficiency of moment estimator of λ relative to its maximum likelihood estimator.

equation

$$
\begin{aligned}
I(\lambda) = {} & \frac{-p^2 q^2 \exp(-\lambda)}{(p^2 + q^2 \exp(-\lambda))} + \frac{q^2}{\lambda}\left(1 - \frac{\exp(-\lambda)\lambda^{k-1}}{(k-1)!}\right) \\
& + \frac{q^2 \dfrac{\exp(-\lambda)\lambda^k}{k!}}{2pq + q^2 \dfrac{\exp(-\lambda)\lambda^k}{k!}} \times \left[\left(\frac{k}{\lambda} - 1\right)^2 q^2 \frac{\exp(-\lambda)\lambda^k}{k!}\right] \\
& - q^2 \frac{\exp(-\lambda)\lambda^k}{k!}\left(\left(\frac{k}{\lambda} - 1\right)^2 - \frac{k}{\lambda^2}\right)
\end{aligned}
$$

evaluated at the maximum likelihood estimates of p and λ. The asymptotic variance $\sigma^2(\widehat{\lambda})$ of maximum likelihood estimate $\widehat{\lambda}$ has a lower bound $1/I(\widehat{\lambda})$. In Sections II.3.2 and II.4.2, we obtained the inverse Godambe information matrix $\boldsymbol{A}$ whose second diagonal element is the asymptotic variance of the moment estimator of λ, $\sigma^2(\widetilde{\lambda})$. Taking the ratio $\sigma^2(\widehat{\lambda})/\sigma^2(\widetilde{\lambda})$ we obtain the relative efficiency of moment estimator with respect to maximum likelihood estimator of p. Table 4 represents efficiencies $e(\widetilde{\lambda},\ \widehat{\lambda})$ calculated for various values of p and λ in the parameter space $0 \leq p \leq 1$

and $1 \leq \lambda \leq 10$ for the DIP model with inflated 0's and 3's. Table 5 represents efficiencies $e(\widetilde{\lambda},\ \widehat{\lambda})$ calculated for various values of p and λ for the DIP model with inflated 0's and 6's. As we can see, the efficiencies $e(\widetilde{\lambda},\ \widehat{\lambda})$ are less than 1 indicating that the moment estimator are less efficient than the maximum likelihood estimators of the parameter λ.

Table 5: Relative Efficiencies for DIP (p, λ) Model for Inflated 0's and 6's

	$e(\widetilde{p},\widehat{p})$				$e(\widetilde{\lambda},\widehat{\lambda})$			
p	$\lambda=3$	$\lambda=5$	$\lambda=7$	$\lambda=9$	$\lambda=3$	$\lambda=5$	$\lambda=7$	$\lambda=9$
0.1	0.1877	0.0152	0.0023	0.1468	0.3659	0.6813	0.0178	0.3197
0.2	0.3803	0.0274	0.1170	0.3544	0.5086	0.6890	0.3013	0.4399
0.3	0.5413	0.1692	0.2978	0.5081	0.4549	0.6916	0.4411	0.4854
0.4	0.6653	0.3464	0.4619	0.6211	0.3350	0.6928	0.5071	0.5055
0.5	0.7617	0.5090	0.5969	0.7105	0.2449	0.6934	0.5442	0.5158
0.6	0.8363	0.6461	0.7074	0.7851	0.1866	0.6939	0.5676	0.5216
0.7	0.8939	0.7595	0.7991	0.8491	0.1486	0.6942	0.5837	0.5251
0.8	0.9382	0.8537	0.8764	0.9053	0.1229	0.6944	0.5953	0.5273
0.9	0.9727	0.9328	0.9426	0.9552	0.1047	0.6945	0.6041	0.5288

NOTE: The efficiencies are calculated for the model with inflated zero's and 6's.

II.4.4 TESTING

Goodness-of-fit statistics help us to check the adequacy of the chosen model fit. In order to test the null hypothesis whether DIP(p, λ) model is a good fit, we can take the observed counts and the expected counts to obtain the Pearson and the likelihood-ratio goodness-of-fit statistics. For the ith cell, denote the observed count by n_i and the expected count by $\widehat{\mu}_i$. The Pearson and the likelihood-ratio goodness-of-fit statistics equal

$$\chi^2 = \sum \frac{(n_i - \widehat{\mu}_i)^2}{\widehat{\mu}_i}$$
$$G^2 = 2\sum n_i \log \frac{n_i}{\widehat{\mu}_i}$$

When the fitted values are not small (exceeding 5), these test statistics have approximate chi-squared distributions. The *df* equal the number of response cell counts minus the number of model parameter(s). A large test statistics and small p-values suggest a poor-model fit.

For performing significance tests of hypothesis H_0: $\beta = 0$ about parameters in the model, we can use the large-sample normality assumption of ML estimates. The test statistic $z = \widehat{\beta}/ASE$, where ASE is the asymptotic standard error of $\widehat{\beta}$, has an approximate standard normal distribution when $\beta = 0$. Equivalently, Wald's statistic z^2, which divides a parameter estimate by its standard error and then squares it, can also be used to test the two-sided hypothesis. One may also use the likelihood-ratio test statistic, $-2(L_0 - L_1)$, where L_0 and L_1 denoted the maximized log-likelihood functions. It has a chi-squared distribution with $df = 1$.

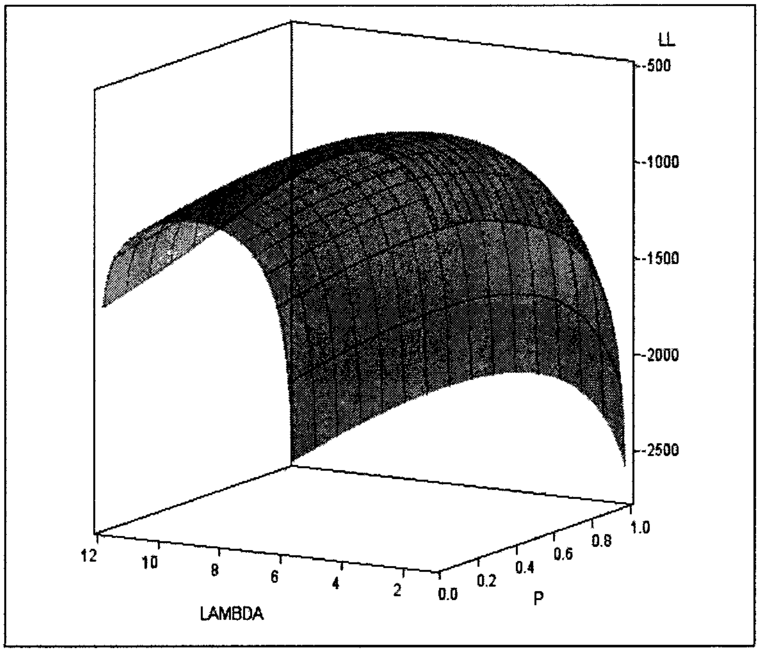

Figure 2: Log-likelihood Graph of LOS data using Doubly Inflated Poisson Model with parameters (p, λ)

II.5 ILLUSTRATION OF METHODS

In this section, we return to our motivating example on hospital stay data and illustrate the analysis of that data using the methods discussed in this chapter. In our hospital stay data, we have grouped frequencies for the count response of length of stay, LOS (see Table 2). We notice that there are inflated counts of patients (n_0 = 55) who receive outpatient treatment and/or care, and inflated counts of patients (n_3 = 75) who possibly under insurance coverage require 3-day inpatient care. We also have 35 patients who stay overnight, 35 patients who stay two nights in a row, and so on.

In this example, we assume that the observations on patients are independent. The count response of LOS can be assumed to have DIP(p, λ) distribution. For the analysis, we obtain the maximum likelihood estimates for the parameters p and λ using the ZIP and DIP (p, λ) probability models. The parameter estimates for p and λ are as given in Table 14. The asymptotic variances for ML estimates using the score function were found to be $\widehat{\sigma^2(\widehat{p})} = 0.002$ and $\widehat{\sigma^2(\widehat{\lambda})} = 0.012$, and the asymptotic covariance of the two parameter estimates was approximately zero. Using the ML estimates for the parameters under both probability models, one can easily compute the expected frequencies for the length of stay data. These fitted frequencies along with the observed frequencies are plotted in Figure 3.

At first, the ZIP model accounts quite well for the first inflation at LOS$=0$, however it fails to account for the second inflated count at LOS $= 3$. Thus, we attempt to fit the observed data using ML estimation for DIP (p, λ) model. Figure 2 plots the log-likelihood values for the observed data using the range of the parameters: $0 \leq p \leq 1$ and $0 < \lambda \leq 12$. The plot shows that an optimal solution is attainable at the peak where the largest value for negative log-likelihood value is observed. As seen from Figure 3, DIP(p, λ) accounts fails to account fairly for the first inflation, however does begin to account the second inflation and also improves on the estimations for LOS > 4. Hence, one needs to seek improvement in the modeling and estimation techniques for the second inflation along with modeling and estimation for the first inflation. The maximum likelihood estimation was done using PROC IML in SAS and results verified using PROC NLP and PROC NLMIXED.

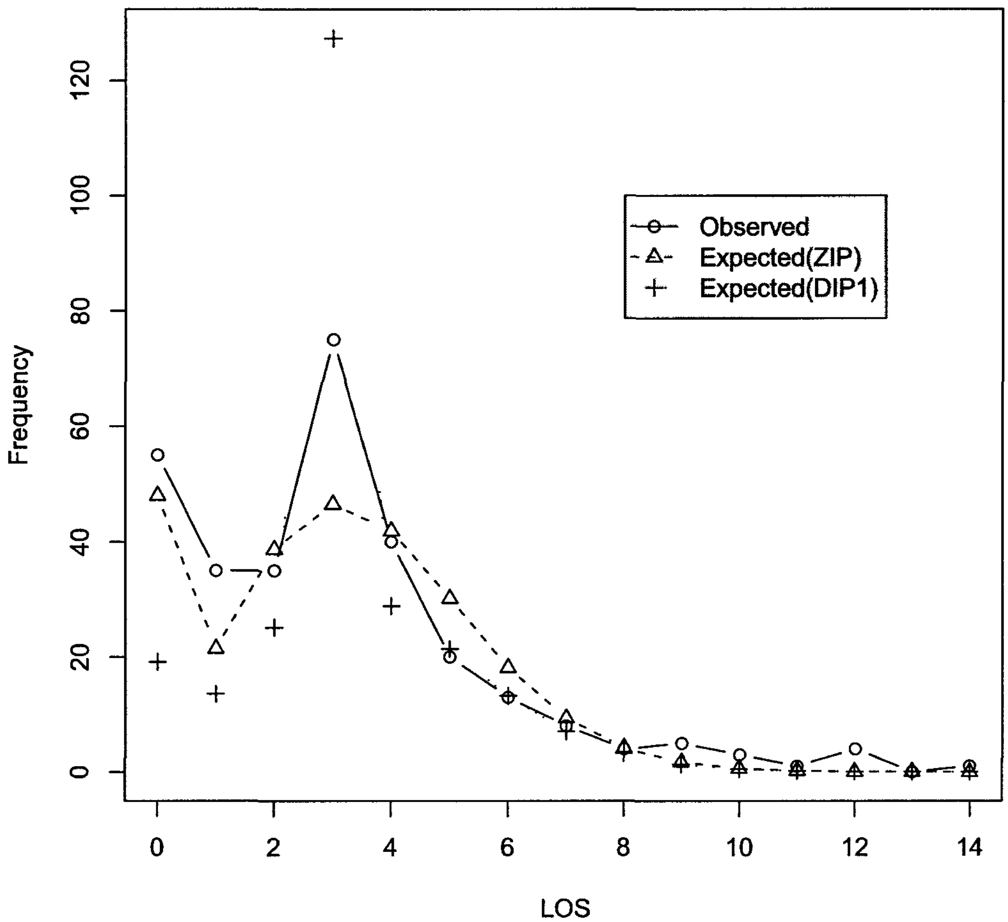

Figure 3: LOS Data: Fitting various Poisson models

CHAPTER III

THE SECOND DOUBLY INFLATED POISSON MODEL

In Chapter II, we introduced a Doubly Inflated Poisson model to deal with count responses with two peaks. In this chapter, we will study another DIP probability that can also be considered as a finite mixture model that accounts for inflated counts of zeros as well as inflated counts of k's.

In Section III.1, we describe in detail the development of Doubly Inflated Poisson model with three parameters p_1, p_2 and λ, using a latent variable. We also discuss the distributional properties of the probability model including moments, expectation, variance, and Fisher information. Section III.2 describes estimation techniques for data that consists of raw count responses using maximum likelihood approach and method of moments. Section III.3 describes parameter estimation methods such as maximum likelihood estimation and moment estimation for data consisting of grouped frequencies of count responses. We also compare the two estimation approaches using relative efficiencies as well as small sample efficiency calculations. In Section III.4, we revisit our illustration on the hospital stay data using DIP (p_1, p_2, λ) probability model and then compare the analysis to the ZIP and DIP (p, λ) models.

III.1 PROBABILITY MODEL

Another approach to modeling a doubly inflated count response variable where a large number of zeros and k's are observed is by introducing another parameter of interest that can explain inflated k's. That is, we let Z be a random variable such that it has following probability mass function:

$$P(Z=z) = \begin{cases} p_1, & \text{if } z=0; \\ p_2, & \text{if } z=1; \\ p_3, & \text{if } z=2, \end{cases}$$

where $p_1 + p_2 + p_3 = 1$ with $p_1, p_2, p_3 \neq 0$. Let Y be a random variable with the following characteristic: Y given $Z=0$ is degenerate at 0, Y given $Z=1$ is degenerate at k, and Y given $Z = 2$ is Poisson process with parameter λ, $\lambda > 0$. That is, the conditional probability function of Y given Z is

$$P(Y = y|Z = z) = \begin{cases} 1, & \text{for } z = 0, y = 0; \\ 1, & \text{for } z = 1, y = k; \\ \dfrac{\exp(-\lambda)\lambda^y}{y!}, & \text{for } z = 2, y = 0, 1, 2, \ldots \end{cases} \tag{25}$$

Then the joint distribution of Y and Z can be obtained as

$$P(Z = z, Y = y) = \begin{cases} p_1, & \text{if } z = 0, y = 0; \\ p_2, & \text{if } z = 1, y = k; \\ p_3\dfrac{\exp(-\lambda)\lambda^y}{y!}, & \text{if } z = 2, y = 0, 1, 2, \ldots \end{cases}$$

Thus, the probability mass function of Y is as follows:

$$P(Y = y) = \begin{cases} p_1 + p_3 \exp(-\lambda), & \text{for } y = 0; \\ \\ p_2 + p_3\dfrac{\exp(-\lambda)\lambda^k}{k!}, & \text{for } y = k; \\ \\ p_3\dfrac{\exp(-\lambda)\lambda^y}{y!}, & \text{for } y = 1, 2, \ldots \neq k. \end{cases} \tag{26}$$

We call this distribution as *Doubly Inflated Poisson* (p_1, p_2, λ), abbreviated as DIP(p_1, p_2, λ). It is that p_1 replaces p^2, p_2 replaces $2pq$, and p_3 replaces q^2 of the DIP (p, λ) probability model. In this probability model, an additional proportion of p_1 would explain the inflated zero counts and an additional proportion of p_2 would explain the inflated k counts. It is quite clear to note that as $p_2 \to 0$, this model reduces to the Zero-Inflated Poisson (ZIP) distribution due to Lambert (1992). A few of the probability mass distributions are generated for a known k using arbitrary values of p_1, p_2, and λ. Similar to the first model as mentioned in Section II.2, the probability mass distributions for this model also shows two peaks for each of the distributions (see Figure 4). One can also obtain the conditional distribution of Z given Y as shown in the Table 6.

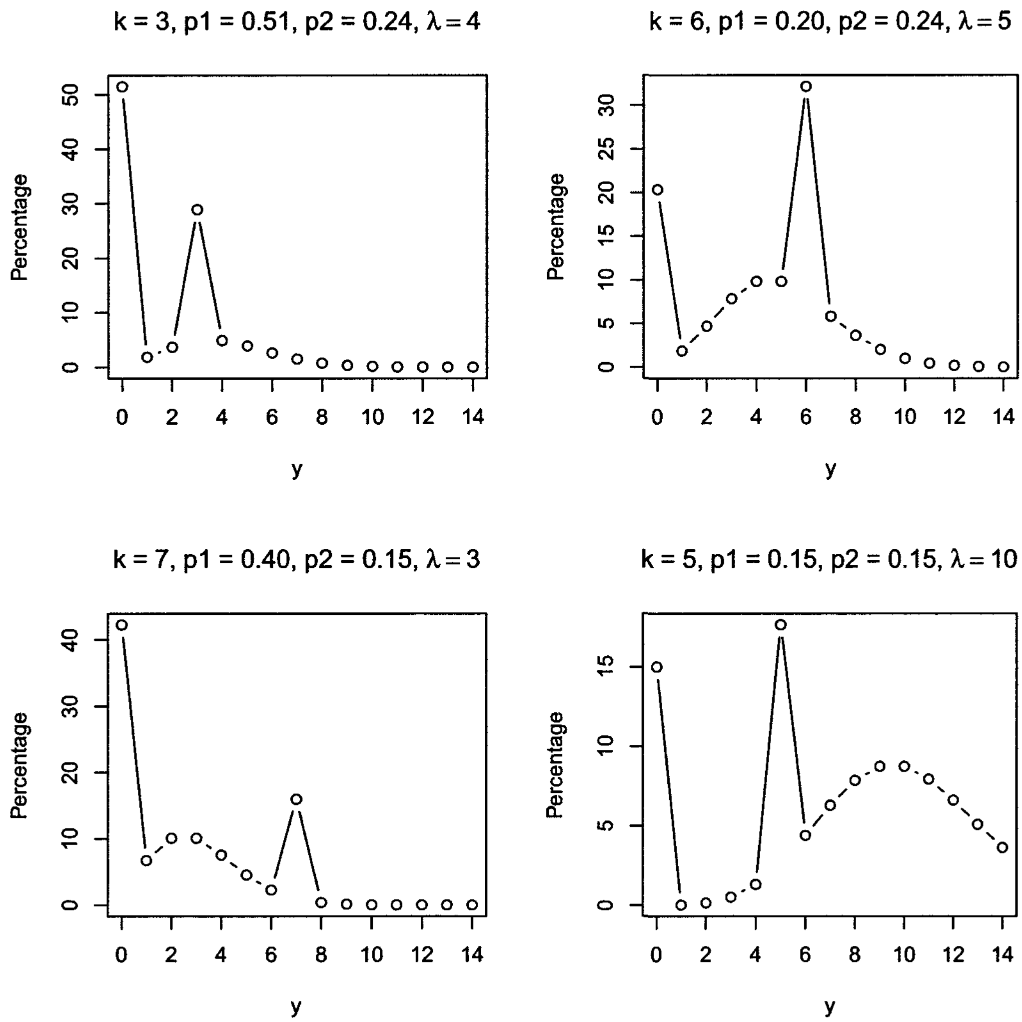

Figure 4: A Few Doubly Inflated Poisson (p_1, p_2, λ) Distributions for a known k

Table 6: $P(Z = z|Y = y)$ of DIP (p_1, p_2, λ)

	Y		
Z	0	k	$1, 2, \ldots \neq k$
0	$\dfrac{p_1}{p_1 + p_3 \exp(-\lambda)}$	0	0
1	0	$\dfrac{p_2(k!)}{p_2(k!) + p_3 \exp(-\lambda)\lambda^k}$	0
2	$\dfrac{p_3 \exp(-\lambda)}{p_1 + p_3 \exp(-\lambda)}$	$\dfrac{p_3 \exp(-\lambda)\lambda^k}{p_2(k!) + p_3 \exp(-\lambda)\lambda^k}$	1

NOTE: The sum of all entries in a column is 1.

III.1.1 MOMENTS, EXPECTATION, AND VARIANCE

Using the conditional distribution of Y given Z as described by equation (25), the conditional expectation of Y given Z is

$$E(Y|Z = z) = \begin{cases} 0, & \text{when } z = 0; \\ k, & \text{when } z = 1; \\ \lambda, & \text{when } z = 2. \end{cases}$$

and, that of Y^2 given Z is

$$E(Y^2|Z = z) = \begin{cases} 0, & \text{when } z = 0; \\ k^2, & \text{when } z = 1; \\ \lambda^2 + \lambda, & \text{when } z = 2. \end{cases}$$

Let μ_r be the rth moment of Y which is distributed as DIP(p_1, p_2, λ). Then

$$\begin{aligned}
\mu_1 &= E(Y) = p_2 k + p_3 \lambda, \\
\mu_2 &= E(Y^2) = p_2 k^2 + p_3(\lambda^2 + \lambda), \\
\mu_3 &= E(Y^3) = p_2 k^3 + p_3(\lambda + 3\lambda^2 + \lambda^3), \text{ and} \\
\mu_4 &= E(Y^4) = p_2 k^4 + p_3(\lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4).
\end{aligned} \tag{27}$$

Thus the expected value $\mu = E(Y)$ and the variance $\sigma^2 = Var(Y)$ are

$$E(Y) = p_2 k + p_3 \lambda, \text{ and} \tag{28}$$

$$\begin{aligned}
Var(Y) &= p_2 k^2 + p_3(\lambda^2 + \lambda) - \mu^2 \\
&= \mu - \mu^2 + p_3 \lambda^2 - p_2(k^2 - k).
\end{aligned} \tag{29}$$

It is clear that the mean and variance of Y are not equal, thereby explaining variability caused by two inflated counts.

