Statistics and Its Interface Volume 9 (2016) 11–32

Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods

Chi Zhang, Guo-Liang Tian*, and Kai-Wang Ng

To model count data with excess zeros and excess ones, in their unpublished manuscript, Melkersson and Olsson (1999) extended the zero-inflated Poisson distribution to a *zero-and-one-inflated Poisson* (ZOIP) distribution. However, the distributional theory and corresponding properties of the ZOIP have not yet been explored, and likelihood-based inference methods for parameters of interest were not well developed. In this paper, we extensively study the ZOIP distribution by first constructing five equivalent stochastic representations for the ZOIP random variable and then deriving other important distributional properties. Maximum likelihood estimates of parameters are obtained by both the Fisher scoring and expectation—maximization algorithms. Bootstrap confidence intervals for parameters of interest and testing hypotheses under large sample sizes are provided. Simulations studies are performed and five real data sets are used to illustrate the proposed methods.

Keywords and phrases: EM algorithm, Fisher scoring algorithm, One-inflated Poisson distribution, Zero-inflated Poisson model, Zero-and-one inflated Poisson model.

1. INTRODUCTION

Poisson model provides a standard and popular framework to model count data. The over-dispersion problem emerges in modeling count data due to the greater incidence of zeros than that permitted by the traditional Poisson distribution. Lambert [17] proposed the *zero-inflated Poisson* (ZIP) regression model to handle count data with excess zeros. Some examples of ZIP distribution can be found in Neyman [22], Cohen [4], Singh [29], Martin and Katti [19], Kemp [16], and Mullahy [21]. Gupta *et al.* [11] developed a zero-adjusted generalized Poisson distribution including the ZIP distribution as a special case. Subsequently, Ridout *et al.* [24] reviewed some useful models for fitting count data with excess zeros. Carrivick *et al.* [3] applied the ZIP model to evaluate the effectiveness of a consultative manual handling workplace risk assessment team in reducing the risk of occupational injury among cleaners within a hospital.

*Corresponding author.

The over-dispersion may result from other phenomena such as a high proportion of both zeros and ones in the count data. For example, Eriksson and Åberg [8] reported a two-year panel data from Swedish Level of Living Surveys in 1974 and 1991, where the visits to a dentist have higher proportions of zeros and ones and one-visit observations are even much more frequent than zero-visits. It is common to visit a dentist for a routine control, e.g., school-children go to a dentist once a year almost as a rule. As the second example, Carrivick *et al.* [3] reported that much of the data collected on occupational safety involves accident or injury counts with many zeros and ones. In general, the safety regulations protect most workers from injury so that many zeros are observed. However, no perfect protection exists, such that there must be somebody suffering an accident or injury. This experience would be a signal to warn other workers to be much more cautious and to avoid more accidents. Hence there will be respectively more one observed indicating the majority of people having only one accident or none than those with two or more injuries.

Thus, the ZIP is no longer an appropriate distribution to model such count data with extra zeros and extra ones. As an extension of the ZIP, a so-called *zero-and-one-inflated Poisson* (ZOIP) distribution proposed by Melkersson and Olsson [20] is then a useful tool to capture the characteristics of such count data. The major goal of Melkersson and Olsson [20] is to fit the dentist visiting data with covariates in Sweden. Later, Saito and Rodrigues [25] presented a Bayesian analysis of the same dentist visiting data without considering any covariates by the data augmentation algorithm (Tanner and Wong, [30]). However, the distributional theory and corresponding properties of the ZOIP have not yet been explored, and likelihood-based methods for parameters of interest were not well developed, because, to our best knowledge, only two papers (i.e., Melkersson and Olsson, [20]; Saito and Rodrigues, [25]) involve the ZOIP to date. The main objective of this paper is to fill the gap.

For convenience, in this paper we denote a degenerate distribution at a single point c by $\xi \sim \text{Degenerate}(c)$, where ξ is a random variable with *probability mass function* (pmf) $\Pr(\xi = c) = 1$ and c is a constant.