III.1.2 FISHER INFORMATION

Suppose Y is distributed as DIP(p_1, p_2, λ) with mass function $p(y) = P(Y = y)$ as given by (26). Then Fisher information matrix is

$$\boldsymbol{I} = \begin{pmatrix} I(p_1) & I(p_1, p_2) & I(p_1, \lambda) \\ I(p_1, p_2) & I(p_2) & I(p_2, \lambda) \\ I(p_1, \lambda) & I(p_2, \lambda) & I(\lambda) \end{pmatrix}, \tag{30}$$

where

$$\begin{aligned}
I(p_1, \lambda) &= -E\left(\frac{\partial^2 \log(p(Y))}{\partial \lambda \, \partial p_2}\right) \\
&= \frac{-\exp(-\lambda)\,(p_1 + p_3)}{\Big(p_1 + p_3 \exp(-\lambda)\Big)} + \frac{p_2 \exp(-\lambda) \lambda^k \left(\frac{k}{\lambda} - 1\right)}{\Big(p_2(k!) + p_3 \exp(-\lambda) \lambda^k\Big)},
\end{aligned}$$

$$I(p_1) = -E\left(\frac{\partial^2 \log(p(Y))}{\partial {p_1}^2}\right)$$

$$= \frac{\Big(1-\exp(-\lambda)\Big)^2}{\Big(p_1 + p_3\exp(-\lambda)\Big)} + \frac{\left(\frac{\exp(-\lambda)\lambda^k}{k!}\right)^2}{\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^k}{k!}\right)}$$

$$+ \frac{1}{p_3}\left(1-\exp(-\lambda) - \frac{\exp(-\lambda)\lambda^k}{k!}\right),$$

$$I(p_1, p_2) = -E\left(\frac{\partial^2 \log(p(Y))}{\partial p_1 \partial p_2}\right)$$

$$= \frac{-\exp(-\lambda)\Big(1-\exp(-\lambda)\Big)}{\Big(p_1 + p_3\exp(-\lambda)\Big)} - \frac{\exp(-\lambda)\lambda^k\left(1-\frac{\exp(-\lambda)\lambda^k}{k!}\right)}{\Big(p_2(k!) + p_3\exp(-\lambda)\lambda^k\Big)}$$

$$+ \frac{1}{p_3}\left(1-\exp(-\lambda) - \frac{\exp(-\lambda)\lambda^k}{k!}\right),$$

$$I(p_2) = -E\left(\frac{\partial^2 \log(p(Y))}{\partial p_2^2}\right)$$

$$= \frac{\exp(-2\lambda)}{\Big(p_1 + p_3\exp(-\lambda)\Big)} + \frac{\left(1-\frac{\exp(-\lambda)\lambda^k}{k!}\right)^2}{\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^k}{k!}\right)}$$

$$+ \frac{1}{p_3}\Big(1-\exp(-\lambda) - \frac{\exp(-\lambda)\lambda^k}{k!}\Big),$$

$$I(p_2, \lambda) = -E\left(\frac{\partial^2 \log(p(Y))}{\partial \lambda \partial p_2}\right)$$

$$= \frac{-p_1\exp(-\lambda)}{\Big(p_1 + p_3\exp(-\lambda)\Big)} + \frac{(1-p_1)\exp(-\lambda)\lambda^k\left(\frac{k}{\lambda}-1\right)}{\Big(p_2(k!) + p_3\exp(-\lambda)\lambda^k\Big)}, \text{ and}$$

$$
\begin{aligned}
I(\lambda) &= -E\left(\frac{\partial^2 \log(p(Y))}{\partial\lambda^2}\right)\\
&= \frac{p_1 p_3 \exp(-\lambda)}{\Big(p_1 + p_3\exp(-\lambda)\Big)} + \frac{p_3}{\lambda^2}\left(1 - \exp(-\lambda) - \frac{\exp(-\lambda)\lambda^k}{k!}\right)\\
&- \frac{p_3\exp(-\lambda)\lambda^k}{\Big(p_2(k!) + p_3\exp(-\lambda)\lambda^k\Big)}\left[p_2\left(\frac{k}{\lambda} - 1\right)^2 - \frac{k}{\lambda^2}\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^k}{k!}\right)\right].
\end{aligned}
$$

Then, the Cramér-Rao lower bounds for the variances of any unbiased estimator of the parameters p_1, p_2, and λ are the diagonal elements of the inverse of the information matrix.

III.2 METHODS OF ESTIMATION FOR RAW COUNT DATA

Suppose our observations include count responses $y_i, i = 1, \ldots, n$. In this section, we will explore estimation techniques for the parameters, including maximum likelihood estimation and moment estimation for raw counts.

III.2.1 MAXIMUM LIKELIHOOD ESTIMATION

Suppose data consists of counts $y_i, i = 1, ..., n$, where $y = 0$ and $y = k$ occur more frequently in the data. The likelihood function is

$$
\begin{aligned}
L(p_1, p_2, \lambda|\boldsymbol{y}) = &\prod_{\{i:y_i=0\}}\Big(p_1 + p_3\exp(-\lambda)\lambda^{y_i}\Big)\prod_{\{i:y_i=k\}}\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)\\
&\prod_{\{i:{y_i=y \atop y\neq 0,k}\}}\left(p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right).
\end{aligned}
$$

Then, the log-likelihood function is given by

$$
\begin{aligned}
\ell(p_1, p_2, \lambda|\boldsymbol{y}) = &\sum_{\{i:y_i=0\}}\log\Big(p_1 + p_3\exp(-\lambda)\lambda^{y_i}\Big) + \sum_{\{i:y_i=k\}}\log\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)\\
&+ \sum_{\{i:{y_i=y \atop y\neq 0,k}\}}\Big(\log(p_3) - \lambda + y_i\log\lambda - \log(y_i!)\Big).
\end{aligned} \tag{32}
$$

The first-order partial derivatives of the log-likelihood function with respect to the parameters p_1, p_2, and λ, are

$$\begin{aligned}
\frac{\partial \ell(p_1,p_2,\lambda)}{\partial p_1} &= \sum_{\{i:y_i=0\}} \frac{\left(1-\exp(-\lambda)\lambda^{y_i}\right)}{\left(p_1+p_3\exp(-\lambda)\lambda^{y_i}\right)} - \sum_{\{i\ y_i=k\}} \frac{\exp(-\lambda)\lambda^{y_i}}{\left(p_2(y_i!)+p_3\exp(-\lambda)\lambda^{y_i}\right)} \\
&\quad - \frac{1}{p_3} \sum_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} 1, \\
\frac{\partial \ell(p_1,p_2,\lambda)}{\partial p_2} &= \sum_{\{i:y_i=0\}} \frac{-\exp(-\lambda)\lambda^{y_i}}{\left(p_1+p_3\exp(-\lambda)\lambda^{y_i}\right)} - \sum_{\{i:y_i=k\}} \frac{(y_i!-\exp(-\lambda)\lambda^{y_i})}{(p_2 y_i!+p_3\exp(-\lambda)\lambda^{y_i})} \\
&\quad - \frac{1}{p_3} \sum_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} 1, \text{ and} \\
\frac{\partial \ell(p_1,p_2,\lambda)}{\partial \lambda} &= \sum_{\{i:y_i=0\}} \frac{\left(-p_3\exp(-\lambda)\lambda^{y_i}\right)}{\left(p_1+p_3\exp(-\lambda)\right)} - \sum_{\{i.y_i=k\}} \frac{p_3\exp(-\lambda)\lambda^{y_i}\left(\frac{y_i}{\lambda}-1\right)}{\left(p_2(y_i!)+p_3\exp(-\lambda)\lambda^{y_i}\right)} \\
&\quad + \sum_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} \left(\frac{y_i}{\lambda}-1\right).
\end{aligned}$$

To obtain the maximum likelihood estimates for the parameters p_1, p_2, and λ, we solve the score functions simultaneously by setting them equal to zero. Here, the constraints on the parameter estimates are

$$\begin{aligned}
0 \leq p_1, p_1 &\leq 1, \\
p_1 + p_2 &\leq 1, \text{ and} \\
\lambda &> 0.
\end{aligned} \tag{33}$$

We can use numerical algorithms such as Newton-Rhapson algorithm to find solutions to the parameter estimates, $\widehat{p_1}, \widehat{p_2}$, and $\widehat{\lambda}$. Taking second order partial derivatives will gives us the following Hessian matrix:

$$\boldsymbol{H} = \begin{pmatrix}
\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial {p_1}^2} & \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_2 \partial p_1} & \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda \partial p_1} \\
\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_1 \partial p_2} & \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial {p_2}^2} & \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda \partial p_2} \\
\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_1 \partial \lambda} & \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_2 \partial \lambda} & \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda^2}
\end{pmatrix} \tag{34}$$

where

$$\frac{\partial^2 \ell}{\partial {p_1}^2} = \sum_{\{i:y_i=0\}} \frac{-\Big(1-\exp(-\lambda)\lambda^{y_i}\Big)^2}{\Big(p_1 + p_3\exp(-\lambda)\lambda^{y_i}\Big)^2} - \sum_{\{i:y_i=k\}} \frac{\exp(-2\lambda)\lambda^{2y_i}}{\Big(p_2(y_i!) + p_3\exp(-\lambda)\lambda^{y_i}\Big)^2}$$
$$-\left(\frac{1}{p_3^2}\right) \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} 1,$$

$$\begin{aligned}\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_2 \partial p_1} &= \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_1 \partial p_2} \\ &= \sum_{\{i:y_i=0\}} \frac{\Big(1-\exp(-\lambda)\Big)\exp(-\lambda)\lambda^{y_i}}{\Big(p_1+p_3\exp(-\lambda)\lambda^{y_i}\Big)^2} \\ &\quad - \sum_{\{i:y_i=k\}} \frac{\Big(1-\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\Big)\Big(\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\Big)}{\Big(p_2+p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\Big)^2} - \left(\frac{1}{p_3^2}\right)\sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} 1,\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda \partial p_1} &= \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_1 \partial \lambda} \\ &= \sum_{\{i:y_i=0\}} \frac{\exp(-\lambda)\lambda^{y_i}(p_1+p_3)}{\Big(p_1+p_3\exp(-\lambda)\lambda^{y_i}\Big)^2} - \sum_{\{i:y_i=k\}} \frac{p_2\left(\frac{y_i}{\lambda}-1\right)\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}}{\left(p_2+p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)^2},\end{aligned}$$

$$\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial {p_2}^2} = \sum_{\{i:y_i=0\}} \frac{\exp(-2\lambda)\lambda^{y_i}}{\Big(p_1+p_3\exp(-\lambda)\lambda^{y_i}\Big)^2} - \sum_{\{i:y_i=k\}} \frac{\left(1-\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)^2}{\left(p_2+p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)^2}$$
$$-\left(\frac{1}{p_3^2}\right) \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} 1,$$

$$\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial\lambda\partial p_2} = \frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_2 \partial\lambda}$$

$$= \sum_{\{i:y_i=0\}} \frac{p_1 \exp(-\lambda)\lambda^{y_i}}{\left(p_1 + p_3 \exp(-\lambda)\lambda^{y_i}\right)^2} - \sum_{\{i:y_i=k\}} \frac{(1-p_1)\left(\frac{y_i}{\lambda}-1\right)\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}}{\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)^2}$$

and

$$\frac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial\lambda^2} = -\sum_{\{i:y_i=0\}} \frac{p_1 p_3 \exp(-\lambda)\lambda^{y_i}}{\left(p_1 + p_3 \exp(-\lambda)\lambda^{y_i}\right)^2} - \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} \left(\frac{y_i}{\lambda^2}\right)$$
$$+ \sum_{\{i:y_i=k\}} \frac{p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}}{\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)^2}\left(p^2\left(\frac{y_i}{\lambda}-1\right)^2\right.$$
$$\left.-\frac{y_i}{\lambda^2}\left(p_2 + p_3\frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}\right)\right).$$

Thus, for n independent observations of raw count data, the observed Hessian matrix is obtained by evaluating Hessian matrix (34) at the maximum likelihood estimates. The observed covariance matrix is the inverse of the negative Hessian matrix given above, where the diagonal elements are the variances $\sigma^2(\widehat{p_1}), \sigma^2(\widehat{p_2})$, and $\sigma^2(\widehat{\lambda})$ of the parameters p_1, p_2, and λ, respectively. The off-diagonal elements will give us the covariances between the pairs of parameters evaluated at ML estimates.

III.2.2 MOMENT ESTIMATION

Suppose n observations consisting of raw counts that are independently distributed as DIP(p_1, p_2, λ). Our goal is to find the moment estimators for the parameters p_1, p_2, and λ. As established in Section III.1.1, the first three population moments are

$$\mu_1 = E(Y) = p_2 k + p_3\lambda,$$
$$\mu_2 = E(Y^2) = p_2 k^2 + p_3\left(\lambda + \lambda^2\right), \text{ and}$$
$$\mu_3 = E(Y^3) = p_2 k^3 + p_3\left(\lambda + 3\lambda^2 + \lambda^3\right).$$

Equating the first three population moments to the first three sample moments, $\overline{y_1}, \overline{y_2}$, and $\overline{y_3}$ where, $\overline{y_1} = (\sum_{i=1}^{n} y_i)/n$, $\overline{y_2} = (\sum_{i=1}^{n} y_i^2)/n$, and $\overline{y_3} = (\sum_{i=1}^{n} y_i^3)/n$, we obtain the following set of equations:

$$\begin{aligned}
\overline{y_1} &= p_2 k + p_3 \lambda \\
\overline{y_2} &= p_2 k^2 + p_3 \left(\lambda + \lambda^2\right) \\
\overline{y_3} &= p_2 k^3 + p_3 \left(\lambda + 3\lambda^2 + \lambda^3\right).
\end{aligned} \tag{36}$$

The set of equations (36) can be solved simultaneously to obtain the moment estimators $(\widetilde{p_1}, \widetilde{p_2}, \widetilde{\lambda})$ of the parameters p_1, p_2, and λ for a known k. Here, the parameter constraints are the same as those given by equation (33).

Let $\boldsymbol{D}$ be the matrix of first order partial derivatives of the first three moments with respect to the three parameters. Thus,

$$\boldsymbol{D} = \begin{pmatrix} \frac{\partial \mu_1}{\partial p_1} & \frac{\partial \mu_1}{\partial p_2} & \frac{\partial \mu_1}{\partial \lambda} \\ \frac{\partial \mu_2}{\partial p_1} & \frac{\partial \mu_2}{\partial p_2} & \frac{\partial \mu_2}{\partial \lambda} \\ \frac{\partial \mu_3}{\partial p_1} & \frac{\partial \mu_3}{\partial p_2} & \frac{\partial \mu_3}{\partial \lambda} \end{pmatrix}.$$

That is,

$$\boldsymbol{D} = \begin{pmatrix} -\lambda & k - \lambda & p_3 \\ -\lambda(\lambda+1) & k^2 - \lambda(\lambda+1) & k^2 - \lambda(\lambda+1) \\ -(\lambda^3 + 3\lambda^2 + \lambda) & k^3 - (\lambda^3 + 3\lambda^2 + \lambda) & p_3(3\lambda^2 + 6\lambda + 1) \end{pmatrix} \tag{37}$$

Note $p_3 = 1 - p_1 - p_2$. Let $\boldsymbol{\Sigma}$ be the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} Var(Y) & Cov(Y, Y^2) & Cov(Y, Y^3) \\ Cov(Y^2, Y) & Var(Y^2) & Cov(Y^2, Y^3) \\ Cov(Y^3, Y) & Cov(Y^3, Y^2) & Var(Y^3) \end{pmatrix}, \tag{38}$$

where,

$$
\begin{aligned}
Var(Y) &= p_2k^2 + p_3\left(\lambda^2+\lambda\right) - \left(p_2k + p_3\lambda\right)^2, \\
Var(Y^2) &= p_2k^4 + p_3\left(\lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda\right) - \left(p_2k^2 + p_3\lambda^2 + p_3\lambda\right)^2, \\
Var(Y^3) &= p_2k^6 + p_3\left(\lambda^6 + 15\lambda^5 + 65\lambda^4 + 90\lambda^3 + 31\lambda^2 + \lambda\right) \\
&\quad - \left(p_2k^3 + p_3\lambda^3 + 3p_3\lambda^2 + p_3\lambda\right)^2, \\
Cov(Y, Y^2) &= Cov(Y^2, Y) \\
&= p_2k^3 + p_3(\lambda^3 + 3\lambda^2 + \lambda) - \Big[\left(p_2k + p_3\lambda\right)\left(p_2k^2 + p_3\lambda^2 + p_3\lambda\right)\Big], \\
Cov(Y, Y^3) &= Cov(Y^3, Y) \\
&= p_2k^4 + p_3\left(\lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda\right) \\
&\quad - \Big[\left(p_2k + p_3\lambda\right)\left(p_2k^3 + p_3\lambda^3 + 3p_3\lambda^2 + p_3\lambda\right)\Big], \text{ and} \\
Cov(Y^2, Y^3) &= Cov(Y^3, Y^2) \\
&= p_2k^5 + p_3\left(\lambda^5 + 10\lambda^4 + 25\lambda^3 + 15\lambda^2 + \lambda\right) \\
&\quad - \Big[\left(p_2k^2 + p_3\lambda^2 + p_3\lambda\right)\left(p_2k^3 + p_3\lambda^3 + 3p_3\lambda^2 + p_3\lambda\right)\Big].
\end{aligned}
$$

The asymptotic covariance matrix of the moments estimators $(\widetilde{p_1}, \widetilde{p_2}, \widetilde{\lambda})$ is given by the inverse Godambe's information matrix $\boldsymbol{A} = \dfrac{1}{n}\left(\boldsymbol{D}\right)^{-1}\Sigma\left(\boldsymbol{D}^{\mathsf{T}}\right)^{-1}$ (Chaganty and Shi 2004). The diagonal elements of covariance matrix $\boldsymbol{A}$ are the asymptotic variances of the moment estimators $\widetilde{p_1}$, $\widetilde{p_2}$, and $\widetilde{\lambda}$, respectively.

III.3 METHODS OF ESTIMATION FOR GROUPED DATA

In this section, we consider obtaining parameter estimates using maximum likelihood estimation as well as moment estimation techniques for grouped data. As described in Section II.4, grouped data consists of frequencies n_j of count $y_i = j, i = 1, \ldots, n$, and $j = 0, \ldots, m$, where counts 0's and k's are highly abundant.

III.3.1 MAXIMUM LIKELIHOOD ESTIMATION

Under the assumption that n independent observations of frequencies n_j of count $y_i = j$ are distributed as DIP (p_1, p_2, λ), the likelihood function is:

$$L(p_1, p_2, \lambda | n_j) = \Big(p_1 + p_3 \exp(-\lambda)\Big)^{n_0} \Big(p_2 + p_3 \frac{\exp(-\lambda)\lambda^k}{k!}\Big)^{n_k} \prod_{\substack{j=1 \\ j \neq k}}^{\infty} \Big(p_3 \frac{\exp(-\lambda)\lambda^j}{j!}\Big)^{n_j}.$$

The log-likelihood function is given by

$$\begin{aligned} \ell(p_1, p_2, \lambda | \boldsymbol{n}_j) = & \, n_0 \log \Big(p_1 + p_3 \exp(-\lambda)\Big) \\ & + n_k \log \Big(p_2 + p_3 \frac{\exp(-\lambda)\lambda^k}{k!}\Big) \\ & + \sum_{\substack{j=1 \\ j \neq k}}^{\infty} n_j \Big(\log(p_3) - \lambda + j \log(\lambda) - \log(j!) \Big). \end{aligned} \tag{39}$$

The score functions are

$$\frac{\partial \ell}{\partial p_1} = \frac{n_0 \Big(1 - \exp(-\lambda)\Big)}{\Big(p_1 + p_3 \exp(-\lambda)\Big)} - \frac{n_k \exp(-\lambda)\lambda^k}{\Big(p_2(k!) + p_3 \exp(-\lambda)\lambda^k\Big)} - \sum_{\substack{j=1 \\ j \neq 0,k}}^{\infty} n_j \left(\frac{1}{p_3}\right),$$

$$\frac{\partial \ell}{\partial p_2} = \frac{-n_0 \exp(-\lambda)}{\Big(p_1 + p_3 \exp(-\lambda)\Big)} + \frac{n_k \left(1 - \frac{\exp(-\lambda)\lambda^k}{k!}\right)}{\left(p_2 + p_3 \frac{\exp(-\lambda)\lambda^k}{k!}\right)} - \sum_{\substack{j=1 \\ j \neq 0,k}}^{\infty} n_j \left(\frac{1}{p_3}\right), \text{ and}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} = & \frac{-n_0 p_3 \exp(-\lambda)}{\Big(p_1 + p_3 \exp(-\lambda)\Big)} + \frac{n_k p_3 \exp(-\lambda)\lambda^k}{\Big(p_2 + p_3 \exp(-\lambda)\lambda^k\Big)} \left(\frac{k}{\lambda} - 1\right) \\ & + \sum_{\substack{j=1 \\ j \neq 0,k}}^{\infty} n_j \left(\frac{j}{\lambda} - 1\right). \end{aligned}$$

In order to solve for the maximum likelihood estimates of p_1, p_2, and λ, we solve the score equations by setting them equal to zero, simultaneously with the same constraints on parameters as those given by equation (33). Since the equations are not of the closed form, one can use Newton-Rhapson algorithm to find ML estimates for the parameters.

The Hessian matrix is

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial {p_1}^2} & \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_2 \partial p_1} & \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda \partial p_1} \\ \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_1 \partial p_2} & \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial {p_2}^2} & \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda \partial p_2} \\ \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_1 \partial \lambda} & \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial p_2 \partial \lambda} & \dfrac{\partial^2 \ell(p_1,p_2,\lambda)}{\partial \lambda^2} \end{pmatrix}, \tag{40}$$

where

$$\frac{\partial^2 \ell}{\partial {p_1}^2} = \frac{-n_0\Big(1-\exp(-\lambda)\Big)^2}{\Big(p_1 + p_3\exp(-\lambda)\Big)^2} - \frac{n_k \exp(-2\lambda)\lambda^{2k}}{\Big(p_2(k!) + p_3\exp(-\lambda)\lambda^k\Big)^2} - \sum_{\substack{j=1 \\ j\neq 0,k}}^{\infty} \frac{n_j}{p_3^2},$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial p_2 \partial p_1} &= \frac{\partial^2 \ell}{\partial p_1 \partial p_2} \\ &= \frac{n_0 \exp(-\lambda)\Big(1-\exp(-\lambda)\Big)}{\Big(p_1+p_3\exp(-\lambda)\Big)^2} + \left[\frac{n_k \exp(-\lambda)\lambda^k}{(k!)\left(p_2 + p_3\dfrac{\exp(-\lambda)\lambda^k}{k!}\right)^2}\right. \\ &\quad \left.\left(1 - \frac{\exp(-\lambda)\lambda^k}{k!}\right)\right] - \sum_{\substack{j=1 \\ j\neq 0,k}}^{\infty} \frac{n_j}{p_3^2},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \lambda \partial p_1} &= \frac{\partial^2 \ell}{\partial p_1 \partial \lambda} \\ &= \frac{n_0 \exp(-\lambda)\,(p_1+p_3)}{\Big(p_1+p_3\exp(-\lambda)\Big)^2} - \frac{n_k p_2 \exp(-\lambda)\lambda^k}{(k!)\left(p_2 + p_3\dfrac{\exp(-\lambda)\lambda^k}{k!}\right)^2}\left(\frac{k}{\lambda} - 1\right),\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial {p_2}^2} = \frac{n_0 \exp(-2\lambda)}{\Big(p_1 + p_3\exp(-\lambda)\Big)^2} - \frac{n_k\Big(1-\exp(-\lambda)\lambda^k\Big)^2}{\Big(p_2(k!) + p_3\exp(-\lambda)\lambda^k\Big)^2} - \sum_{\substack{j=1 \\ j\neq 0,k}}^{\infty} \frac{n_j}{p_3^2},$$

$$\frac{\partial^2 \ell}{\partial \lambda \partial p_2} = \frac{\partial^2 \ell}{\partial p_2 \partial \lambda}$$
$$= \frac{n_0 p_1 \exp(-\lambda)}{\left(p_1 + p_3 \exp(-\lambda)\right)^2} - \frac{n_k (1-p_1)\exp(-\lambda)\lambda^k}{(k!)\left(p_2 + p_3 \frac{\exp(-\lambda)\lambda^k}{k!}\right)^2}\left(\frac{k}{\lambda} - 1\right), \text{ and}$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = \frac{-n_0 p_1 p_3 \exp(-\lambda)}{\left(p_1 + p_3 \exp(-\lambda)\right)^2} + \frac{n_k p_3 \exp(-\lambda)\lambda^k}{(k!)\left(p_2 + p_3 \frac{\exp(-\lambda)\lambda^k}{k!}\right)^2} \times \left[p^2\left(\frac{k}{\lambda} - 1\right)^2\right.$$
$$\left. - \frac{k}{\lambda^2}\left(p_2 + p_3 \frac{\exp(-\lambda)\lambda^k}{k!}\right)\right] - \sum_{\substack{j=1 \\ j \neq 0,k}}^{\infty} n_j \left(\frac{j}{\lambda^2}\right).$$

Thus, for data that consists of n_j frequency for count j, the observed Hessian matrix can be obtained by evaluating (40) at the maximum likelihood estimates $(\widehat{p_1}, \widehat{p_2}, \widehat{\lambda})$. The covariance matrix can be obtained by taking the inverse of observed information matrix which is negative of the observed Hessian matrix. The standard errors of the ML estimates can be computed from the observed variance matrix.

III.3.2 MOMENT ESTIMATION

Suppose that our data consists of n independent observations of frequencies n_j of count j, $j = 0, \ldots, m$. We assume that the frequencies n_0 and n_k are inflated frequencies of counts 0's and k's. Our goal here is to obtain the moment estimates for the parameters p_1, p_2, and λ under the assumption that our observations are distributed as DIP (p_1, p_2, λ). Since the observations are in grouped form, we first compute the sample moments $\overline{n_1}, \overline{n_2}$, and $\overline{n_3}$, where $\overline{n_i} = (\sum_{j=0}^{m} j^i n_j)/n$. The sum of n_j's is the total number of observations. Equating the sample moments to those of the population moments as given by equation (27), we obtain the following

$$\begin{aligned} \overline{n_1} &= p_2 k + p_3 \lambda \\ \overline{n_2} &= p_2 k^2 + p_3 \left(\lambda + \lambda^2\right) \\ \overline{n_3} &= p_2 k^3 + p_3 \left(\lambda + 3\lambda^2 + \lambda^3\right). \end{aligned} \tag{42}$$

To obtain the moment estimators $\widetilde{p_1}$, $\widetilde{p_2}$, and $\widetilde{\lambda}$, we solve equation (42) by numerical algorithms since the solution is not in a closed form. The parameter constraints are the same as those given by (33).

Let $\boldsymbol{D}$ be the matrix of first order partial derivatives of the first three moments with respect to the three parameters as given by equation (37). And, let Σ be the covariance matrix as represented in (38). The asymptotic covariance matrix of the method of moments estimates of p_1, p_2, and λ is given by the inverse Godambe's information matrix, that is $\boldsymbol{A} = \frac{1}{n} \cdot (\boldsymbol{D})^{-1} \Sigma (\boldsymbol{D}^{\intercal})^{-1}$ (Chaganty and Shi 2004). The diagonal elements of $\boldsymbol{A}$ will give us the asymptotic variances for the moment estimator of the parameters p_1, p_2, and λ.

III.3.3 ASYMPTOTIC RELATIVE EFFICIENCY COMPARISONS

The performance of the two estimators, maximum likelihood and moment, can be compared by calculating the asymptotic variance of the estimators. The asymptotic relative efficiency (ARE) of moment estimators $\widetilde{p_1}$, $\widetilde{p_2}$, and $\widetilde{\lambda}$ with respect to the maximum likelihood estimators $\widehat{p_1}$, $\widehat{p_2}$, and $\widehat{\lambda}$ are the ratios of their respective asymptotic variances. Thus, we calculate $e(\widetilde{p_1}, \widehat{p_1})$ by taking the ratio $\sigma^2(\widehat{p_1})/\sigma^2(\widetilde{p_1})$, $e(\widetilde{p_2}, \widehat{p_2})$ by taking the ratio $\sigma^2(\widehat{p_2})/\sigma^2(\widetilde{p_2})$, and $e(\widetilde{\lambda}, \widehat{\lambda})$ by taking the ratio $\sigma^2(\widehat{\lambda})/\sigma^2(\widetilde{\lambda})$. If the efficiencies are less than 1, then we conclude that moment estimates are less efficient than the maximum likelihood estimates and vice versa if the efficiencies are greater than 1.

First, we calculated the efficiencies for the DIP (p_1, p_2, λ) model with inflated 0's and 3's. Table 7 lists efficiencies $e(\widetilde{p_1}, \widehat{p_1})$ for the parameter p_1 for various values of p_1 and p_2 with a fixed $\lambda = 4$. Table 8 lists efficiencies $e(\widetilde{p_2}, \widehat{p_2})$ for the parameter p_2 for various values of p_1 and p_2 for a fixed $\lambda = 4$. Table 9 lists efficiencies $e(\widetilde{\lambda}, \widehat{\lambda})$ for the parameter λ for various values of p_1 and p_2 for a fixed $\lambda = 4$. Figure 5 which plots the relative efficiencies $e(\widetilde{p_1}, \widehat{p_1})$ shows that efficiencies are less than 1. Figure 6 which plots the relative efficiencies $e(\widetilde{p_2}, \widehat{p_2})$ shows that efficiencies are less than 1. The relative efficiencies for other values of λ also show similar results. Next, we calculated the efficiencies for the DIP (p_1, p_2, λ) model with inflated 0's and 14's. Table 10 lists efficiencies $e(\widetilde{p_1}, \widehat{p_1})$ for the parameter p_1 for various values of p_1 and p_2 with a fixed $\lambda = 7$. Table 11 lists efficiencies $e(\widetilde{p_2}, \widehat{p_2})$ for the parameter p_2 for various values of p_1 and p_2 for a fixed $\lambda = 7$. Table 12 lists efficiencies $e(\widetilde{\lambda},\ \widehat{\lambda})$ for the parameter λ for various values of p_1 and p_2 for a fixed $\lambda = 7$. Figure 7 which plots the relative efficiencies $e(\widetilde{p_1},\ \widehat{p_1})$ shows that efficiencies are less than 1. Figure 8 which plots the relative efficiencies $e(\widetilde{p_2},\ \widehat{p_2})$ shows that efficiencies are less than 1. We can conclude that the ML estimators for the parameters are far more efficient than the moment estimators since the efficiencies are less than 1.

Table 7: Relative Efficiencies of p_1 for DIP (p_1, p_2, λ) Model with $\lambda = 4$

		p_2				
λ	p_1	0.1	0.3	0.5	0.7	0.9
4	0.1	0.0369	0.0456	0.0593	0.0839	
	0.2	0.0738	0.0952	0.1310	0.2012	
	0.3	0.1186	0.1588	0.2303		
	0.4	0.1740	0.2429	0.3745		
	0.5	0.2445	0.3585			
	0.6	0.3371	0.5247			
	0.7	0.4637				
	0.8	0.6469				
	0.9					

NOTE: The efficiencies are calculated for DIP (p_1, p_2, λ) model with inflated 0's and 3's.

III.4 ANALYSIS OF LENGTH OF STAY DATA

The patients' length of stay data, as discussed in Section II.1, shows that the observations of 261 patients includes inflated frequencies of counts 0 and inflated frequencies of count 3's. Under the assumption that the parameters are not a function of covariates, we fit Lambert's zero-inflated Poisson (ZIP) to the observed data. The maximum likelihood estimation for the parameters p and λ was done using PROC IML in SAS and results verified using PROC NLP and PROC NLMIXED. Table 13 shows the expected frequencies due to ML estimation using ZIP for parameters p and λ. As one can note, the ZIP model, in this case, accommodates the first inflation at LOS$=0$ fairly well, however, fails to accommodate the second inflation at LOS$=3$. Thus, we obtained the maximum likelihood estimates for DIP (p, λ) model. Using

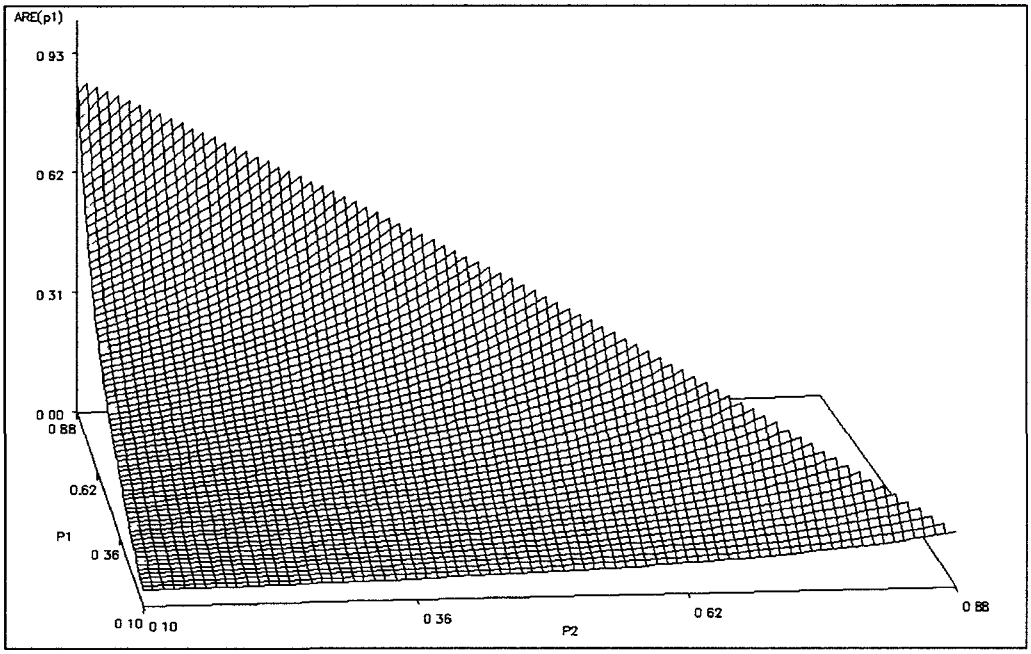

Figure 5: Plot of $e(\widetilde{p_1}, \widehat{p_1})$ for DIP (p_1, p_2, λ) for $\lambda = 4$

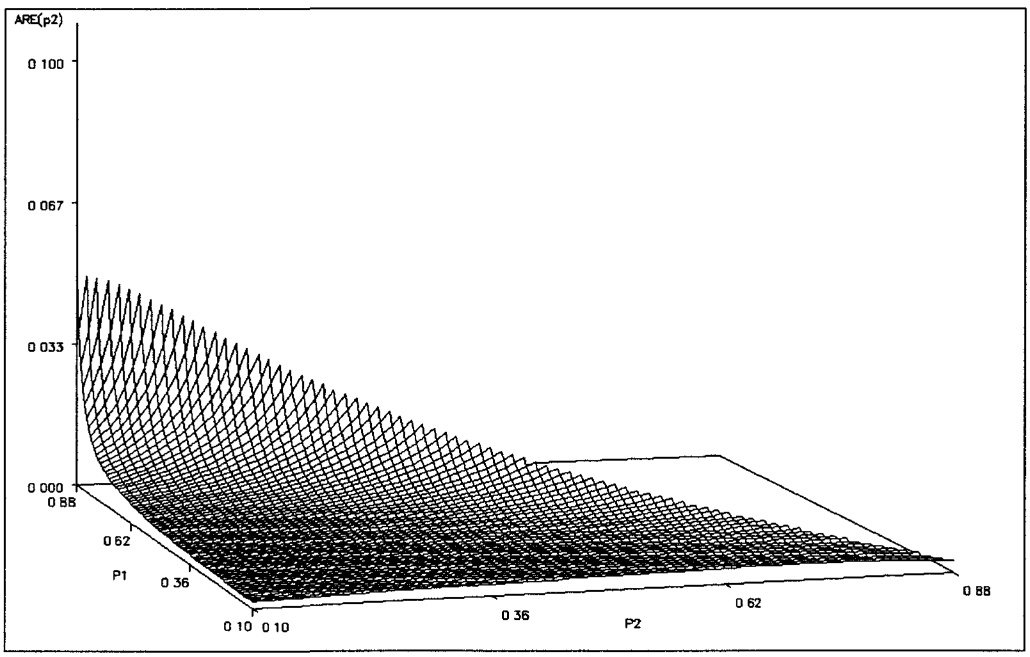

Figure 6: Plot of $e(\widetilde{p_2}, \widehat{p_2})$ for DIP (p_1, p_2, λ) for $\lambda = 4$

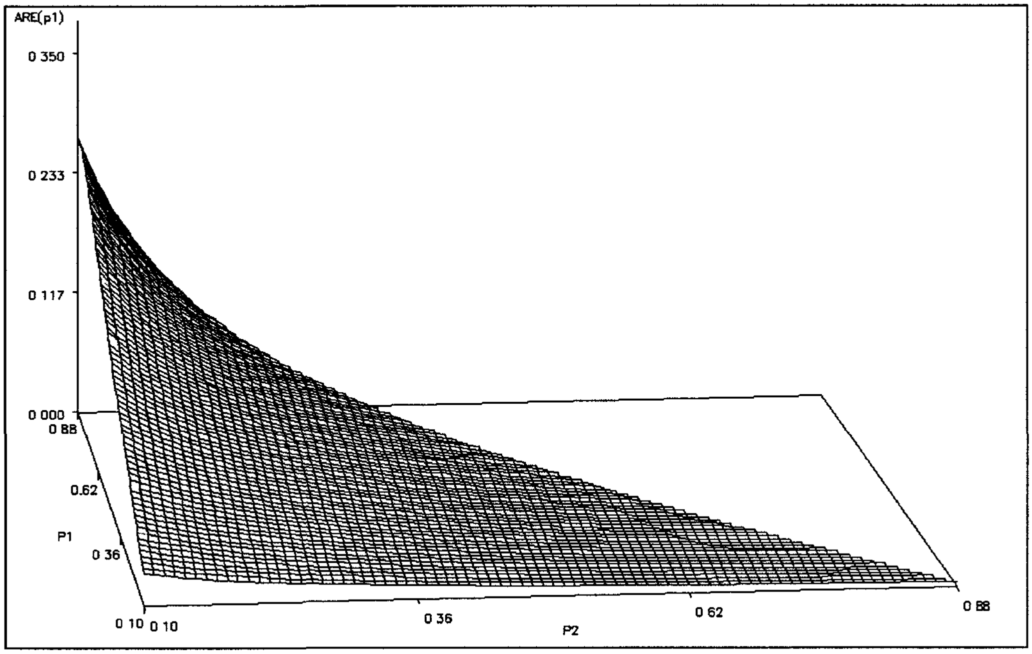

Figure 7: Plot of $e(\widetilde{p_1}, \widehat{p_1})$ for DIP (p_1, p_2, λ) for $\lambda = 7$

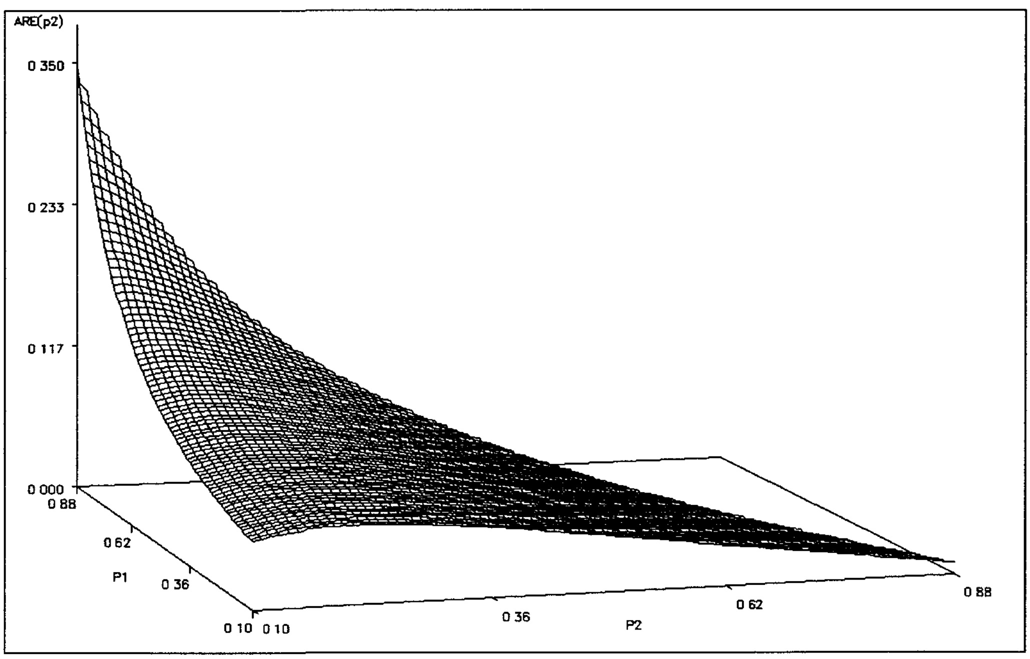

Figure 8: Plot of $e(\widetilde{p_2}, \widehat{p_2})$ for DIP (p_1, p_2, λ) for $\lambda = 7$

Table 8: Relative Efficiencies of p_2 for DIP (p_1, p_2, λ) Model with $\lambda = 4$

		p_2				
λ	p_1	0.1	0.3	0.5	0.7	0.9
4	0.1	0.0015	0.0023	0.0030	0.0034	0.0033
	0.2	0.0017	0.0028	0.0039	0.0050	
	0.3	0.0020	0.0035	0.0055	0.0068	
	0.4	0.0024	0.0047	0.0089		
	0.5	0.0030	0.0070	0.0110		
	0.6	0.0041	0.0128			
	0.7	0.0062	0.0138			
	0.8	0.0124				
	0.9	0.0051				

NOTE: The efficiencies are calculated for DIP (p_1, p_2, λ) model with inflated 0's and 3's.

PROC IML in SAS, we get the ML estimates $\widehat{p}$ and $\widehat{\lambda}$ as 0.244 and 3.710, respectively. Using these ML estimates, the expected frequencies can be obtained as seen in Table 13. As discussed in Section II.5, the expected frequencies due to DIP (p, λ) model accounts for the inflation at LOS = 3, but not as well for the LOS = 0. To investigate if DIP (p_1, p_2, λ) is an improvement for the observed data, we compute the ML estimates for the parameters using this model. The ML estimates $\widehat{p_1}, \widehat{p_2}$, and $\widehat{\lambda}$ are 0.166, 0.097, and 3.707, respectively by fitting DIP (p_1, p_2, λ) model to the observed data. We note that the expected frequencies due to DIP (p_1, p_2, λ) MLE not only accommodates the inflation at zero as well as 3 counts but as fairly for the other counts as well. Figure 9 which shows the observed frequencies and the expected frequencies due to all three models also indicates that the DIP (p_1, p_2, λ) is a better fit. Table 14 summarizes the parameter estimates obtained by ML estimation for all the three models. All the parameter estimates are significant at a 0.01 significance level. Also, DIP (p_1, p_2, λ) model has a significantly improved log-likelihood value as compared to the others. The asymptotic variances for ML estimates using the score functions were found to be $\widehat{\sigma^2(\widehat{p_1})} = 0.004$, $\widehat{\sigma^2(\widehat{p_2})} = 0.001$ and $\widehat{\sigma^2(\widehat{\lambda})} = 0.017$, and the asymptotic covariance of the two parameter estimates was approximately zero. Here, Pearson chi-square and goodness-of-fit tests are unreliable as measures of lack of fit since many observed total cell frequency counts are small, several equaling zero.

Table 9: Relative Efficiencies of λ for DIP (p_1, p_2, λ) Model with $\lambda = 4$

		p_2				
λ	p_1	0.1	0.3	0.5	0.7	0.9
4	0.1	0.0235	0.0227	0.0207	0.0158	0.0105
	0.2	0.0254	0.0245	0.0217	0.0126	
	0.3	0.0285	0.0271	0.0221	0.0156	
	0.4	0.0325	0.0303	0.0197		
	0.5	0.0381	0.0337	0.0257		
	0.6	0.0460	0.0349			
	0.7	0.0579	0.0499			
	0.8	0.0767				
	0.9	0.4578				

NOTE: The efficiencies are calculated for DIP (p_1, p_2, λ) model with inflated 0's and 3's.

Table 10: Relative Efficiencies of p_1 for DIP (p_1, p_2, λ) Model with $\lambda = 7$

		p_2				
λ	p_1	0.1	0.3	0.5	0.7	0.9
7	0.1	0.0280	0.0142	0.0087	0.0059	0.0050
	0.2	0.0577	0.0283	0.0171	0.0114	
	0.3	0.0893	0.0422	0.0249	0.0200	
	0.4	0.1229	0.0555	0.0318		
	0.5	0.1583	0.0677	0.0493		
	0.6	0.1948	0.0776			
	0.7	0.2307	0.1306			
	0.8	0.2601				
	0.9	0.0144				

NOTE: The efficiencies are calculated for DIP (p_1, p_2, λ) model with inflated 0's and 14's.

Table 11: Relative Efficiencies of p_2 for DIP (p_1, p_2, λ) Model with $\lambda = 7$

		p_2				
λ	p_1	0.1	0.3	0.5	0.7	0.9
7	0.1	0.0509	0.0562	0.0412	0.0239	0.0155
	0.2	0.0598	0.0634	0.0454	0.0259	
	0.3	0.0714	0.0721	0.0504	0.0391	
	0.4	0.0869	0.0828	0.0564		
	0.5	0.1081	0.0964	0.0780		
	0.6	0.1385	0.1138			
	0.7	0.1839	0.1664			
	0.8	0.2560				
	0.9	0.0131				

NOTE: The efficiencies are calculated for DIP (p_1, p_2, λ) model with inflated 0's and 14's.

Table 12: Relative Efficiencies of λ for DIP (p_1, p_2, λ) Model with $\lambda = 7$

		p_2				
λ	p_1	0.1	0.3	0.5	0.7	0.9
4	0.1	0.1250	0.0470	0.0192	0.0066	0.0028
	0.2	0.1312	0.0447	0.0160	0.0036	
	0.3	0.1373	0.0413	0.0119	0.0048	
	0.4	0.1432	0.0362	0.0068		
	0.5	0.1477	0.0286	0.0780		
	0.6	0.1485	0.0173			
	0.7	0.1424	0.0352			
	0.8	0.1080				
	0.9	0.0318				

NOTE: The efficiencies are calculated for DIP (p_1, p_2, λ) model with inflated 0's and 14's.

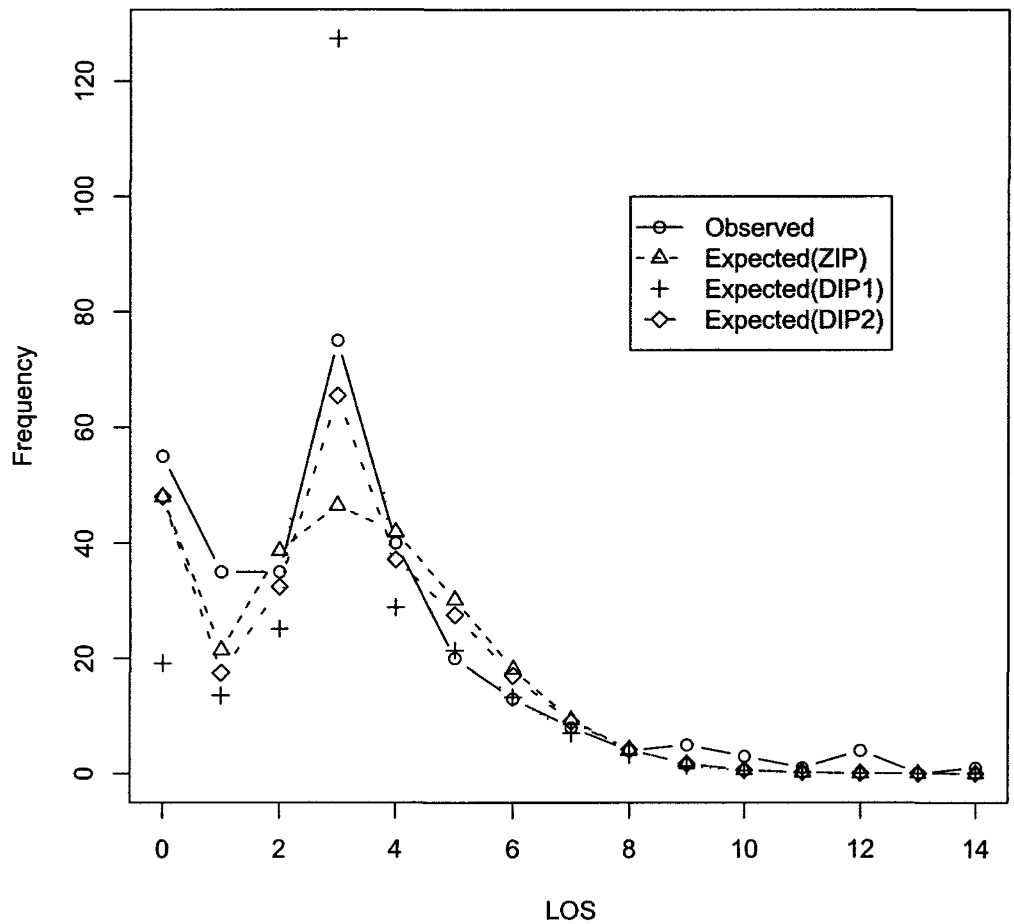

Figure 9: Observed vs Expected Frequencies for LOS Data

Table 13: Observed and Expected Frequencies for LOS Data

		Expected Frequencies		
LOS	Observed Frequency	ZIP	DIP 1	DIP 2
0	55	48	19	48
1	35	21	14	18
2	35	39	25	32
3	75	46	127	65
4	40	42	29	37
5	20	30	21	28
6	13	18	13	17
7	8	9	7	9
8	4	4	3	4
9	5	2	1	2
10	3	1	0	1
11	1	0	0	0
12	4	0	0	0
13	0	0	0	0
14	1	0	0	0

Table 14: Parameter Estimation for LOS Data

	ZIP		DIP Model 1		DIP Model 2	
Parameter	Est.(S.E.)	p-value	Est.(S.E.)	p-value	Est.(S.E.)	p-value
p	0.161(0.023)	0.001	0.244(0.025)	0.001	-	-
p_1	-	-	-	-	0.166(0.023)	0.001
p_2	-	-	-	-	0.097(0.031)	0.002
λ	3.604(0.126)	0.001	3.710(0.162)	0.001	3.707(0.140)	0.001
NegLL	666.000	-	703.955	-	660.524	-

CHAPTER IV

DOUBLY INFLATED POISSON REGRESSION MODELS

In statistical methodology, the generalized linear model (GLM) generalizes linear regression by allowing the linear model to be related to the response variable via link function(s) (McCulloch and Searle 1989). That is, it specifies how the mean of the random variable relates to the explanatory variables in the linear predictor. For count responses, a common approach is to model a Poisson log-linear model which assumes Poisson distribution for count responses and uses the log link function for its linear predictors. However, since many real-life count data may not fulfill the underlying assumptions for Poisson regression models, many researchers have considered alternatives to Poisson regression models (Coxe et al. 2009; Bae et al. 2005). Lambert (1992) illustrated ZIP regression model using analysis of data consisting of incidences of defective manufacturing items.

Numerous researchers have since then extended ZIP regression models to many applications (Böhning et al. 1997; Böhning et al. 1999; Carrivick et al. 2003; Hall and Shen 2010; Slymen et al. 2006) as well as to extensions of zero-inflated regression models (Hall 2000; Finkelman et al. 2009; Karazsia et al. 2008 Li et al. 1999; Min and Agresti 2005; Xiang et al. 2007). In this chapter, DIP regression models that mix counts of zeros, counts of k's, and Poisson counts are described in detail using logit and log-linear link functions. For binomial responses, logistic regression model that assumes binomial distribution for binary responses and uses the logit link function is quite common. Another approach is use probit link function for binary responses. However, interpretation of logistic regression models deems more useful in practice. Maximum likelihood parameter estimation for both DIP regression models are discussed in length for data consisting of raw counts as well as grouped frequencies. These methods help us investigate effects of explanatory variables on doubly inflated count response variables.

In Section IV.1, we begin with a description of dental epidemiology data to demonstrate a need for DIP regression models. Section IV.2 describes Doubly Inflated Poisson regression models for data that consists of raw counts using maximum likelihood estimation techniques. Section IV.3 describes two DIP regression models for data that consists of grouped frequencies of count responses by maximizing log-likelihood function. Section IV.4 discusses inference and testing for model parameters that helps us evaluate what factors affect the inflated counts as well as the Poisson mean. In Section IV.5, we illustrate the use of DIP regression models on the dental cavities data. Section IV.6 discusses usage of baseline-category logits for DIP (p_1, p_2, λ) regression models.

IV.1 ILLUSTRATING EXAMPLE

We will use dental epidemiology data as a motivating example in this chapter. As discussed in Section I.2.4, DMFT-index is a count response variable measuring the dental health of a person. For the $n = 1013$ children in a sample study, DMFT-index was measured at the beginning as well as the end of the study, resulting in the change of DMFT-index, $\delta DMFT$. The interested reader is pointed to Böhning et al. 1997 and Böhning et al. 1999. The line of argument followed in dental epidemiology uses the fact that the DMFT-index is a count variable, and argues that typically Poisson distributions are used for count data, finally leading to log-linear models to include covariates.

Table 15: Observed and Expected Frequencies for DMFT Data

		Expected Frequencies		
δDMFT Count	Observed Frequency	ZIP	DIP (p, λ)	DIP (p_1, p_2, λ)
0	231	231	148	231
1	379	276	535	379
2	140	251	106	140
3	116	151	96	120
4	70	69	66	77
5	55	25	36	40
6	22	8	16	17
7	0	2	6	6
8	0	0	2	2

The first and second columns of Table 15 show us the observed frequencies of the δDMFT, the counts for change in DMFT-index. It is evident that DMFT-index has improved for many children. However, we notice an extra spike in proportions of zero counts and another spike in proportions of count one. The zero count (δDMFT $= 0$) corresponds to those children showing no improvement and/or having consistent dental care. The one count (δDMFT $= 1$) corresponds to those children showing improvement in one cavity. Applying the maximum likelihood estimation methods from Chapters II and III, we can easily obtain the expected frequencies for two DIP models, DIP (p, λ) and DIP (p_1, p_2, λ) and compare them to the expected frequencies using the ML estimates for ZIP Model (see Table 15). The expected frequencies due to ML estimates for DIP (p_1, p_2, λ) are quite comparable to the observed frequencies. The parameter estimates and log-likelihood value is given in Table 16.

Table 16: Parameter Estimation for various Poisson Models

	ZIP		DIP Model 1		DIP Model 2	
Parameter	Est.(S.E.)	p-value	Est.(S.E.)	p-value	Est.(S.E.)	p-value
p	0.078(0.019)	0.001	0.344(0.017)	0.001	-	-
p_1	-	-	-	-	0.186(0.015)	0.001
p_2	-	-	-	-	0.266(0.020)	0.001
λ	1.813(0.055)	0.001	2.727(0.105)	0.001	2.566(0.098)	0.001
NegLL	1749.845	-	1740.673	-	1686.805	-

Our goal, however, is to compare the treatments as mentioned in Section I.2.4 which can be accomplished using DIP regression models. Each of the six schools of children were given one of the following treatments: oral health education, enrichment of the school diet with rice bran, mouthwash with 0.2% of sodium flouride solution, oral hygiene, all of the four treatments combined, and a control group. Other covariates such as gender and ethnicity groups (White, Black, Others consisting of predominantly Hispanic) were also considered.

IV.2 REGRESSION MODELS FOR RAW DATA

In this section, we discuss fitting DIP regression models for independent observations consisting of raw count responses, y_i, for subject i, $i = 1, \ldots, n$. Table 17 shows the general layout for the raw counts that will be used in this section. Let l be the number

Table 17: General Layout of Raw Count Data

Subject	Response	Covariates		
1	y_1	x_{11}	$\cdots$	x_{1l}
2	y_2	x_{21}	$\cdots$	x_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_n	x_{n1}	$\cdots$	x_{nl}

of covariates, and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{il})$ be the corresponding covariate vector associated with the subject i. Thus, for a particular covariate, $\boldsymbol{x}_i$, our data consists of count response y_i. Using this data layout, Section IV.2.1 discusses maximum likelihood estimation for DIP (p, λ) regression, and Section IV.2.2 discusses maximum likelihood estimation for DIP (p_1, p_2, λ) regression.

IV.2.1 DIP (p, λ) REGRESSION MODEL

Assuming that the count responses $Y_1, \ldots, Y_n$ independently distributed as Doubly Inflated Poisson with parameters (p_i, λ_i), $i = 1, \ldots, n$. We model this data using DIP (p, λ) regression model using logit and log link functions for the parameters. That is, the count responses $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ are independent and

$$Y_i = \begin{cases} 0, & \text{with probability } p_i^2 + q_i^2 \exp(-\lambda_i); \\ k, & \text{with probability } 2p_i q_i + q_i^2 \left(\dfrac{\exp(-\lambda_i)\lambda_i^k}{k!} \right); \\ y, & \text{with probability } q_i^2 \left(\dfrac{\exp(-\lambda_i)\lambda_i^y}{y!} \right), \ y = 1, 2, \ldots \neq 0, k \end{cases}$$

Moreover, the parameters p_i and λ_i satisfy

$$\operatorname{logit}(p_i) = \boldsymbol{G}_i \boldsymbol{\gamma} \text{ and } \qquad \log(\lambda_i) = \boldsymbol{B}_i \boldsymbol{\beta} \tag{43}$$

where $\boldsymbol{G}_i, \boldsymbol{B}_i$ are the ith row of design matrices $\boldsymbol{G}$ and $\boldsymbol{B}$. Also, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the regression parameters. That is,

$$p_i = \frac{\exp(\boldsymbol{G}_i \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{G}_i \boldsymbol{\gamma})} \text{ and } \qquad \lambda_i = \exp(\boldsymbol{B}_i \boldsymbol{\beta}) \tag{44}$$

Since $q_i = 1 - p_i$, $q_i = (1 + \exp(\boldsymbol{G}_i \boldsymbol{\gamma}))^{-1}$. Here, p_i and λ_i represent subject-specific binomial probability and Poisson mean, respectively.

The covariates that affect the Poisson mean λ_i may or may not be the same as the covariates that affect the probability p_i. If they are same, and if p_i and λ_i are not related, then DIP (p, λ) regression requires twice as many parameters as Poisson regression. On the contrary, when the probability p_i does not depend on the covariates, $\boldsymbol{G}_i$ is a column vector of ones, and DIP (p, λ) regression requires only one more parameter than the Poisson regression. In either case, DIP (p, λ) regression will require as many parameters as ZIP regression. If the covariates affecting p_i and λ_i are the same, then we can think of p_i as a function of λ_i thereby reducing the number of parameters required for estimation. But in most real-life data, usually prior information on how p_i and λ_i are related and how the covariates affect p_i and λ_i is usually unknown. Our goal, usually is to deduce the effects of covariates on the count response variable and to know if they affect the data behavior of two peaks.

Thus, for data consisting of raw count responses, $y_1, \ldots, y_n$, the log-likelihood function of DIP (p, λ) regression when p_i and λ_i, assumed to be unrelated, is

$$\begin{aligned} \ell(p_i, \lambda_i | \boldsymbol{y}) = & \sum_{\{i: y_i = 0\}} \log\left(p_i^2 + q_i^2 \exp(-\lambda_i)\right) + \sum_{\{i: y_i = k\}} \log\left(2 p_i q_i + q_i^2 \frac{\exp(-\lambda_i) \lambda_i^k}{k!}\right) \\ & + \sum_{\{i: \substack{y_i = y \\ y \neq 0, k}\}} \log\left(q_i^2 \frac{\exp(-\lambda_i) \lambda_i^{y_i}}{y_i!}\right), \end{aligned} \tag{45}$$

which can be further written in terms of the regression parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ as

$$\ell(\gamma,\boldsymbol{\beta}) = -2\sum_{i=1}^{n}\log\Big(1+\exp(\boldsymbol{G}_i\gamma)\Big) + \sum_{\{i\; y_i=0\}}\log\Big(\exp(2\boldsymbol{G}_i\gamma)+\exp(-\exp(\boldsymbol{B}_i\beta))\Big)$$
$$+ \sum_{\{i\; y_i=k\}}\log\Big(2\exp(\boldsymbol{G}_i\gamma)+\frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta}-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\Big)$$
$$+ \sum_{\{i\; \substack{y_i=y \\ y\neq 0,k}\}}\Big(y_i\boldsymbol{B}_i\boldsymbol{\beta}-\exp(\boldsymbol{B}_i\boldsymbol{\beta})-\log(y_i!)\Big).$$

The maximum likelihood estimate $(\widehat{\gamma},\widehat{\boldsymbol{\beta}})$ is the solution to the first-order derivatives

$$\frac{\partial\ell}{\partial\gamma}=0 \text{ and } \qquad \frac{\partial\ell}{\partial\boldsymbol{\beta}}=0,$$

where

$$\frac{\partial\ell(\gamma,\boldsymbol{\beta})}{\partial\gamma} = -2\sum_{i=1}^{n}\frac{\boldsymbol{G}_i\exp(\boldsymbol{G}_i\gamma)}{1+\exp(\boldsymbol{G}_i\gamma)} + \sum_{\{i\; y_i=0\}}\frac{2\boldsymbol{G}_i\exp(2\boldsymbol{G}_i\gamma)}{\exp(2\boldsymbol{G}_i\gamma)+\exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}$$
$$+ \sum_{\{i\; y_i=k\}}\frac{2\boldsymbol{G}_i\exp(\boldsymbol{G}_i\gamma)}{2\exp(\boldsymbol{G}_i\gamma)+\dfrac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta}-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}},$$

and

$$\frac{\partial\ell(\gamma,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{\{i\; y_i=0\}}\frac{-\boldsymbol{B}_i\exp(\boldsymbol{B}_i\boldsymbol{\beta}-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{\exp(2\boldsymbol{G}_i\gamma)+\exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}$$
$$+ \sum_{\{i\; y_i=k\}}\frac{\Big(k\boldsymbol{B}_i-\boldsymbol{B}_i\exp(\boldsymbol{B}_i\boldsymbol{\beta})\Big)\exp(k\boldsymbol{B}_i\boldsymbol{\beta}-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!\Big(2\exp(\boldsymbol{G}_i\gamma)+\dfrac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta}-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\Big)}$$
$$+ \sum_{\{i\; \substack{y_i=y \\ y\neq 0,k}\}}\Big(y_i\boldsymbol{B}_i-\boldsymbol{B}_i\exp(\boldsymbol{B}_i\boldsymbol{\beta})\Big).$$

The second order partial derivatives of the log-likelihood function are given by the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2\ell(\gamma,\boldsymbol{\beta})}{\partial\gamma^2} & \dfrac{\partial^2\ell(\gamma,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\gamma} \\ \dfrac{\partial^2\ell(\gamma,\boldsymbol{\beta})}{\partial\gamma\partial\boldsymbol{\beta}} & \dfrac{\partial^2\ell(\gamma,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^2} \end{pmatrix},$$

where

$$\frac{\partial^2 \ell(\gamma,\boldsymbol{\beta})}{\partial \gamma^2} = -2\sum_{i=1}^{n} \frac{\boldsymbol{G}_i^2 \exp(\boldsymbol{G}_i\gamma)}{\Big(1+\exp(\boldsymbol{G}_i\gamma)\Big)^2} + \sum_{\{i\ y_i=0\}} \frac{4\boldsymbol{G}_i^2 \exp(2\boldsymbol{G}_i\gamma - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{\Big(\exp(2\boldsymbol{G}_i\gamma) + \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))\Big)^2}$$
$$+ \sum_{\{i\ y_i=k\}} \frac{2\boldsymbol{G}_i^2 \exp(\boldsymbol{G}_i\gamma)\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{(k!)\left(2\exp(\boldsymbol{G}_i\gamma) + \dfrac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right)^2},$$

$$\frac{\partial^2 \ell(\gamma,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial\gamma} = \frac{\partial^2 \ell(\gamma,\boldsymbol{\beta})}{\partial \gamma\partial\boldsymbol{\beta}}$$
$$= \sum_{\{i\ y_i=0\}} \frac{2\boldsymbol{B}_i\boldsymbol{G}_i \exp(2\boldsymbol{G}_i\gamma + \boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{\Big(\exp(2\boldsymbol{G}_i\gamma) + \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))\Big)^2}$$
$$- \sum_{\{i\ y_i=k\}} \frac{2\boldsymbol{B}_i\boldsymbol{G}_i\Big(k-\exp(\boldsymbol{B}_i\boldsymbol{\beta})\Big)\exp\left(\boldsymbol{G}_i\gamma + k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta})\right)}{k!\left(2\exp(\boldsymbol{G}_i\gamma) + \dfrac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right)^2},$$

and

$$\frac{\partial^2 \ell(\gamma,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = \sum_{\{i\ y_i=0\}} \frac{-\boldsymbol{B}_i^2 \exp(\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{\Big(\exp(2\boldsymbol{G}_i\gamma) + \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))\Big)^2}$$
$$\times \Big(\exp(2\boldsymbol{G}_i\gamma) + \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta})) - \exp(2\boldsymbol{G}_i\gamma + \boldsymbol{B}_i\boldsymbol{\beta})\Big)$$
$$+ \sum_{\{i\ y_i=k\}} \frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!\left(2\exp(\boldsymbol{G}_i\gamma) + \dfrac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right)^2}$$
$$\times \Big(2\exp(\boldsymbol{G}_i\gamma)\,(k\boldsymbol{B}_i - \boldsymbol{B}_i\exp(\boldsymbol{B}_i\boldsymbol{\beta}))^2\Big)$$
$$- \sum_{\{i\ y_i=k\}} \frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!\left(2\exp(\boldsymbol{G}_i\gamma) + \dfrac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right)^2}$$
$$\times \left(\boldsymbol{B}_i^2\exp(\boldsymbol{B}_i\boldsymbol{\beta})\left(2\exp(\boldsymbol{G}_i\gamma) + \frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right)\right)$$
$$- \sum_{\left\{i\ \substack{y_i=y \\ y\neq 0,k}\right\}} \Big(\boldsymbol{B}_i^2\exp(\boldsymbol{B}_i\boldsymbol{\beta})\Big).$$

The covariance matrix can be obtained by taking the inverse of the negative Hessian matrix calculated at the ML estimates. In large samples, the MLE's $(\widehat{\gamma}, \widehat{\boldsymbol{\beta}})$ for DIP

(p, λ) regression are approximately normal with means $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ and variances to the diagonal elements of the observed covariance matrix. The estimated standard errors are the square roots of the asymptotic variances of ML estimates.

IV.2.2 DIP (p_1, p_2, λ) REGRESSION MODEL

For DIP (p_1, p_2, λ) regression model, assume that count responses $Y_1, \ldots, Y_n$ are independent and

$$Y_i = \begin{cases} 0, & \text{with probability } p_{1i} + p_{3i}\exp(-\lambda_i); \\ k, & \text{with probability } p_{2i} + p_{3i}\left(\dfrac{\exp(-\lambda_i)\lambda_i^k}{k!}\right); \\ y, & \text{with probability } p_{3i}\left(\dfrac{\exp(-\lambda_i)\lambda_i^y}{y!}\right),\ y = 1, 2, \ldots \neq 0, k \end{cases}$$

Note that the sum of the probabilities p_{1i}, p_{2i}, and p_{3i} equals 1. Also, p_{1i}, p_{2i}, and λ_i, $i = 1, \ldots, n$, be subject-specific parameters. Moreover, the probabilities p_{1i} and p_{2i} are regressed over set(s) of covariates using their respective logit link functions, and the mean parameter λ_i can be regressed over a set of covariates using log link function. That is,

$$\text{logit}(p_{1i}) = \boldsymbol{A}_i\boldsymbol{\alpha}, \qquad \text{logit}(p_{2i}) = \boldsymbol{G}_i\boldsymbol{\gamma}, \text{ and} \qquad \log(\lambda_i) = \boldsymbol{B}_i\boldsymbol{\beta}$$

Thus,

$$p_{1i} = \frac{\exp(\boldsymbol{A}_i\boldsymbol{\alpha})}{1 + \exp(\boldsymbol{A}_i\boldsymbol{\alpha})}, \qquad p_{2i} = \frac{\exp(\boldsymbol{G}_i\boldsymbol{\gamma})}{1 + \exp(\boldsymbol{G}_i\boldsymbol{\gamma})}, \text{ and} \qquad \lambda_i = \exp(\boldsymbol{B}_i\boldsymbol{\beta}) \tag{47}$$

Here, $\boldsymbol{A}_i, \boldsymbol{G}_i$, and $\boldsymbol{B}_i$ are the ith row of design matrices $\boldsymbol{A}$, $\boldsymbol{G}$, and $\boldsymbol{B}$, respectively. Also, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$ are regression parameters.

The covariates that affect the Poisson mean may or may not be the same as the covariates affecting the probabilities p_{1i} and p_{2i}. If they are same, and the Poisson mean and probabilities are not related, then DIP (p_1, p_2, λ) regression requires thrice as many parameters as an ordinary Poisson regression. On the contrary, if the probabilities are not a function of covariates, then DIP (p_1, p_2, λ) requires only two more parameters than Poisson regression. Besides, DIP (p_1, p_2, λ) regression requires one more set of covariates as compared to the covariates necessary for DIP (p, λ) and ZIP regression models.

Suppose data consists of $y_1, \ldots, y_n$ indepedently distributed as Doubly Inflated Poisson with parameters p_{1i}, p_{2i}, and λ_i, $i = 1, \ldots, n$. Under the assumption that the parameters p_{1i}, p_{2i}, and λ_i are not related, the marginal log-likelihood function can be constructed as

$$\begin{aligned}\ell(p_{1i}, p_{2i}, \lambda_i|\boldsymbol{y}) = &\sum_{\{i:y_i=0\}} \log\left[p_{1i} + p_{3i}\exp(-\lambda_i)\right] + \sum_{\{i\, y_i=k\}} \log\left[p_{2i} + p_{3i}\frac{\exp(-\lambda_i)\lambda_i^k}{k!}\right] \\ &+ \sum_{\{i:{y_i=y \atop y\neq 0,k}\}} \log\left[p_{3i}\frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}\right]. \end{aligned} \tag{48}$$

Substituting equations (47), the log-likelihood function is therefore

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \sum_{\{i:y_i=0\}} \log \ell_1 + \sum_{\{i\, y_i=k\}} \log \ell_2 + \sum_{\{i:{y_i=y \atop y\neq 0,k}\}} (\log \ell_3 + \ell_4) + \sum_{i=1}^{n} (\log \ell_5 + \log \ell_6)$$

where

$$\ell_1 = \exp(\boldsymbol{A}_i\boldsymbol{\alpha})(1 + \exp(\boldsymbol{G}_i\boldsymbol{\gamma})) + \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))\left(1 - \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma})\right),$$

$$\ell_2 = \exp(\boldsymbol{G}_i\boldsymbol{\gamma})\big(1 + \exp(\boldsymbol{A}_i\boldsymbol{\alpha})\big) + \frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\big(1 - \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma})\big),$$

$$\ell_3 = 1 - \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma}), \qquad \ell_4 = (y_i\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta})) - \log(y_i!),$$

$$\ell_5 = 1 + \exp(\boldsymbol{A}_i\boldsymbol{\alpha}), \text{ and } \qquad \ell_6 = 1 + \exp(\boldsymbol{G}_i\boldsymbol{\gamma}).$$

The DIP (p_1, p_2, λ) MLE $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ is the solution to the first-order partial derivatives

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = 0, \qquad \frac{\partial \ell}{\partial \boldsymbol{\gamma}} = 0, \text{ and } \qquad \frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0,$$

where

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = \sum_{\{i:y_i=0\}} \frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \boldsymbol{\alpha}} + \sum_{\{i:y_i=k\}} \frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \boldsymbol{\alpha}} + \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} \frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \boldsymbol{\alpha}} - \sum_{i=1}^{n} \frac{1}{\ell_5}\frac{\partial \ell_5}{\partial \boldsymbol{\alpha}},$$

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \sum_{\{i:y_i=0\}} \frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \boldsymbol{\gamma}} + \sum_{\{i:y_i=k\}} \frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \boldsymbol{\gamma}} + \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} \frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \boldsymbol{\gamma}} - \sum_{i=1}^{n} \frac{1}{\ell_6}\frac{\partial \ell_6}{\partial \boldsymbol{\gamma}}, \text{ and}$$

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{\{i:y_i=0\}} \frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \boldsymbol{\beta}} + \sum_{\{i:y_i=k\}} \frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \boldsymbol{\beta}} + \sum_{\{i:\substack{y_i=y\\ y\neq 0,k}\}} \frac{\partial \ell_4}{\partial \boldsymbol{\beta}}$$

Note

$$\frac{\partial \ell_1}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_i \left[\exp(\boldsymbol{A}_i\boldsymbol{\alpha}) + \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma})\big(1 - \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))\big)\right],$$

$$\frac{\partial \ell_2}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_i \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma})\left(1 - \frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right),$$

$$\frac{\partial \ell_3}{\partial \boldsymbol{\alpha}} = -\boldsymbol{A}_i \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma}), \qquad \frac{\partial \ell_5}{\partial \boldsymbol{\alpha}} = -\boldsymbol{A}_i \exp(\boldsymbol{A}_i\boldsymbol{\alpha}),$$

$$\frac{\partial \ell_1}{\partial \boldsymbol{\gamma}} = \boldsymbol{G}_i \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma})\big(1 - \exp(-\exp(\boldsymbol{B}_i\boldsymbol{\beta}))\big),$$

$$\frac{\partial \ell_2}{\partial \boldsymbol{\gamma}} = \boldsymbol{G}_i \left[\exp(\boldsymbol{G}_i\boldsymbol{\gamma}) + \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma})\left(1 - \frac{\exp(k\boldsymbol{B}_i\boldsymbol{\beta} - \exp(\boldsymbol{B}_i\boldsymbol{\beta}))}{k!}\right)\right],$$

$$\frac{\partial \ell_3}{\partial \boldsymbol{\gamma}} = -\boldsymbol{G}_i \exp(\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{G}_i\boldsymbol{\gamma}), \qquad \frac{\partial \ell_6}{\partial \boldsymbol{\gamma}} = \frac{-\boldsymbol{G}_i \exp(\boldsymbol{G}_i\boldsymbol{\gamma})}{1 + \exp(\boldsymbol{G}_i\boldsymbol{\gamma})},$$

$$\frac{\partial \ell_1}{\partial \boldsymbol{\beta}} = -\boldsymbol{B}_\imath \exp(\boldsymbol{B}_\imath \beta - \exp(\boldsymbol{B}_\imath \beta))\big(1 - \exp(\boldsymbol{A}_\imath \boldsymbol{\alpha} + \boldsymbol{G}_\imath \gamma)\big),$$

$$\frac{\partial \ell_2}{\partial \boldsymbol{\beta}} = (k\boldsymbol{B}_\imath - \boldsymbol{B}_\imath \exp(\boldsymbol{B}_\imath \beta))\big(1 - \exp(\boldsymbol{A}_\imath \boldsymbol{\alpha} + \boldsymbol{G}_\imath \gamma)\big)\frac{\exp(k\boldsymbol{B}_\imath \boldsymbol{\beta} - \exp(\boldsymbol{B}_\imath \boldsymbol{\beta}))}{k!}, \text{ and}$$

$$\frac{\partial \ell_4}{\partial \boldsymbol{\beta}} = y_\imath \boldsymbol{B}_\imath - \boldsymbol{B}_\imath \exp(\boldsymbol{B}_\imath \beta).$$

The second order partial derivatives of the log-likelihood function are given by the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}^2} & \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\alpha}} & \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}} \\ \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\gamma}} & \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}^2} & \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}} \\ \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}} & \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}} & \frac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \end{pmatrix},$$

where

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2} = \sum_{\{\imath\ y_\imath = 0\}} \frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{\imath\ y_\imath = k\}} \frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{\imath\ {y_\imath = y \atop y \neq 0,k}\}} \frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \boldsymbol{\alpha}}\right] - \sum_{\imath=1}^{n} \frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{\ell_5}\frac{\partial \ell_5}{\partial \boldsymbol{\alpha}}\right],$$

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\alpha}} = \sum_{\{\imath\ y_\imath = 0\}} \frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{\imath\ y_\imath = k\}} \frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{\imath\ {y_\imath = y \atop y \neq 0,k}\}} \frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \boldsymbol{\alpha}}\right] - \sum_{\imath=1}^{n} \frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{\ell_5}\frac{\partial \ell_5}{\partial \boldsymbol{\alpha}}\right],$$

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}} = \sum_{\{\imath\ y_\imath = 0\}} \frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{\imath\ y_\imath = k\}} \frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{\imath\ {y_\imath = y \atop y \neq 0,k}\}} \frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \boldsymbol{\alpha}}\right] - \sum_{\imath=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{\ell_5}\frac{\partial \ell_5}{\partial \boldsymbol{\alpha}}\right],$$

$$\frac{\partial^2 \ell}{\partial \gamma^2} = \sum_{\{i:y_i=0\}} \frac{\partial}{\partial \gamma}\left[\frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \gamma}\right] + \sum_{\{i:y_i=k\}} \frac{\partial}{\partial \gamma}\left[\frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \gamma}\right] + \sum_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} \frac{\partial}{\partial \gamma}\left[\frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \gamma}\right] - \sum_{i=1}^{n} \frac{\partial}{\partial \gamma}\left[\frac{1}{\ell_6}\frac{\partial \ell_6}{\partial \gamma}\right],$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \gamma} = \sum_{\{i:y_i=0\}} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \gamma}\right] + \sum_{\{i:y_i=k\}} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \gamma}\right] + \sum_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \gamma}\right] - \sum_{i=1}^{n} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_6}\frac{\partial \ell_6}{\partial \gamma}\right],$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = \sum_{\{i:y_i=0\}} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_1}\frac{\partial \ell_1}{\partial \beta}\right] + \sum_{\{i:y_i=k\}} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_2}\frac{\partial \ell_2}{\partial \beta}\right] + \sum_{\{i:\substack{y_i=y \\ y\neq 0,k}\}} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_3}\frac{\partial \ell_3}{\partial \beta}\right] - \sum_{i=1}^{n} \frac{\partial}{\partial \beta}\left[\frac{1}{\ell_6}\frac{\partial \ell_6}{\partial \beta}\right],$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta} = \frac{\partial^2 \ell}{\partial \beta \partial \alpha}, \qquad \frac{\partial^2 \ell}{\partial \gamma \partial \beta} = \frac{\partial^2 \ell}{\partial \beta \partial \gamma}, \text{ and } \qquad \frac{\partial^2 \ell}{\partial \alpha \partial \gamma} = \frac{\partial^2 \ell}{\partial \gamma \partial \alpha}.$$

The large sample estimated covariance matrix for the ML parameter estimates is the inverse of the observed information matrix evaluated at the ML estimates. The observed information matrix contains the negative Hessian matrix. The estimated standard errors are obtained by taking the square root of the diagonal elements of the estimated covariance matrix.

IV.3 REGRESSION MODELS FOR GROUPED DATA

In practice, sometimes the data collected are in terms of frequencies as shown in Table 18. Let u be the number of covariates, and $\boldsymbol{x}_l = (x_{l1}, \ldots, x_{lu}), l = 1, \ldots, s$ be the corresponding vector of covariates. Assume that there are s such distinct sets of vectors of covariates. Thus, for a particular lth covariate, $\boldsymbol{x}_l$, our data consists of

Table 18: General Layout of Grouped Data

Response	Covariates		
$n_{01}, n_{11}, \ldots, n_{m1}$	x_{11}	$\cdots$	x_{1u}
$n_{02}, n_{12}, \ldots, n_{m2}$	x_{21}	$\cdots$	x_{2u}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$n_{0s}, n_{1s}, \ldots, n_{ms}$	x_{s1}	$\cdots$	x_{su}

n independent observations of the frequencies $n_{0l}, n_{1l}, \ldots, n_{ml}$. Here, n_{jl} represents the frequencies of count $j, j = 0, \ldots, m$, and the sum of n_{jl}'s is the total number of observations n. That is, $n = \sum_{l=1}^{s} \sum_{j=1}^{m} n_{jl}$. We will use this layout for the grouped data in this section. Section IV.3.1 discusses the ML estimation for DIP (p, λ) regression models, and Section IV.3.2 discusses the ML estimation for DIP (p_1, p_2, λ) regression models.

IV.3.1 DIP (p, λ) REGRESSION MODEL

Suppose data consists of n observations that are independently distributed as doubly inflated Poisson counts with parameters (p_l, λ_l), $l = 1, \ldots, s$. The binomial probability p_l and mean λ_l can be parameterized by using the logit and log link functions as follows:

$$\operatorname{logit}(p_l) = \boldsymbol{G}_l \boldsymbol{\gamma} \qquad \text{and} \qquad \log(\lambda_l) = \boldsymbol{B}_l \boldsymbol{\beta} \tag{51}$$

where $\boldsymbol{G}_l, \boldsymbol{B}_l$ are lth vector of covariates; and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are regression parameters. Under the assumption the covariate-specific parameters p_l and λ_l are not related, the log-likelihood function is

$$\begin{aligned} \ell(p_l, \lambda_l | n_{jl}) = & \sum_{\{l\ j=0\}} n_{jl} \log\left(p_l^2 + q_l^2 \exp(-\lambda_l)\right) + \sum_{\{l\ j=k\}} n_{jl} \log\left(2p_l q_l + q_l^2 \frac{\exp(-\lambda_l)\lambda_l^k}{k!}\right) \\ & + \sum_{\{l\ {j=1 \atop \neq k}\}}^{m} n_{jl} \log\left(q_l^2 \frac{\exp(-\lambda_l)\lambda_l^j}{j!}\right). \end{aligned}$$

Note that $n = \sum_{l=1}^{s} \sum_{j=0}^{m} n_{jl}$. The log-likelihood function, written in terms of the regression parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, is thus

$$\begin{aligned}
\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) = & -2 \sum_{l} \sum_{j=0}^{m} n_{jl} \log\Big(1 + \exp(\boldsymbol{G}_l \gamma)\Big) \\
& + \sum_{\{l\ j=0\}} n_{jl} \log\Big(\exp(2\boldsymbol{G}_l \gamma) + \exp(-\exp(\boldsymbol{B}_l \beta))\Big) \\
& + \sum_{\{l\ j=k\}} n_{jl} \log\Big(2\exp(\boldsymbol{G}_l \gamma) + \frac{\exp(k\boldsymbol{B}_l \beta - \exp(\boldsymbol{B}_l \beta))}{k!}\Big) \\
& + \sum_{\{l\ \substack{j=1 \\ \neq k}\}}^{m} n_{jl}\Big(j\boldsymbol{B}_l \beta - \exp(\boldsymbol{B}_l \beta) - \log(j!)\Big)
\end{aligned}$$

The ML estimate $(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ is the solution to the first-order derivatives

$$\frac{\partial \ell}{\partial \boldsymbol{\gamma}} = 0 \text{ and } \qquad \frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0,$$

where

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}} = & -2 \sum_{l} \sum_{j=0}^{m} n_{jl} \frac{\boldsymbol{G}_l \exp(\boldsymbol{G}_l \gamma)}{1 + \exp(\boldsymbol{G}_l \gamma)} + \sum_{\{l\ j=0\}} n_{jl} \frac{2\boldsymbol{G}_l \exp(2\boldsymbol{G}_l \gamma)}{\exp(2\boldsymbol{G}_l \gamma) + \exp(-\exp(\boldsymbol{B}_l \beta))} \\
& + \sum_{\{l\ j=k\}} n_{jl} \frac{2\boldsymbol{G}_l \exp(\boldsymbol{G}_l \gamma)}{2\exp(\boldsymbol{G}_l \gamma) + \dfrac{\exp(k\boldsymbol{B}_l \beta - \exp(\boldsymbol{B}_l \beta))}{k!}}, \text{ and} \\
\frac{\partial \ell(\boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = & - \sum_{\{l\ j=0\}} n_j \frac{\boldsymbol{B}_l \exp(\boldsymbol{B}_l \beta - \exp(\boldsymbol{B}_l \beta))}{\exp(2\boldsymbol{G}_l \gamma) + \exp(-\exp(\boldsymbol{B}_l \beta))} \\
& + \sum_{\{l\ j=k\}} n_{jl} \frac{(k\boldsymbol{B}_l - \boldsymbol{B}_l \exp(\boldsymbol{B}_l \beta)) \exp(k\boldsymbol{B}_l \beta - \exp(\boldsymbol{B}_l \beta))}{k!\Big(2\exp(\boldsymbol{G}_l \gamma) + \dfrac{\exp(k\boldsymbol{B}_l \beta - \exp(\boldsymbol{B}_l \beta))}{k!}\Big)} \\
& + \sum_{\{l\ \substack{j=1 \\ \neq k}\}}^{m} n_{jl} \left(j\boldsymbol{B}_l - \boldsymbol{B}_l \exp(\boldsymbol{B}_l \boldsymbol{\beta})\right).
\end{aligned}$$

The second order partial derivatives of the log-likelihood function are given by the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2 \ell(\boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}^2} & \dfrac{\partial^2 \ell(\boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}} \\ \dfrac{\partial^2 \ell(\boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell(\boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \end{pmatrix},$$

where

$$\frac{\partial^2\ell(\gamma,\boldsymbol{\beta})}{\partial\gamma^2} = -2\sum_{l\ j=0}^{m} n_{jl}\frac{\boldsymbol{G}_l^2\exp(\boldsymbol{G}_l\gamma)}{(1+\exp(\boldsymbol{G}_l\gamma))^2} + \frac{4\sum_{\{l\ j=0\}} n_{jl}\boldsymbol{G}_l^2\exp(2\boldsymbol{G}_l\gamma - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{(\exp(2\boldsymbol{G}_l\gamma)+\exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta})))^2}$$
$$+ \frac{2\sum_{\{l\ j=k\}} n_{jl}\boldsymbol{G}_l^2\exp(\boldsymbol{G}_l\gamma)\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}}{\left(2\exp(\boldsymbol{G}_l\gamma)+\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right)^2},$$

$$\frac{\partial^2\ell(\gamma,\beta)}{\partial\beta\partial\gamma} = \frac{\partial^2\ell(\gamma,\beta)}{\partial\gamma\partial\beta}$$
$$= \frac{2\sum_{\{l\ j=0\}} n_{jl}\boldsymbol{B}_l\boldsymbol{G}_l\exp(2\boldsymbol{G}_l\gamma+\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{(\exp(2\boldsymbol{G}_l\gamma)+\exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta})))^2}$$
$$- \frac{2\sum_{\{l\ j=k\}} n_{jl}\boldsymbol{B}_l\boldsymbol{G}_l\,(k-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))\exp(\boldsymbol{G}_l\gamma+k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!\left[2\exp(\boldsymbol{G}_l\gamma)+\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right]^2}, \text{ and}$$

$$\frac{\partial^2\ell(\gamma,\beta)}{\partial\beta^2} = \sum_{\{l\ j=0\}} n_{jl}\frac{-\boldsymbol{B}_l^2\exp(\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{(\exp(2\boldsymbol{G}_l\gamma)+\exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta})))^2}$$
$$\times\left[\exp(2\boldsymbol{G}_l\gamma)+\exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))-\exp(2\boldsymbol{G}_l\gamma+\boldsymbol{B}_l\boldsymbol{\beta})\right]$$
$$+\sum_{\{l\ j=k\}} n_{jl}\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!\left[2\exp(\boldsymbol{G}_l\gamma)+\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right]^2}$$
$$\times\left[2\exp(\boldsymbol{G}_l\gamma)\left(k\boldsymbol{B}_l-\boldsymbol{B}_l\exp(\boldsymbol{B}_l\beta)\right)^2\right]$$
$$-\sum_{\{l\ j=k\}} n_{jl}\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!\left[2\exp(\boldsymbol{G}_l\gamma)+\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right]^2}$$
$$\times\left[\boldsymbol{B}_l^2\exp(\boldsymbol{B}_l\boldsymbol{\beta})\left(2\exp(\boldsymbol{G}_l\gamma)+\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right)\right]$$
$$-\sum_{\{l\ \substack{j=1\\ \neq k}\}}^{m} n_{jl}\left(\boldsymbol{B}_l^2\exp(\boldsymbol{B}_l\beta)\right)$$

The estimated covariance matrix can be obtained by taking the inverse of the observed information matrix. The observed information matrix is the negative Hessian matrix evaluated at the ML estimates. The standard errors for the parameter estimates are obtained by taking the square roots of the diagonal elements of the asymptotic variances of the parameter estimates.

IV.3.2 DIP (p_1, p_2, λ) REGRESSION MODEL

Suppose that our observations on n subjects are independently distributed as doubly inflated poisson counts with parameters $(p_{1l}, p_{2l}, \lambda_l)$, $l = 1, \ldots, s$. Assume that the covariate-specific probabilities p_{1l} and p_{2l} can be parameterized by using the logistic regression, and the mean parameter λ_l can be parameterized by using the log link function. That is,

$$\begin{aligned} \operatorname{logit}(p_{1l}) &= \boldsymbol{A}_l, \boldsymbol{\alpha}, \qquad \operatorname{logit}(p_{2l}) = \boldsymbol{G}_l, \boldsymbol{\gamma}, \text{ and} \\ \log(\lambda_l) &= \boldsymbol{B}_l, \boldsymbol{\beta}, \end{aligned} \tag{53}$$

where $\boldsymbol{A}_l, \boldsymbol{G}_l, \boldsymbol{B}_l$ are the lth vector of the design matrices $\boldsymbol{A}, \boldsymbol{G}$ and $\boldsymbol{B}$. Also, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$ are regression parameters. Under the assumption that the covariate-specific parameters p_l and λ_l are not related, the log-likelihood function is given by

$$\begin{aligned} \ell(p_{1l}, p_{2l}, \lambda_l | \boldsymbol{n}_j) = & \sum_{\{l.j=0\}} n_{jl} \log\Big(p_{1l} + p_{3l}\exp(-\lambda_l)\Big) \\ & + \sum_{\{l:j=k\}} n_{jl} \log\left(p_{2l} + p_{3l}\frac{\exp(-\lambda_l)\lambda_l^k}{k!}\right) \\ & + \sum_{\{l: {j=1 \atop \neq k}\}}^{m} n_{jl} \log\left(p_{3l}\frac{\exp(-\lambda_l)\lambda_l^j}{j!}\right), \end{aligned} \tag{54}$$

which, when written in terms of the regression parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$, is as follows:

$$\begin{aligned} \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}) = & \sum_{\{l \cdot j=0\}} n_{jl} \log M_1 + \sum_{\{l:j=k\}} n_{jl} \log M_2 + \sum_{\{l: {j=1 \atop \neq k}\}}^{m} n_{jl} \left(\log M_3 + M_4\right) \\ & + \sum_{\{l.j=0\}}^{m} n_{jl} \left(\log M_5 + \log M_6\right). \end{aligned}$$

Note that $n = \sum_{l=1}^{s} \sum_{j=0}^{m} n_{jl}$, and

$$M_1 = \exp(\boldsymbol{A}_l\boldsymbol{\alpha})(1+\exp(\boldsymbol{G}_l\boldsymbol{\gamma})) + \exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))\left(1-\exp(\boldsymbol{A}_l\boldsymbol{\alpha}+\boldsymbol{G}_l\boldsymbol{\gamma})\right),$$

$$M_2 = \exp(\boldsymbol{G}_l\boldsymbol{\gamma})(1+\exp(\boldsymbol{A}_l\boldsymbol{\alpha})) + \frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}(1-\exp(\boldsymbol{A}_l\boldsymbol{\alpha}+\boldsymbol{G}_l\boldsymbol{\gamma}))$$

$$M_3 = 1-\exp(\boldsymbol{A}_l\boldsymbol{\alpha}+\boldsymbol{G}_l\boldsymbol{\gamma}), \qquad M_4 = j\boldsymbol{B}_l\boldsymbol{\beta}-\exp(\boldsymbol{B}_l\boldsymbol{\beta})-\log(j!),$$

$$M_5 = 1+\exp(\boldsymbol{A}_l\boldsymbol{\alpha}), \text{ and } \qquad M_6 = 1+\exp(\boldsymbol{G}_l\boldsymbol{\gamma}).$$

The maximum likelihood estimate of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$ is the solution to the likelihood equations

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = 0, \qquad \frac{\partial \ell}{\partial \boldsymbol{\gamma}} = 0, \text{ and } \qquad \frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0,$$

where

$$\begin{aligned}
\frac{\partial \ell}{\partial \boldsymbol{\alpha}} &= \sum_{\{l:j=0\}} n_{jl}\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\alpha}} + \sum_{\{l:j=k\}} n_{jl}\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\alpha}} + \sum_{\{l\, {j=1 \atop \neq k}\}}^{m} n_{jl}\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\alpha}} \\
&\quad + \sum_{\{l\, j=0\}}^{m} n_{jl}\frac{1}{M_5}\frac{\partial M_5}{\partial \boldsymbol{\alpha}}, \\
\frac{\partial \ell}{\partial \boldsymbol{\gamma}} &= \sum_{\{l:j=0\}} n_{jl}\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\gamma}} + \sum_{\{l:j=k\}} n_{jl}\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\gamma}} + \sum_{\{l\, {j=1 \atop \neq k}\}}^{m} n_{jl}\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\gamma}} \\
&\quad + \sum_{\{l\, j=0\}}^{m} n_{jl}\frac{1}{M_5}\frac{\partial M_5}{\partial \boldsymbol{\gamma}}, \text{ and} \\
\frac{\partial \ell}{\partial \boldsymbol{\beta}} &= \sum_{\{l\, j=0\}} n_{jl}\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\beta}} + \sum_{\{l\, j=k\}} n_{jl}\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\beta}} + \sum_{\{l\, {j=1 \atop \neq k}\}}^{m} n_{jl}\frac{\partial M_4}{\partial \boldsymbol{\beta}}.
\end{aligned}$$

Here,

$$\frac{\partial M_1}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_l\Big[\exp(\boldsymbol{A}_l\boldsymbol{\alpha}) + \exp(\boldsymbol{A}_l\boldsymbol{\alpha}+\boldsymbol{G}_l\boldsymbol{\gamma})\big(1-\exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))\big)\Big],$$

$$\frac{\partial M_2}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_l \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma})\left(1 - \frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta} - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right),$$

$$\frac{\partial M_3}{\partial \boldsymbol{\alpha}} = -\boldsymbol{A}_l \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma}), \qquad \frac{\partial M_5}{\partial \boldsymbol{\alpha}} = -\boldsymbol{A}_l \exp(\boldsymbol{A}_l\boldsymbol{\alpha}),$$

$$\frac{\partial M_1}{\partial \boldsymbol{\gamma}} = \boldsymbol{G}_l \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma})\big(1 - \exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))\big),$$

$$\frac{\partial M_2}{\partial \boldsymbol{\gamma}} = \boldsymbol{G}_l\left[\exp(\boldsymbol{G}_l\boldsymbol{\gamma}) + \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma})\left(1 - \frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta} - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\right)\right],$$

$$\frac{\partial M_3}{\partial \boldsymbol{\gamma}} = -\boldsymbol{G}_l \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma}), \qquad \frac{\partial M_6}{\partial \boldsymbol{\gamma}} = \frac{-\boldsymbol{G}_l \exp(\boldsymbol{G}_l\boldsymbol{\gamma})}{1 + \exp(\boldsymbol{G}_l\boldsymbol{\gamma})},$$

$$\frac{\partial M_1}{\partial \boldsymbol{\beta}} = -\boldsymbol{B}_l \exp(\boldsymbol{B}_l\boldsymbol{\beta} - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))\big(1 - \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma})\big),$$

$$\frac{\partial M_2}{\partial \boldsymbol{\beta}} = (k\boldsymbol{B}_l - \boldsymbol{B}_l \exp(\boldsymbol{B}_l\boldsymbol{\beta}))\big(1 - \exp(\boldsymbol{A}_l\boldsymbol{\alpha} + \boldsymbol{G}_l\boldsymbol{\gamma})\big)\frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta} - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}, \text{ and}$$

$$\frac{\partial M_4}{\partial \boldsymbol{\beta}} = j\boldsymbol{B}_l - \boldsymbol{B}_l \exp(\boldsymbol{B}_l\boldsymbol{\beta}).$$

The second order partial derivatives of the log-likelihood function are given by the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}^2} & \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\alpha}} & \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}} \\ \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\gamma}} & \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}^2} & \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\gamma}} \\ \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \end{pmatrix},$$

where

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\alpha}^2} = \sum_{\{l\ j=0\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ j=k\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ \substack{j=1\\ \neq k}\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ j=0\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{1}{M_5}\frac{\partial M_5}{\partial \boldsymbol{\alpha}}\right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\alpha}} = \sum_{\{l\ j=0\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ j=k\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ \substack{j=1\\ \neq k}\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ j=0\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_5}\frac{\partial M_5}{\partial \boldsymbol{\alpha}}\right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}} = \sum_{\{l\ j=0\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ j=k\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ \substack{j=1\\ \neq k}\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\alpha}}\right] + \sum_{\{l\ j=0\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_5}\frac{\partial M_5}{\partial \boldsymbol{\alpha}}\right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma}^2} = \sum_{\{l\ j=0\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\gamma}}\right] + \sum_{\{l\ j=k\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\gamma}}\right] + \sum_{\{l\ \substack{j=1\\ \neq k}\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\gamma}}\right] + \sum_{\{l\ j=0\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\gamma}}\left[\frac{1}{M_6}\frac{\partial M_6}{\partial \boldsymbol{\gamma}}\right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\gamma}} = \sum_{\{l\ j=0\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_1}\frac{\partial M_1}{\partial \boldsymbol{\gamma}}\right] + \sum_{\{l\ j=k\}} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_2}\frac{\partial M_2}{\partial \boldsymbol{\gamma}}\right] + \sum_{\{l\ \substack{j=1\\ \neq k}\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_3}\frac{\partial M_3}{\partial \boldsymbol{\gamma}}\right] + \sum_{\{l\ j=0\}}^{m} n_{jl}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\frac{1}{M_6}\frac{\partial M_6}{\partial \boldsymbol{\gamma}}\right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} = \sum_{\{l:j=0\}} n_{jl} \frac{\partial}{\partial \boldsymbol{\beta}} \left[\frac{1}{M_1} \frac{\partial M_1}{\partial \boldsymbol{\beta}} \right] + \sum_{\{l:j=k\}} n_{jl} \frac{\partial}{\partial \boldsymbol{\beta}} \left[\frac{1}{M_2} \frac{\partial M_2}{\partial \boldsymbol{\beta}} \right]$$
$$+ \sum_{\{l:\substack{j=1 \\ \neq k}\}}^{m} n_{jl} \frac{\partial}{\partial \boldsymbol{\beta}} \left[\frac{1}{M_4} \frac{\partial M_4}{\partial \boldsymbol{\beta}} \right],$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}} = \frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}}, \qquad \frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}} = \frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}}, \text{ and}$$

$$\frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\gamma}} = \frac{\partial^2 \ell(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\alpha}}.$$

The covariance matrix is obtained by taking the inverse of negative of the Hessian matrix calculated at the ML estimates. The standard errors for the parameter estimates are obtained from the diagonal elements of the covariance matrix.

IV.4 INFERENCE AND TESTING

Under the correct specification of the model, we know that the maximum likelihood estimates are the most efficient estimates. However, it is important to check for model adequacy. A common approach to check for the model adequacy is to compare the working model with the most complex model. The likelihood-ratio test statistic equals $-2(L_0 - L_1)$, where L_0 is the maximized log-likelihood under the simpler model. The *df* for the large-sample chi-squared distribution equals the number of extra parameters in the more complex model. A small p-value suggests that the more general model be used as a new working model (Agresti 2002).

In DIP regression models, the design matrices contain potentially different sets of experimental factors and covariate effects that pertain to the probabilities of count zero's and count k's as well as the Poisson mean, respectively. Therefore, the parameters have interpretations in terms of a covariate or factor level effect on the probabilities of count zero's and count k's as well as the Poisson mean. For DIP

(p, λ) regression model, the γ's have interpretations in terms of a covariate or factor's effect on the binomial probability and the β's have interpretations in terms of a covariate or factor's effect on the Poisson mean of the count responses. For DIP (p_1, p_2, λ) regression model, the α's and γ's have interpretations in terms of a covariate or factor's effect on the probabilities p_1 and p_2, respectively, and the β's have interpretations in terms of a covariate or factor's effect on the Poisson mean of the count responses. In either of the two models, interpretations of parameter estimates must be done with care, especially for those relate to the probabilities that lead to the additional proportions of inflated counts of zeros and k's.

Depending upon the fit of the model, we can make an inference about the model parameters using the ML estimates, their estimated standard errors, and their maximized log-likelihood function. For the lth parameter, say β_l, a 95% Wald confidence interval for the parameter is $\widehat{\beta}_l \pm 1.96(s.e.(\widehat{\beta}_l))$, where s.e.$(\widehat{\beta}_l$ is the estimated standard error of $\widehat{\beta}_l$, the lth component of $\boldsymbol{\beta}$. For testing $\boldsymbol{H_0}$: $\beta_l = 0$, we can use

$$z = \frac{\widehat{\beta}_l}{s.e.(\widehat{\beta}_l)}$$

or z^2, which (under $\boldsymbol{H_0}$) has an asymptotic chi-squared distribution with $df = 1$.

IV.4.1 INTERPRETING DIP REGRESSION MODELS

Consider DIP (p, λ) regression model as discussed in Sections IV.2.1, where the parameters p_i and λ_i are parameterized by the logit link and log link function, respectively (see equation (43)). Thus, for the u covariates, the logit of the probability has the following linear form:

$$\text{logit}(p_i) = \gamma_0 + \gamma_1 x_{i1} + \ldots + \gamma_u x_{iu}, \qquad i = 1, \ldots, N. \tag{57}$$

Hence, $\exp(\gamma_l)$ is the multiplicative effect on the odds of 1-unit increase in the lth covariate for ith subject, x_{il}, at the fixed levels of the other x_{il}'s, $l = 1, \ldots, u$.

Since p_i is the binomial probability, for a significant covariate, γ_l, the factor $\exp(2\gamma_l)$ explains the odds of additional proportions, p^2, for the inflated frequencies of zero counts. It is important to note that the probability of Doubly Inflated Poisson counts is as given by equation (4). Thus, interpretation and inference of the parameter estimates must be done carefully. It is also of importance to notice that if any of the variables x_{il} in the above equation (57) are dummy variables for the categorical explanatory variables, then $\exp(\gamma_l)$ describes the conditional odds ratio between the response and explanatory variables. That is, the odds of "success" at $x_{il} = 1$ is $\exp(\gamma_l)$ times the odds of "success" at $x_{il} = 0$. Similar interpretation of the parameters can also be done for the data with grouped frequencies, where p_l and λ_l are covariate specific parameter (see Section IV.3).

Consider now DIP(p_1, p_2, λ) regression model as discussed in Section IV.2.2, where the parameters p_{1i} and p_{2i} are parameterized by logistic regression, and λ_i by log-linear regression, respectively (see equation (47)). Thus, for the u covariates, the logit of the probabilities have the following linear form:

$$\begin{aligned} \text{logit}(p_{1i}) &= \alpha_0 + \alpha_1 x_{i1} + \ldots + \alpha_u x_{iu} \qquad && \text{and} \\ \text{logit}(p_{2i}) &= \gamma_0 + \gamma_1 x_{i1} + \ldots + \gamma_u x_{iu}, && i = 1, \ldots, N. \end{aligned} \tag{58}$$

Hence, $\exp(\alpha_l)$ and $\exp(\gamma_l)$ are the respective multiplicative effect on the odds of 1-unit increase in the lth covariate for the ith subject, x_{il}, at the fixed levels of the other x_{il}'s, $l = 1, \ldots, u$, for the each of the logit link functions. Since p_{1i} is probability that explains the additional proportions of zero counts, for a significant covariate, α_l, the odds of additional proportions of zero counts increase multiplicatively by $\exp(\alpha_l)$ for every one-unit increase in x_{il}. Similarly, for a significant covariate, γ_l, the odds of additional proportions of k counts increase multiplicatively by $\exp(\gamma_l)$ for every one-unit increase in x_{il}.

In DIP (p, λ) as well as DIP (p_1, p_2, λ) regression models, where the Poisson mean λ_i is parameterized by the log link function of the covariates, the log-linear model has the following form:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_u x_{iu} \tag{59}$$

for u covariates. Thus, a one-unit increase in x_{il} has a multiplicative impact of $\exp(\beta_l)$ on λ_i. If $\beta_l > 0$, then the mean λ_i increases as x_{il} increases. If $\beta_l < 0$, then the Poisson mean λ_i decreases as x_{il} decreases. Similar interpretation of the parameters can also be done for the data with grouped frequencies, where p_l and λ_l are covariate specific parameter (see Section IV.3).

IV.5 ILLUSTRATION OF METHODS

Returning to the DMFT data discussed earlier in the chapter, one of the questions of interest is to assess whether any of the treatments are significant in improving dental cavities amongst children of different gender and ethnicity. We would also like to assess, if possible, which of the treatments and other factors impact the substantial increase in count zeros and in count 1's. To answer such questions, we consider fitting two Doubly Inflated Poisson regression models discussed in this chapter and then compare to ZIP regression model.

Suppose that the response variable, number of faulty teeth, for each child i to be independently distributed as doubly inflated Poisson counts with parameters $(p_i, \lambda_i), i = 1, \ldots, n$ for DIP (p, λ) as well as ZIP Regression Models. The relationship between p and λ is unknown. Also, there is no evidence that the covariates affecting p and λ are the same. Thus, we fit DIP (p, λ) regression model (see Section IV.2.1) and ZIP regression model (Lambert 1992). Since the log-likelihood function is not of closed form, we consider finding first-order solutions for ML estimation using the Newton-Rhapson method implemented by SAS NLPNRA routine in PROC IML. Table 19 lists the log-likelihood values as well as the parameter estimates for ZIP and DIP(p, λ) regression models. Next, we consider that the number of faulty teeth to be independently distributed as DIP (p_1, p_2, λ) counts with parameters $(p_{1i}, p_{2i}, \lambda_i), i = 1, ..., n$. Fitting the regression model as described in Section IV.2.2, we can obtain parameter estimates by maximizing the log-likelihood function. Table 20 lists log-likelihood value as well as the parameter estimates for the DIP (p_1, p_2, λ) regression model.

Comparing two regression models, ZIP and DIP (p, λ), one can see that the log-likelihood value is a slight improvement. It is also interesting to note that the only covariate significant in estimating the binomial probability p for both regression models have a different estimate. Also, in DIP (p, λ) regression model, the treatment group "All" is significant in estimating binomial probability and the treatment group "Oral Health Education" is significant in estimating the Poisson mean parameter.

Under DIP (p_1, p_2, λ) model, the treatment group "All" and "Rinse" are significant in estimating p_1, the "Black" ethnicity group is significant in estimating p_2, and the treatments "All" and "Oral Health Education" are significant in estimating the Poisson mean λ. The log-likelihood value has significantly improved in this case. All the results have been verified using PROC NLP as well as PROC NLMIXED in SAS. The results show that the odds of consistent dental care amongst those who rinse with 0.2% sodium flouride solution is $\exp(-1.041) = 0.35$ or approximately 1/3 to the odds of those who do not. The odds of improvement in one dental cavity is $\exp(-1.078) = 0.34$ amongst blacks compared to the others that are predominantly Hispanic. Additionally, the treatment of oral health education helps reduce the dental cavities amongst children by 30% on an average. All four treatments help reduce the dental cavities by 23% on an average.

One may also consider fitting DIP regression models under the assumption that only the Poisson mean is a function of a given set of covariates via log link. That is, the probabilities p_1 and p_2 are a constant function rather than a function of set of covariates. Thus, only one set of covariates needs to be estimated using the log-linear regression model for the Poisson mean. Results obtained by using above explained model for DIP (p_1, p_2, λ) regression show that all variables are significant except for diet enriched with rice bran, gender, and white ethnicity group. The negative log-likelihood value was found to be 1668.270. The results are as shown in Table 21.

IV.6 BASELINE-CATEGORY LOGITS FOR DIP (P_1, P_2, λ) REGRESSION MODELS

Usually, for nominal-level response variables, the standard logits are the baseline-category logits, also known as the generalized logit models (Agresti 2002). When the last category c is the baseline, the baseline-category logits are

$$\log\left(\frac{\pi_j}{\pi_c}\right), j = 1, \ldots, c-1$$

For $c = 3$, the logit model uses $\log(\pi_1/\pi_3)$ and $\log(\pi_2/\pi_3)$. Hence, we model the log odds of being at any particular level j as compared to being in the reference class c, and this relationship is allowed to be different across the covariates. The difference between the logistic regression model and the generalized logit model is how a researcher prefers to explain the effects of covariates. The logistic regression model describes effects within each individual response category, whereas the generalized logit model describes effects with the pairs of individual categories and a baseline category.

In order to apply the generalized logit models to DIP (p_1, p_2, λ) regression models, we can assume the probability of p_3 for the reference category that describes counts under Poisson distribution, and compare the probabilities of p_1 and p_2 to this reference class. Note: $p_3 = 1 - p_1 - p_2$. Thus, for data consisting of raw counts as described in Section IV.2, DIP (p_1, p_2, λ) regression model with generalized logit links is as follows:

$$\log\left(\frac{p_{1\imath}}{p_{3\imath}}\right) = \boldsymbol{A}_\imath \boldsymbol{\alpha}, \qquad \log\left(\frac{p_{2\imath}}{p_{3\imath}}\right) = \boldsymbol{G}_\imath \boldsymbol{\gamma}, \text{ and } \qquad \log(\lambda_\imath) = \boldsymbol{B}_\imath \boldsymbol{\beta} \tag{60}$$

The log-likelihood function given by equation (54) can also be expressed as follows for the baseline-category logit links:

$$\begin{aligned} \ell(\alpha, \gamma, \beta) = & \sum_{\{\imath\, y_\imath = 0\}} \log\left[\exp(\boldsymbol{A}_\imath \boldsymbol{\alpha}) + \exp(-\exp(\boldsymbol{B}_\imath \boldsymbol{\beta}))\right] \\ & + \sum_{\{\imath : y_\imath = k\}} \log\left[\exp(\boldsymbol{G}_\imath \boldsymbol{\gamma}) + \frac{\exp(k\boldsymbol{B}_\imath \boldsymbol{\beta} - \exp(\boldsymbol{B}_\imath \boldsymbol{\beta}))}{k!}\right] \\ & + \sum_{\{\imath : \substack{y_\imath = y \\ y \neq 0, k}\}} \log\left[\frac{\exp(y_\imath \boldsymbol{B}_\imath \boldsymbol{\beta} - \exp(\boldsymbol{B}_\imath \boldsymbol{\beta}))}{y_\imath !}\right] + \sum_{\{\imath\, y_\imath = y\}} p_{3\imath} \end{aligned} \tag{61}$$

For data consisting of grouped frequencies as described in Section IV.3, DIP (p_1, p_2, λ) regression model with generalized logit links is as follows:

$$\log\left(\frac{p_{1l}}{p_{3l}}\right) = \boldsymbol{A}_l, \boldsymbol{\alpha}, \qquad \log\left(\frac{p_{2l}}{p_{3l}}\right) = \boldsymbol{G}_l, \boldsymbol{\gamma}, \text{ and} \log(\lambda_l) = \boldsymbol{B}_l, \boldsymbol{\beta} \tag{62}$$

The log-likelihood function given by equation (54) can also be expressed as follows for DIP (p_1, p_2, λ) regression using the baseline-category logit links:

$$\begin{aligned}\ell(\alpha,\gamma,\beta) = &\sum_{\{l:j=0\}} n_{jl} \log\left[p_{3l}\Big(\exp(\boldsymbol{A}_l\boldsymbol{\alpha}) + \exp(-\exp(\boldsymbol{B}_l\boldsymbol{\beta}))\Big)\right] \\ &+ \sum_{\{l:j=k\}} n_{jl} \log\left[p_{3l}\Big(\exp(\boldsymbol{G}_l\boldsymbol{\gamma}) + \frac{\exp(k\boldsymbol{B}_l\boldsymbol{\beta} - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{k!}\Big)\right] \\ &+ \sum_{\{l:\substack{j=1\\ \neq k}\}}^{m} n_{jl} \log\left[p_{3l}\frac{\exp(j\boldsymbol{B}_l\boldsymbol{\beta} - \exp(\boldsymbol{B}_l\boldsymbol{\beta}))}{j!}\right]\end{aligned} \tag{63}$$

. Since the log-likelihood functions given by equations (61) and (63) are not in a closed form, numerical algorithms such as Newton-Rhapson algorithm can be used to find solutions for ML estimates $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\beta}})$.

Returning to the dental epidemiology data, a researcher may want to explain the effects of covariates on the inflated counts of 0's and 1's in comparison to those of Poisson counts. Table 22 shows the results for the dental data obtained by maximizing log-likelihood function for DIP(p_1, p_2, λ) regression model with Poisson counts as reference category. It is interesting to note that the demographical factors such as gender and race are significant when comparings the inflated counts to those of Poisson counts. For the significant treatments, the interpretation is as follows: The odds of consistent dental care is approximately 3.5 times to those with Poisson counts amongst the children using the rinsing solution; the odds of improvement in one dental cavity is 3.75 times the odds of Poisson counts for those who received oral health education; and, the odds of improvement in one dental cavity is 3.9 times the odds of Poisson counts for those who received oral hygiene treatment.

Table 19: Parameter Estimates for ZIP and DIP Model (p, λ)

	ZIP		DIP Model (p, λ)	
Parameter	Est.(S.E.)	p-value	Est.(S.E.)	p-value
Logit Link for p				
Constant	2.857(1.281)	0.026	1.484(0.677)	0.029
Treatment:				
Educ	0.461(1.684)	0.784	1.161(0.807)	0.151
Enrich	−0.116(1.614)	0.943	0.847(0.772)	0.273
Rinse	−1.001(0.792)	0.207	−0.928(0.544)	0.088
Hygiene	−0.237(1.275)	0.853	−0.760(0.572)	0.185
All	−0.827(0.959)	0.388	−1.195(0.543)	0.028
Gender:				
Male	0.184(0.627)	0.769	0.077(0.236)	0.774
Ethnicity:				
White	−0.416(1.325)	0.754	−0.181(0.354)	0.609
Black	12.230(3.191)	0.001	−0.766(0.362)	0.035
Log Link for λ				
Constant	0.713(0.116)	0.001	0.878(0.190)	0.001
Treatment:				
Educ	−0.232(0.108)	0.032	−0.451(0.169)	0.008
Enrich	−0.082(0.116)	0.518	−0.254(0.180)	0.158
Rinse	−0.214(0.106)	0.044	−0.003(0.159)	0.983
Hygiene	−0.262(0.137)	0.055	−0.041(0.200)	0.838
All	−0.448(0.129)	0.001	−0.097(0.167)	0.561
Gender:				
Male	0.107(0.074)	0.152	0.113(0.085)	0.184
Ethnicity:				
White	0.134(0.124)	0.279	0.175(0.109)	0.111
Black	−0.273(0.111)	0.014	0.060(0.137)	0.659
NegLL	1713.094		1700.094	

Table 20: Parameter Estimates for DIP (p_1, p_2, λ) Model

	Logit (p_1)		**Logit (p_2)**		**Log(λ)**	
Parm	Est.(S.E.)	p-value	Est.(S.E.)	p-value	Est.(S.E.)	p-value
Const	2.047(0.388)	0.001	1.356(0.372)	0.001	0.930(0.108)	0.001
Trt						
Educ	0.319(0.634)	0.614	0.306(0.518)	0.555	−0.332(0.128)	0.009
Enrich	−0.166(0.445)	0.710	−0.018(0.380)	0.962	−0.068(0.110)	0.534
Rinse	−1.041(0.374)	0.005	−0.006(0.367)	0.987	−0.096(0.111)	0.390
Hyg	−0.696(0.404)	0.085	−0.517(0.381)	0.175	−0.059(0.141)	0.673
All	−1.096(0.379)	0.004	−0.222(0.379)	0.559	−0.266(0.138)	0.055
Gender						
Male	0.113(0.213)	0.595	−0.047(0.227)	0.832	0.113(0.074)	0.127
Eth						
White	−0.027(0.260)	0.918	0.116(0.303)	0.703	0.105(0.087)	0.231
Black	−0.114(0.303)	0.707	−1.078(0.315)	0.001	0.053(0.133)	0.694
NegLL	1648.925					

Table 21: Parameter Estimates for DIP (p_1, p_2, λ) Model with constants p_1 and p_2

Parameter	Est.(S.E.)	p-value
Constant	1.012(0.080)	0.001
Treatment		
Educ	−0.243(0.090)	0.007
Enrich	−0.090(0.085)	0.292
Rinse	−0.256(0.098)	0.009
Hyg	−0.235(0.114)	0.040
All	−0.500(0.125)	0.081
Gender		
Male	0.107(0.061)	0.081
Ethnicity		
White	0.112(0.066)	0.100
Black	−0.242(0.106)	0.023

Table 22: Parameter Estimates using Baseline-Category Logit p_3

	Logit (p_1)		**Logit** (p_2)		**Log**(λ)	
Parm	Est.(S.E.)	p-value	Est.(S.E.)	p-value	Est.(S.E.)	p-value
Const	4.337(0.636)	0.001	7.022(1.559)	0.001	−1.418(1.191)	0.883
Trt						
Educ	0.649(0.617)	0.859	1.324(0.545)	0.008	−0.105(0.061)	0.043
Enrich	0.754(0.602)	0.106	1.279(0.567)	0.0119	−0.074(0.059)	0.106
Rinse	1.277(0.604)	0.017	1.320(1.556)	0.848	−0.490(0.059)	0.001
Hyg	0.871(0.603)	0.075	1.364(0.556)	0.007	−0.642(0.072)	0.001
All	1.130(0.597)	0.083	1.332(1.258)	0.145	−1.145(0.067)	0.001
Gender						
Male	2.073(0.216)	0.001	2.900(0.258)	0.001	0.176(0.041)	0.001
Eth						
White	0.871(0.231)	0.001	1.951(0.208)	0.001	−1.039(0.067)	0.001
Black	2.126(0.296)	0.001	3.719(0.250)	0.001	0.165(0.037)	0.001
NegLL	1797.693					

CHAPTER V

FUTURE CONSIDERATIONS

Since most real-life count data are overdispersed, analysts typically employ alternatives to the Poisson model such as the negative binomial probability model which uses an additional parameter in describing overdispersed count responses (Greene 2007; Lawless 1987). For count data consisting of excess zeros, zero-inflated negative binomial (ZINB) regression models have been studied extensively (Yau et al. 2003). For data consisting of two inflated counts, 0 and k, one may also ask if the overdispersed count responses can be explained by the two peaks and/or by the negative binomial distribution. This can be accomplished by using Doubly Inflated Negative Binomial (DINB) models. In this chapter, we first begin with a review of negative binomial distribution in Section V.1. Section V.2 describes two Doubly Inflated Negative Binomial (DINB) models with a brief discussion of their distributional properties. In Section V.3, we describe the maximum likelihood estimation technique for both DINB models.

V.1 NEGATIVE BINOMIAL DISTRIBUTION

The most commonly used Negative Binomial distribution is the NB2 model, which has the following probability mass function:

$$g(y|\mu,\eta) = \frac{\Gamma(y+\eta^{-1})}{\Gamma(y+1)\Gamma(\eta^{-1})}\left(\frac{\eta^{-1}}{\mu+\eta^{-1}}\right)^{\eta^{-1}}\left(\frac{\mu}{\mu+\eta^{-1}}\right)^{y}, \tag{64}$$

where $\eta > 0$ and $y = 0, 1, 2, \ldots$. The mean and variance of NB2 model are μ and $\mu(1+\eta\mu)$, respectively. If we replace η^{-1} with $\eta^{-1}\mu$, then we get the NB1 model which has the following density

$$g(y|\mu,\eta) = \frac{\Gamma(y+\mu\eta^{-1})}{\Gamma(y+1)\Gamma(\mu\eta^{-1})}\left(\frac{\eta^{-1}}{1+\eta^{-1}}\right)^{\mu\eta^{-1}}\left(\frac{1}{1+\eta^{-1}}\right)^{y}, \tag{65}$$

where $\eta > 0$ and $y = 0, 1, 2, \ldots$. The mean and variance of NB1 model are μ and $\mu(1+\eta)$, respectively.

Cameron and Trivedi (1998) discussed the NBr parameterization where $\tau = \mu^{2-r}\eta^{-1}$ and $\xi = \mu^{r-1}\eta$, where $1 \leq r \leq 2$. Thus, if we substitute $\tau = \eta^{-1}$ and $\xi = \mu\eta$ in equation (64) for NB2 model, and $\tau = \mu\eta^{-1}$ and $\xi = \eta$ in equation (65) for NB1 model, then we obtain the following Negative Binomial distribution with parameters (τ, ξ):

$$g(y|\tau, \xi) = \frac{\Gamma(\tau + y)}{\Gamma(\tau)\Gamma(y+1)} \frac{\xi^y}{(1+\xi)^{\tau+y}}, \text{ for } y = 0, 1, 2, \ldots \tag{66}$$

where, τ is a real positive number and $\xi > 0$. Note that $\xi = 1/(\pi - 1)$ where π is a Bernoulli parameter. The mean is $\mu = \tau\xi$ and variance is $\sigma^2 = \tau\xi(1+\xi)$. The overdispersion index is $(1+\xi)$. One could also consider the generalized NBr model where $1 \leq r \leq 2$ in which case r must be estimated.

V.2 DOUBLY INFLATED NEGATIVE BINOMIAL DISTRIBUTIONS

If the observed data is overdispersed then negative binomial may be used for modeling purposes as it uses an additional parameter in describing the variance of the count response variable as compared to Poisson distribution. However, if the data has inflated frequencies of zero's and inflated frequencies another count value k, then a Doubly Inflated Negative Binomial (DINB) may also be a good fit. The construction of DINB probability model can be paralleled to that of DIP models as discussed in Chapters II and III. Let Z be a latent random variable distributed as Binomial $(2, p)$, $0 \leq p \leq 1$. Under the assumptions that Y given $Z = 2$ is degenerate at 0; Y given $Z = 1$ is degenerate at k; and, Y given $Z = 0$ is a negative binomial with real-valued positive parameters τ and ξ, then one can easily construct the first DINB distribution as follows:

$$f(y|p, \tau, \xi) = \begin{cases} p^2 + q^2 g(0|\tau, \xi), & \text{for } y = 0; \\ 2pq + q^2 g(k|\tau, \xi), & \text{for } y = k; \\ q^2 g(y|\tau, \xi), & \text{for } y = 1, 2, \ldots \neq k. \end{cases} \tag{67}$$

Here, $g(y|\tau, \xi)$ is assumed to be negative binomial distribution as described by equation (66), and $p + q = 1$. We call this distribution as *Doubly Inflated Negative Binomial* (p, τ, ξ), abbreviated as DINB (p, τ, ξ). The mean and variance of DINB (p, τ, ξ) are $E(Y) = 2pqk + q^2\tau\xi$ and $Var(Y) = 2pqk^2 + q^2\tau\xi(2 + \xi) - (2pqk + q^2\tau\xi)^2$, respectively. It is clear that as $p \to 0$, this model reduces to an ordinary Negative Binomial distribution with parameters (τ, ξ). Further, since τ is a function of η, if $\eta \to 0$, it will further reduce to an ordinary Poisson distribution.

Similarly, DINB (p_1, p_2, τ, ξ) can be constructed by assuming that the latent variable Z taking three values with probabilities p_1, p_2, and p_3 and Y conditional upon Z is distributed as described above. Thus, the probability mass function of Doubly Inflated Negative Binomial (p_1, p_2, τ, ξ) distribution is therefore:

$$f(y|p_1, p_2, \tau, \xi) = \begin{cases} p_1 + p_3 g(0|\tau, \xi), & \text{for } y = 0; \\ p_2 + p_3 g(k|\tau, \xi), & \text{for } y = k; \\ p_3 g(y|\tau, \xi), & \text{for } y = 1, 2, \ldots \neq k. \end{cases} \tag{68}$$

where $p_1 + p_2 + p_3 = 1$. Here, $g(y|\tau, \xi)$ is assumed to be negative binomial distribution as described by (66). We call this distribution as *Doubly Inflated Negative Binomial* (p_1, p_2, τ, ξ), abbreviated as DINB (p_1, p_2, τ, ξ). The mean and variance of DINB (p_1, p_2, τ, ξ) are $E(Y) = p_2 k + p_3\tau\xi$ and $Var(Y) = p_2 k^2 + p_3\tau\xi(2 + \xi) - (p_2 k + p_3\tau\xi)^2$, respectively. It is also interesting to note that as $p_2 \to 0$, this model reduces to a Zero-Inflated Negative Binomial (ZINB) distribution. Further, since τ is a function of η, if $\eta \to 0$, this distribution reduces further to a Zero-Inflated Poisson (ZIP) distribution. A more detailed discussion of distributional properties for DINB models will be covered elsewhere.

V.3 DINB REGRESSION MODELS

Here we investigate if the explanatory variables have an effect on the inflated counts as well as on the mean parameter of the negative binomial distribution. For both

NB1 and NB2 case, the log-likelihood function for DINB (p, τ, ξ) is as follows:

$$\ell = n_0 \log\left(p^2 + q^2 g(0|\tau,\xi)\right) + n_k \log\left(2pq + q^2 g(k|\tau,\xi)\right) + \sum_{\substack{j=1 \\ j\neq k}}^{m} n_j \log\left(q^2 g(j|\tau,\xi)\right) \tag{69}$$

Under the assumption that the mean μ_i and p_i can be parameterized by their respective log and logit link functions, DINB (p, τ, ξ) regression can be modeled as

$$\text{logit}(p_i) = \boldsymbol{B_i}\,\boldsymbol{\beta}, \text{ and } \qquad \log(\mu_i) = \boldsymbol{G_i}\,\gamma \tag{70}$$

where, $\mu_i = \tau_i \xi_i$ is the mean of negative binomial distribution. Here, $\boldsymbol{B_i}$ and $\boldsymbol{G_i}$ are ith vector of covariates and β and γ are regression parameters. Then, the log-likelihood function for data consisting of grouped frequencies can be written in terms of regression parameters is

$$\begin{aligned}\ell = n_0 \log\Big(\exp(\boldsymbol{2B_i\beta}) + g(0|\tau_i,\xi_i)\Big) + n_k \log\Big(2\exp(\boldsymbol{B_i\beta}) + g(k|\tau_i,\xi_i)\Big) \\ + \sum_{\substack{j=1 \\ j\neq k}}^{m} n_j \log\Big(g(j|\tau_i,\xi_i)\Big) - 2\sum_{j=0}^{m} n_j \log\Big(1 + \exp(\boldsymbol{B_i\beta})\Big)\end{aligned} \tag{71}$$

It is interesting to note that for NB1 model, ξ does not depend on the set of covariates, and for NB2 model, τ does not depend on the set of covariates. As mentioned earlier, one could also use the generalized NBr model where $1 \leq r \leq 2$. However, it will add an additional parameter that needs to be estimated and will be discussed elsewhere. The ML estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\gamma}$ for DINB (p, τ, ξ) model are the solution to the first-order derivatives

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0, \text{ and } \qquad \frac{\partial \ell}{\partial \gamma} = 0,$$

where

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = n_0 \frac{1}{\boldsymbol{V}_1}\frac{\partial \boldsymbol{V}_1}{\partial \boldsymbol{\beta}} + n_k \frac{1}{\boldsymbol{V}_2}\frac{\partial \boldsymbol{V}_2}{\partial \boldsymbol{\beta}} - \sum_{j=0}^{m} n_j \frac{1}{\boldsymbol{V}_3}\frac{\partial \boldsymbol{V}_3}{\partial \boldsymbol{\beta}}, \text{ and}$$

$$\frac{\partial \ell}{\partial \gamma} = n_0 \frac{1}{\boldsymbol{V}_1} \frac{\partial \boldsymbol{V}_1}{\partial \gamma} + n_k \frac{1}{\boldsymbol{V}_2} \frac{\partial \boldsymbol{V}_2}{\partial \gamma} - \sum_{\substack{j=1 \\ j \neq k}}^{m} n_j \frac{1}{g(j)} \frac{\partial g(j|\tau_i, \xi_i)}{\partial \gamma}.$$

Here,

$$\begin{aligned}
&\boldsymbol{V}_1 = \exp(2\boldsymbol{B_i\beta}) + g(0|\tau_i, \xi_i), && \boldsymbol{V_2} = 2\exp(\boldsymbol{B_i\beta}) + g(k|\tau_i, \xi_i), \\
&\boldsymbol{V}_3 = (1 + \exp(\boldsymbol{B_i\beta})), && \frac{\partial \boldsymbol{V_1}}{\partial \boldsymbol{\beta}} = 2\boldsymbol{B_i\beta}\exp(2\boldsymbol{B_i\beta}), \\
&\frac{\partial \boldsymbol{V_2}}{\partial \boldsymbol{\beta}} = 2\boldsymbol{B_i\beta}\exp(\boldsymbol{B_i\beta}), && \frac{\partial \boldsymbol{V_3}}{\partial \boldsymbol{\beta}} = \boldsymbol{B_i\beta}\exp(\boldsymbol{B_i\beta}), \\
&\frac{\partial \boldsymbol{V_1}}{\partial \gamma} = \frac{\partial}{\partial \gamma} g(0|\tau_i, \xi_i), && \frac{\partial \boldsymbol{V_2}}{\partial \gamma} = \frac{\partial}{\partial \gamma} g(k|\tau_i, \xi_i), \text{ and } \quad \frac{\partial \boldsymbol{V_3}}{\partial \gamma} = 0.
\end{aligned}$$

The second order partial derivatives of the log-likelihood function are given by the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \gamma} \\ \dfrac{\partial^2 \ell}{\partial \gamma \partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2} \end{pmatrix}$$

The covariance matrix is obtained by taking the inverse of negative of the Hessian matrix calculated at the ML estimates. Thus, the standard errors for parameter estimates $(\widehat{\boldsymbol{\beta}}, \widehat{\gamma})$ can be easily obtained by taking square root of the asymptotic variances. Similarly, for both NB1 and NB2 case, the log-likelihood function for DINB (p_1, p_2, τ, ξ) is as follows:

$$\ell = n_0 \log\left(p_1 + p_3 g(0|\tau, \xi)\right) + n_k \log\left(p_2 + p_3 g(k|\tau, \xi)\right) + \sum_{\substack{j=1 \\ j \neq k}}^{m} n_j \log\left(p_3 g(y|\tau, \xi)\right). \tag{73}$$

To fit DINB (p_1, p_2, τ, ξ) regression model, let

$$\text{logit}(p_{1i}) = \boldsymbol{A_i}\, \boldsymbol{\alpha}, \qquad \text{logit}(p_{2i}) = \boldsymbol{B_i}\, \boldsymbol{\beta}, \qquad \text{and} \qquad \log(\mu_i) = \boldsymbol{G_i}\, \gamma \tag{74}$$

where $\mu_i = \tau_i \xi_i$ is the mean of negative binomial distribution, and $\boldsymbol{A}_i, \boldsymbol{B}_i$, and $\boldsymbol{G}_i$ are ith vector of covariates and α, β, and γ are regression parameters. Thus, DINB (p_1, p_2, τ, ξ) regression model can be expressed in terms of regression parameters as follows:

$$\begin{aligned}
\ell = n_0 \log &\Bigg(\exp(\boldsymbol{A_i\alpha}) + \big(1 - g(0|\tau_\imath, \xi_\imath)\big) \exp(\boldsymbol{A_i\alpha} + \boldsymbol{B_i\beta}) + g(0|\tau_\imath, \xi_\imath) \Bigg) \\
+ n_k \log &\Bigg(\exp(\boldsymbol{A_i\alpha}) + \big(1 - g(k|\tau_\imath, \xi_\imath)\big) \exp(\boldsymbol{A_i\alpha} + \boldsymbol{B_i\beta}) + g(k|\tau_\imath, \xi_\imath) \Bigg) \\
+ \sum_{\substack{\jmath=1 \\ \jmath \neq k}}^{m} n_\jmath \log &\Bigg(g(\jmath|\tau_\imath, \xi_\imath)\big(1 - \exp(\boldsymbol{A_i\alpha} + \boldsymbol{B_i\beta})\big) \Bigg) \\
+ \sum_{\jmath=0}^{m} n_\jmath \log &\Bigg(\big(1 + \exp(\boldsymbol{B_i\beta})\big)\big(1 + \exp(\boldsymbol{A_i\alpha})\big) \Bigg)
\end{aligned} \tag{75}$$

If we let

$$\begin{aligned}
\boldsymbol{U_1} &= \exp(\boldsymbol{A_i\alpha}) + \big(1 - g(0|\tau_\imath, \xi_\imath)\big) \exp(\boldsymbol{A_i\alpha} + \boldsymbol{B_i\beta}) + g(0|\tau_\imath, \xi_\imath), \\
\boldsymbol{U_2} &= \exp(\boldsymbol{A_i\alpha}) + \big(1 - g(k|\tau_\imath, \xi_\imath)\big) \exp(\boldsymbol{A_i\alpha} + \boldsymbol{B_i\beta}) + g(k|\tau_\imath, \xi_\imath), \\
\boldsymbol{U_3} &= g(\jmath|\tau_\imath, \xi_\imath)\big(1 - \exp(\boldsymbol{A_i\alpha} + \boldsymbol{B_i\beta})\big), \text{ and} \\
\boldsymbol{U_4} &= \big(1 + \exp(\boldsymbol{B_i\beta})\big)\big(1 + \exp(\boldsymbol{A_i\alpha})\big),
\end{aligned}$$

then (75) can simply be written as:

$$\ell = n_0 \log \boldsymbol{U_1} + n_k \log \boldsymbol{U_2} + \sum_{\substack{\jmath=1 \\ \jmath \neq k}}^{m} n_\jmath \log \boldsymbol{U_3} + \sum_{\jmath=0}^{m} n_\jmath \log \boldsymbol{U_4}.$$

Thus, the ML estimate $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ for DINB (p_1, p_2, τ, ξ) model, is the solution to the first-order derivatives

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = 0, \qquad \frac{\partial \ell}{\partial \boldsymbol{\beta}} = 0, \text{ and} \qquad \frac{\partial \ell}{\partial \gamma} = 0$$

where

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = n_0 \frac{1}{\boldsymbol{U}_1} \frac{\partial \boldsymbol{U}_1}{\partial \boldsymbol{\beta}} + n_k \frac{1}{\boldsymbol{U}_2} \frac{\partial \boldsymbol{U}_2}{\partial \boldsymbol{\beta}} + \sum_{\substack{\jmath=1 \\ \jmath \neq k}}^{m} n_\jmath \frac{1}{\boldsymbol{U}_3} \frac{\partial \boldsymbol{U}_3}{\partial \boldsymbol{\beta}} + \sum_{\jmath=0}^{m} n_\jmath \frac{1}{\boldsymbol{U}_4} \frac{\partial \boldsymbol{U}_4}{\partial \boldsymbol{\beta}},$$

$$\frac{\partial \ell}{\partial \boldsymbol{\alpha}} = n_0 \frac{1}{\boldsymbol{U}_1}\frac{\partial \boldsymbol{U}_1}{\partial \boldsymbol{\alpha}} + n_k \frac{1}{\boldsymbol{U}_2}\frac{\partial \boldsymbol{U}_2}{\partial \boldsymbol{\alpha}} + \sum_{\substack{j=1\\ j\neq k}}^{m} n_j \frac{1}{\boldsymbol{U}_3}\frac{\partial \boldsymbol{U}_3}{\partial \boldsymbol{\alpha}} + \sum_{j=0}^{m} n_j \frac{1}{\boldsymbol{U}_4}\frac{\partial \boldsymbol{U}_4}{\partial \boldsymbol{\alpha}}, \text{ and}$$

$$\frac{\partial \ell}{\partial \gamma} = n_0 \frac{1}{\boldsymbol{U}_1}\frac{\partial \boldsymbol{U}_1}{\partial \gamma} + n_k \frac{1}{\boldsymbol{U}_2}\frac{\partial \boldsymbol{U}_2}{\partial \gamma} + \sum_{\substack{j=1\\ j\neq k}}^{m} n_j \frac{1}{\boldsymbol{U}_3}\frac{\partial \boldsymbol{U}_3}{\partial \gamma} + \sum_{j=0}^{m} n_j \frac{1}{\boldsymbol{U}_4}\frac{\partial \boldsymbol{U}_4}{\partial \gamma}.$$

Here,

$$\frac{\partial \boldsymbol{U}_1}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_i\Bigg(\exp(\boldsymbol{A}_i\boldsymbol{\alpha}) + \big(1-g(0|\tau_i,\xi_i)\big)\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta})\Bigg),$$

$$\frac{\partial \boldsymbol{U}_2}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_i\Bigg(\exp(\boldsymbol{A}_i\boldsymbol{\alpha}) + \big(1-g(k|\tau_i,\xi_i)\big)\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta})\Bigg),$$

$$\frac{\partial \boldsymbol{U}_3}{\partial \boldsymbol{\alpha}} = -\boldsymbol{A}_i g(j|\tau_i,\xi_i)\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta}), \qquad \frac{\partial \boldsymbol{U}_4}{\partial \boldsymbol{\alpha}} = \boldsymbol{A}_i\exp(\boldsymbol{A}_i\boldsymbol{\alpha})\big(1+\exp(\boldsymbol{B}_i\boldsymbol{\beta})\big),$$

$$\frac{\partial \boldsymbol{U}_1}{\partial \boldsymbol{\beta}} = \boldsymbol{B}_i\big(1-g(0|\tau_i,\xi_i)\big)\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta}),$$

$$\frac{\partial \boldsymbol{U}_2}{\partial \boldsymbol{\beta}} = \boldsymbol{B}_i\big(1-g(k|\tau_i,\xi_i)\big)\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta}),$$

$$\frac{\partial \boldsymbol{U}_3}{\partial \boldsymbol{\beta}} = -\boldsymbol{B}_i g(j|\tau_i,\xi_i)\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta}), \qquad \frac{\partial \boldsymbol{U}_4}{\partial \boldsymbol{\beta}} = \boldsymbol{B}_i\exp(\boldsymbol{B}_i\boldsymbol{\beta})\big(1+\exp(\boldsymbol{A}_i\boldsymbol{\alpha})\big),$$

$$\frac{\partial \boldsymbol{U}_1}{\partial \gamma} = \frac{\partial g(0|\tau_i,\xi_i)}{\partial \gamma}\Bigg(1-\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta})\Bigg),$$

$$\frac{\partial \boldsymbol{U}_2}{\partial \gamma} = \frac{\partial g(k|\tau_i,\xi_i)}{\partial \gamma}\Bigg(1-\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta})\Bigg),$$

$$\frac{\partial \boldsymbol{U}_3}{\partial \gamma} = \frac{\partial g(j|\tau_i,\xi_i)}{\partial \gamma}\Bigg(1-\exp(\boldsymbol{A}_i\boldsymbol{\alpha}+\boldsymbol{B}_i\boldsymbol{\beta})\Bigg), \qquad \text{and} \qquad \frac{\partial \boldsymbol{U}_4}{\partial \gamma} = 0.$$

The second order partial derivatives of the log-likelihood function are given by the following matrix:

$$\boldsymbol{H} = \begin{pmatrix} \dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha}^2} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}} & \dfrac{\partial^2 \ell}{\partial \gamma\partial \boldsymbol{\alpha}} \\ \dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\beta}} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2} & \dfrac{\partial^2 \ell}{\partial \gamma\partial \boldsymbol{\beta}} \\ \dfrac{\partial^2 \ell}{\partial \boldsymbol{\alpha}\partial \gamma} & \dfrac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \gamma} & \dfrac{\partial^2 \ell}{\partial \gamma^2} \end{pmatrix}$$

The covariance matrix is obtained by taking the inverse of negative of the Hessian matrix calculated at the ML estimates. Thus, the standard errors for parameter estimates $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\gamma})$ can be easily obtained by taking square root of the asymptotic variances. One can easily compute the ML estimates for the regression parameters using popular software packages. The ML estimation can be done for data consisting of raw counts as well as for grouped data. A more detailed discussion as well as application to DINB regression models will be pursued elsewhere in future.

CHAPTER VI

CONCLUSIONS

Traditionally, Poisson regression models have been commonly used for data consisting of count responses. For data consisting of inflated counts of zeros and inflated counts of k, we presented an in-depth study of two Doubly Inflated Poisson (DIP) models, DIP (p, λ) and DIP (p_1, p_2, λ) in this thesis. These two discrete mixture models can accommodate not only the inflated frequencies of count zero, but also inflated frequencies of count k. For both probability models, we discussed the distributional properties and two parameters estimation techniques, maximum likelihood and method of moments, for data consisting of raw counts as well as grouped frequencies. Efficiency comparisons show that ML estimators perform better than the moment estimators for both models. We also discussed DIP regression models using logistic and log-linear link functions to investigate the effects of covariates on the two inflated counts as well as Poisson counts. Maximum likelihood estimates for regression parameters were discussed in length for data consisting of raw counts as well as grouped frequencies. Applications to DIP models were illustrated using sample data on patient's length of stay in a hospital as well as on dental cavities. A brief introduction to Doubly Inflated Negative Binomial Distributions (DINB) was also presented in this thesis.

BIBLIOGRAPHY

Agresti, A. (2002), *Categorical Data Analysis*, New York: Wiley Series.

Bae, S., Famoye, F., Wulu, J. T. , Bartolucci, A. A., and Singh, K. P. (2005), "A Rich Family of Generalized Poisson Regression Models with Applications", *Mathematics and Computers in Simulations*, 69, 4-11.

Böhning, D., Dietz, E., and Schlattmann, P. (1997), "Zero-Inflated Count Models and their Applications in Public Health and Social Science", In: *Applications of Latent Trait and Latent Class Models in the Social Sciences*, 333-44.

Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L., and Kirchner, U. (1999), "The Zero-Inflated Poisson Model and the Decayed, Missing, and Filled Teeth Index in Dental Epidemiology", *Journal of the Royal Statistical Society*, 162, 195-209.

Cameron, A. C., and Trivedi, P. K. (1998), *Regression Analysis of Count Data*, New York: Cambridge.

Carrivick, P. J. W., Lee, A. H., and Yau, K. K. W. (2003), "Zero-Inflated Poisson Modeling to Evaluate Occupational Safety Interventions", *Safety Science*, 41, 53-63.

Casella, G., and Berger, R. L. (2002), *Statistical Inference*, Thompson Learning Inc.

Chaganty, N. R., and Shi, G. (2004), "A Note on the Estimation of Autocorrelation in Repeated Measurements", *Communication in Statistics - Theory and Methods*, 33, 1157-1170 .

Coxe, S., West, S., and Aiken, L. (2009), "The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternative", *Journal of Personality Assessment*, 91, 121-136.

Finkelman, M. D., Green J. G., Gruber, M. J., and Zaslavsky, A. M. (2009), "A Zero- and K-Inflated Mixture Model for Health Questionaire Data", *Statistics in Medicine*, 30, 1028-1043.

Greene, W. (2007), "Functional Forms for the Negative Binomial Model for Count Data", *Economics Letters*, 99, 585-590.

Hall, D. (2000), "Zero-Inflated Poisson and Binomial Regression with Random Effects: a Case Study", *Biometrics*, 56, 1030-1039.

Hall, D., and Shen, J. (2010), "Robust Estimation for Zero-Inflated Poisson Regression", *Scandinavian Journal of Statistics*, 37, 237-252.

Karazsia, B. T., and Dulmen, M. H. M.(2008), "Regression Models for Count Data: Illustrations using Longitudinal Predictors of Childhood Injury", *Journal of Pediatric Psychology*, 33, 1076-1084.

Lambert, D. (1992), "Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing", *Technometrics*, 34, 1-14 .

Lawless, J. (1987), "Negative Binomial and Mixed Poisson Regression", *The Canadian Journal of Statistics*, 15, 209-225.

Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation*, Springer.

Li, C., Lu, J., and Park, J. (1999), "Multivariate Zero-Inflated Poisson Models and Their Applications", *Technometrics*, 41, 29-38.

McCulloch, C. E., and Searle, S. R. (1989), *Generalized, Linear, and Mixed Models*, New York: Wiley & Sons.

Min, Y., and Agresti, A. (2005), "Random Effect Models for Repeated Measures of Zero-Inflated Count Data", *Statistical Modelling*, 5, 1-19.

Slymen, D. J., Ayala, G. X., Arredondo, E. M., and Elder, J. P. (2006), "A Demonstration of Modeling Count Data with an Application to Physical Activity", *Epidemiologic Perspectives and Innovations*, 3, 1-9.

Xiang, L., Lee, A. H., Yau,K. K. W., and McLachlan, G. J. (2007), "A Score Test for Overdispersion in Zero-Inflated Poisson Mixed Regression Model", *Statistics in Medicine*, 26, 1608-1622.

Yau, K. K. W., Wang, K., and Lee, A. H. (2003), "Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros", *Biometrical Journal*, 45, 437-452.

APPENDIX A

SELECTED SAS PROGRAMS

A.1 ML ESTIMATION

```
/*Loglikelihood Function Module*/;
start ll_fn(pr) global(y);
n = nrow(y); p1 = pr[1]; p2 = pr[2]; p3 = 1-p1-p2; lambda = pr[3];
f3=0.; f=0.; do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
if i = 1 then f1 = y[1]* log(p1 + p3*exp(-lambda));
else if i = 4 then f2 = y[4]*log(p2 + p3*poi);
else f3 = f3 + y[i]*log(p3*poi);
end; f = f1 + f2 + f3;
return(f); finish ll_fn;

/*Jacobian of Loglikelihood function Module*/;
start jacobian(pr) global(y);
n = nrow(y); p1 = pr[1]; p2 = pr[2]; p3=1 - p1 - p2; lambda=pr[3];
*derivative wrt p1; j3 = 0.; j=0;
do i = 1 to n;
poi = (exp(-lambda))*(lambda**(i-1))/fact(i-1);
if i = 1 then j1 = y[1]*(1-exp(-lambda))/(p1 + p3*exp(-lambda));
else if i = 4 then j2 = -1 * y[4]*poi/(p2 + p3*poi);
else j3 = j3 + (-1*y[i]/p3);
end; jp1 = j1 + j2 + j3;
*derivative wrt p2;
j6 = 0.; jp2 = 0.;
do i = 1 to n;
poi = (exp(-lambda))*(lambda**(i-1))/fact(i-1);
if i = 1 then j4 = -1*y[1]*exp(-lambda)/(p1 + p3*exp(-lambda));
else if i = 4 then j5 = y[4]*(1 - poi)/(p2 + p3*poi);
else j6 = j6 + (-1*y[i]/p3);
```

```
end; jp2 = j4 + j5 + j6;
*derivative wrt lambda;
jlambda = 0; j9 = 0;
do i = 1 to n;
poi = (exp(-lambda))*(lambda**(i-1))/fact(i-1);
if i = 1 then j7 = -1*y[1]*p3*exp(-lambda)/(p1 + p3*exp(-lambda));
else if i = 4 then j8 = y[4]*p3*poi*((i-1)/lambda - 1)/(p2 + p3*poi);
else j9 = j9 + y[i]*((i-1)/lambda - 1);
end; jlambda = j7 + j8 + j9;
j= jp1//jp2//jlambda;
return(j);
finish jacobian;

/*Hessian Matrix for loglikelihood Module*/
start hessian(pr) global(y);
p1 = pr[1]; p2 = pr[2]; lambda = pr[3]; p3 = 1 - p1 - p2;
n=nrow(y);
h11=j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
div1= p1 + p3*exp(-lambda);
div2=p2+p3*poi ;
if i = 1 then h11[1] = -1*y[1]*((1 - exp(-lambda))**2)/div1**2;
else if i = 4 then h11[4] = -1*y[4]*poi*poi/(div2**2);
else h11[i] = -1*y[i]/(p3**2);
end;
h12 = j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
div1 = p1 + p3*exp(-lambda);
div2 = p2+p3*poi ;
if i = 1 then h12[1] = y[1]*exp(-lambda)*(1 - exp(-lambda))/(div1**2);
else if i = 4 then h12[4] = y[4]*poi*(1 - poi)/(div2**2);
else h12[i]=-1*y[i]/(p3**2);
```

```
end;
h13 = j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
div1= p1 + p3*exp(-lambda);
div2=p2+p3*poi ;
if i = 1 then h13[1]=y[1]*exp(-lambda)*(p1 + p3)/(div1**2);
else if i = 4 then h13[4]=-1*y[4]*p2*((3/lambda) - 1)*poi/(div2**2);
else h13[i]=0;
end;
h21 = j(n,1,.);
h21 = h12;
h22 = j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
div1= p1 + p3*exp(-lambda);
div2=p2+p3*poi ;
if i = 1 then h22[1]=-1*y[1]*exp(-2*lambda)/(div1**2);
else if i = 4 then h22[4] = -1*y[4]*((1-poi)**2)/(div2**2);
else h22[i] = -1*y[i]/(p3**2);
end;
h23 = j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
div1= p1 + p3*exp(-lambda);
div2=p2+p3*poi ;
if i = 1 then h23[1]=y[1]*p1*exp(-lambda)/(div1**2);
else if i = 4 then h23[4]=y[4]*poi*((i-1)/lambda - 1)*(1
-p1)/(div2**2);
else h23[i]=0;
end;
h31 = j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
```

```
div1= p1 + p3*exp(-lambda);
div2=p2+p3*poi ;
if i = 1 then h31[1]=y[1]*exp(-lambda)*(p1+p3)/(div1**2);
else if i = 4 then h31[4]=-1*y[4]*p2*((i-1)/lambda - 1)*poi/(div2**2);
else h31[i]=0;
end;
h32=h23;
h33 = j(n,1,.);
do i = 1 to n;
poi = (exp(-lambda)) * (lambda**(i-1))/(fact(i-1));
div1= p1 + p3*exp(-lambda);
div2=p2+p3*poi ;
div3 = ((((i-1)/lambda - 1)**2)*p2 - (i-1)* div2/(lambda**2));
if i = 1 then h33[1] = -1*y[1]*p1*p3*exp(-lambda);
else if i = 4 then h33[4] = y[4]*p3*poi*div3/(div2**2);
else h33[i] = -1*y[i]*(i-1)/(lambda**2);
end;
H = (h11[+]||h12[+]||h13[+])//(h21[+]||h22[+]||h23[+])//
(h31[+]||h32[+]||h33[+]);
return(H);
finish hessian;

con = {0.001 0.001 1 . ., 0.999 0.999 . . ., 1 1  . -1 0.5};
*constraints on 0=<p1, p2=<1 and lambda > 0;
pr=j(1,3,.); pr[1]=0.1; pr[2]=0.1; pr[3]=1.1;
optn={1 9};
call nlpnra(rc, xr, "ll_fn", pr, optn, con) grd="gradient"
hes="hessian" ;
```

A.2 MOMENT ESTIMATION

```
start g_mom(theta) global(sample);
n=nrow(sample);
p1 = theta[1]; p2 = theta[2]; p3 = 1 - p1 - p2; lambda=theta[3];
```

```
k = 3; sampletotal = 0; sampletotal2 = 0; sampletotal3 = 0;
do i = 1 to n;
sumcount = (i-1)*sample[i]; sumcount2 = sample[i]*(i - 1)*(i-1);
sumcount3 = sample[i]*(i - 1)*(i-1)*(i-1);
sampletotal = sampletotal + sumcount; sampletotal2 = sampletotal2
+ sumcount2;
sampletotal3 = sampletotal3 + sumcount3;
end;
nbar1 = sampletotal/sample[+]; nbar2 = sampletotal2/sample[+];
nbar3 = sampletotal3/sample[+];
*print samplemean sampletotal;
f1 = k*p2 + p3*lambda - nbar1;
f2 = (k**2)*p2 + p3*lambda*(lambda + 1) - nbar2;
f3 = (k**3)*p2 + p3*lambda*(lambda**2 + 3*lambda +1) - nbar3;
g1 = 0.5*f1*f1`; g2 = 0.5*f2*f2`; g3 = 0.5*f3*f3`;
g = g1+g2+g3;
return(g);
finish g_mom;
```

A.3 RELATIVE EFFICIENCY

```
\* Relative Efficiencies of MOM vs MLE *\
proc iml ;
k = 14; n = 250;

create arenew2 var {p1 p2 p3 lambda I MLE_Cov
Det_D MOM_Cov are_p1 are_p2 are_lambda};
do p1 = 0.1 to 0.9 by 0.01; *p2 = 0.2; *lambda = 3.7;
do p2 = 0.1 to 0.9 by 0.01;
lambda = 7; *do lambda = 1 to 20 by 0.1;
p3 = 1 - p1 - p2;
poi = (exp(-lambda)) * (lambda**(k))/(fact(k));
deno1 = p1 + p3*exp(-lambda);
deno2 = p2 + p3*poi;
```

```
a = 1 - exp(-lambda) - poi;
b = (p2*((k/lambda)-1)**2) - deno2*k/(lambda**2);

/*Computing Variance Matrix for MLE estimators*/;
if p3 ^=0 then do;
I_11 = ((1 - exp(-lambda))**2)/deno1 + (poi*poi/deno2) + a/p3 ;
I_12 = -1*exp(-lambda)*(1-exp(-lambda))/deno1 - (poi*(1-poi)/deno2)
+ a/p3;
I_21 = I_12;
I_13 = (-1*exp(-lambda)*(p1+p3)/deno1) + p2*poi*((k/lambda)-1)/deno2;
I_31 = I_13;
I_22 = exp(-2*lambda)/deno1 + ((1 - poi)**2)/deno2 + a /p3;
I_23 = (-1*p1*exp(-lambda)/deno1) + poi*(1-p1)*((k/lambda)-1)/deno2;
I_32 = I_23;
I_33 = p1*p3*exp(-lambda)/deno1 - (p3*poi*b/deno2) + p3*a/(lambda**2);
I = (I_11 || I_12 || I_13)//(I_21 || I_22 || I_23)
//(I_31 || I_32 || I_33);
det_I = det(I);
if det_I ^=0 then do; MLE_Cov = (1/n)*inv(I); end;
else print p1 p2 lambda Det_I; end;

Exp_Y = p2*k + p3*lambda;
Exp_Y2 = p2*(k**2) + p3*(lambda**2 + lambda);
Exp_Y3 = p2*(k**3) + p3*(lambda**3 + 3*lambda**2 +lambda);
/*Computing Variance for MOM estimators of p */;
Det_D = -3*p3*(k**2)*(lambda**2) + 2*p3*k*(lambda**2)
- 2*p3*(k**2)*(lambda**3)+ p3*(lambda**2)*(k**3)
+ p3*k*(lambda**4);
D = j(3,3,.);
D[1,1]=-1*lambda;
D[1,2]=k-lambda;
D[1,3]=p3;
D[2,1]=-1*(lambda**2 + lambda);
D[2,2]=(k**2) - (lambda**2+lambda);
```

```
D[2,3]=2*p3*(lambda+1);
D[3,1]=-1*(lambda**3 +3*lambda**2+lambda);
D[3,2]=(k**3)-(lambda**3 +3*lambda**2+lambda);
D[3,3]=p3*(3*lambda**2 + 6*lambda+1);
if det(D) ^= 0 then do;
Sigma = j(3,3,0);
Sigma[1,1] = Exp_Y2 - (Exp_Y**2);
Sigma[1,2] = p2*(k**3) + p3*(lambda**3 + 3*lambda**2 + lambda)
 - (Exp_Y*Exp_Y2);
Sigma[1,3]= p2*(k**4) + p3*(lambda**4 + 6*lambda**3 + 7*lambda**2
+lambda);
Sigma[2,1] = Sigma[1,2];
Sigma[2,2] = p2*(k**4) + p3*(lambda**4 + 6*lambda**3 + 7*lambda**2
+lambda) - (Exp_Y2**2) ;
Sigma[2,3]= p2*(k**5) + p3*(lambda**5 + 10*lambda**4 + 25*lambda**3
+ 15*lambda**2 +lambda) - (Exp_Y2*Exp_Y3);
Sigma[3,1]= Sigma[1,3]; Sigma[3,2]= Sigma[2,3];
Sigma[3,3]= p2*(k**6) + p3*(lambda**6 + 15*lambda**5 + 65*lambda**4
+ 90*lambda**3 + 31*lambda**2 + lambda) - (Exp_Y3**2);
MOM_Cov = j(3,3,0); MOM_Cov = (1/n)*inv(D)*Sigma*inv(D`);
are_p1 = MLE_Cov[1,1] / MOM_Cov[1,1];
are_p2 = MLE_Cov[2,2]/MOM_Cov[2,2];
are_lambda = MLE_Cov[3,3] / MOM_Cov[3,3];
end; else  print p1 p2 lambda Det_I;

append;
*if Det_I = 0 then print p1 p2 lambda Det_I;
*if Det_D = 0 then print p1 p2 lambda Det_D;
end;  end; quit; run;
```

A.4 ML ESTIMATION USING PROC NLP AND NLMIXED

```
*****************************************************************
These are nice programs that were used to verify results obtained
```

```
using PROC IML, but can also be used to obtain ML Est for RegModels
******************************************************************
/* DIP 2 with covariates */
proc nlp data=new2;
max ll;
parameters b0=0, b1=0, b2=0, b3 = 0, b4=0, b5=0, b6=0, b7=0, b8=0,
a0=1, a1 = 1, a2 = 1, a3 = 1, a4=1, a5=1, a6=1, a7=1, a8=1,
g0=1, g1 = 1, g2 = 1, g3 = 1, g4=1, g5=1, g6=1, g7=1, g8=1;
linp1infl = a0 + a1*educ + a2*enrich + a3*rinse + a4*hyg + a5*all
+ a6*white + a7*black + a8*gender;
p1  = exp(-linp1infl)/(1+exp(-linp1infl));
linp2infl = g0 + g1*educ + g2*enrich + g3*rinse + g4*hyg + g5*all
+ g6*white + g7*black + g8*gender;
p2  = exp(-linp2infl)/(1+exp(-linp2infl)); p3 = 1 - p1 - p2;
lambda   = exp(b0 + b1*educ + b2*enrich + b3*rinse + b4*hyg + b5*all
+ b6*white + b7*black + b8*gender);
*bounds 0.01 < p1 < 0.99, 0.01< p2<0.99, lambda >1e-6;
*lincon p1+p2<=1;
if ddmft=0 then ll = log(p1 + p3*exp(-lambda));
   else if ddmft=1 then ll = log(p2 + p3*exp(-lambda)*lambda);
   else ll = log(p3) - lambda  + ddmft*log(lambda) -
   lgamma(ddmft + 1);
run;

proc nlmixed data=new2 maxiter=10000;
parameters b0=0, b1=0, b2=0, b3 = 0, b4=0, b5=0, b6=0, b7=0, b8=0,
a0=1, a1 = 1, a2 = 1, a3 = 1, a4=1, a5=1, a6=1, a7=1, a8=1,
g0=1, g1 = 1, g2 = 1, g3 = 1, g4=1, g5=1, g6=1, g7=1, g8=1, p3=0.5;
Aialpha = a0 + a1*educ + a2*enrich + a3*rinse + a4*hyg + a5*all +
a6*white + a7*black + a8*gender;
Gigamma = g0 + g1*educ + g2*enrich + g3*rinse + g4*hyg + g5*all +
g6*white + g7*black + g8*gender;
Bibeta = b0 + b1*educ + b2*enrich + b3*rinse + b4*hyg + b5*all +
b6*white + b7*black + b8*gender;
```

```
lambda   = exp(b0 + b1*educ + b2*enrich + b3*rinse + b4*hyg + b5*all
+ b6*white + b7*black + b8*gender);
   /* Build the DIP log likelihood */
   if dmft2=0 then ll = log(exp(Aialpha) + exp(-lambda)) + log(p3);
   else if dmft2=1
   then ll = ll = log(exp(Gigamma) + (1/2)*exp(Bibeta - lambda))
    + log(p3);
   else ll = log((1/lgamma(dmft2+1))*exp(dmft2*Bibeta - lambda))
    + log(p3);
   model dmft2 ~ general(ll);
run
```

VITA

Manasi Sheth-Chandra
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education/Professional Certifications

- Ph.D. in Applied and Computational Mathematics (Statistics), Old Dominion University, Norfolk, VA (December 2011)
- M.S. in Statistics, Old Dominion University, Norfolk, VA (May 2010)
- Master of Science, Mathematics, Loyola University, Chicago, IL (Dec 2003)
- Bachelor of Science, Mathematics and Biochemistry, Loyola University, Chicago, IL (May 2002)

Experience

- Teaching Assistant, Old Dominion University, Norfolk, VA (since Aug 2007)
- Data Analyst, Effective School-wide Discipline, VDOE (since Aug 2009)
- Mathematics Instructor, University of Wisconsin-Whitewater, WI (Aug 2004 to May 2007)

Awards

- Best Student Paper Presentation Award, Statistics Section, Virginia Academy of Sciences Meeting, May 2011
- Philip R. Wohl Scholarship Award, Department of Mathematics and Statistics, Old Dominion University, May 2010
- Modeling and Simulation Grant, Aug 2008 - June 2011

Typeset using $\LaTeX$.