Let $\xi_0 \sim \text{Degenerate}(0)$, $\xi_1 \sim \text{Degenerate}(1)$, $X \sim \text{Poisson}(\lambda)$ and they are independent. A discrete random variable Y is said to follow a ZOIP distribution, denoted by $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$, if its pmf is (Melkersson and Olsson, [20])

$$
\begin{aligned}
(1.1) \qquad & f(y|\phi_0, \phi_1; \lambda) \\
= \; & \phi_0 \Pr(\xi_0 = y) + \phi_1 \Pr(\xi_1 = y) + \phi_2 \Pr(X = y) \\
= \; & \begin{cases} \phi_0 + \phi_2 \mathrm{e}^{-\lambda}, & \text{if } y = 0, \\ \phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}, & \text{if } y = 1, \\ \phi_2 \dfrac{\lambda^y \mathrm{e}^{-\lambda}}{y!}, & \text{if } y = 2, 3, \ldots \end{cases} \\
= \; & (\phi_0 + \phi_2 \mathrm{e}^{-\lambda}) I(y = 0) + (\phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}) I(y = 1) \\
& + \left(\phi_2 \frac{\lambda^y \mathrm{e}^{-\lambda}}{y!} \right) I(y \geqslant 2),
\end{aligned}
$$

where $\phi_0 \in [0, 1)$ and $\phi_1 \in [0, 1)$ respectively denote the unknown proportions for incorporating extra-zeros and extra-ones than those allowed by the traditional Poisson distribution, and $\phi_2 \mathrel{\hat=} 1 - \phi_0 - \phi_1 \in (0, 1]$. The ZOIP distribution is a mixture of two degenerate distributions with all mass at zero and one, respectively and a Poisson(λ) distribution. In particular, when $\phi_0 = 0$, the ZOIP distribution reduces to *one-inflated Poisson* (OIP) distribution (denoted by $\text{OIP}(\phi_1, \lambda)$); when $\phi_1 = 0$, the ZOIP distribution reduces to ZIP distribution (denoted by $\text{ZIP}(\phi_0, \lambda)$); when $\phi_0 = \phi_1 = 0$, the ZOIP distribution becomes the traditional Poisson distribution.

The remainder of the paper is organized as follows. Section 2 presents five equivalent stochastic representations for the ZOIP random variable. Section 3 develops other distributional properties. In Section 4, we derive the Fisher scoring algorithm and the *expectation–maximization* (EM) algorithm for finding the MLEs of parameters. Bootstrap confidence intervals are also provided. The likelihood ratio test and the score test for one inflation, zero inflation and simultaneous zero-and-one inflation are developed in Section 5. Simulation studies to compare the likelihood ratio test with the score test are conducted in Section 6. In Section 7, five real data sets are used to illustrate the proposed methods. A discussion is given in Section 8.

2. VARIOUS STOCHASTIC REPRESENTATIONS AND THEIR EQUIVALENCE

In this section, we present five different *stochastic representation* (SR) for the random variable $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ and show their equivalence. These SRs reveal the relationship of the $\text{ZOIP}(\phi_0, \phi_1; \lambda)$ distribution with Degenerate(0), Degenerate(1), Poisson(λ), $\text{Bernoulli}(\phi_1/(\phi_0 + \phi_1))$, $\text{ZIP}(\phi_0/(1 - \phi_1), \lambda)$, ZTP($\lambda$) and OTP($\lambda$).

2.1 The first stochastic representation

Let $\mathbf{z} = (Z_0, Z_1, Z_2)^\top \sim \text{Multinomial}\,(1; \phi_0, \phi_1, \phi_2)$, $X \sim \text{Poisson}(\lambda)$, and $\mathbf{z}$ and X be mutually independent (denoted by $\mathbf{z} \perp\!\!\!\perp X$). It is easy to show that the first SR of the random variable $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ is given by

$$
\begin{aligned}
(2.1) \qquad Y \; & \stackrel{d}{=} \; Z_0 \cdot 0 + Z_1 \cdot 1 + Z_2 X = Z_1 + Z_2 X \\
& = \begin{cases} 0, & \text{with probability } \phi_0, \\ 1, & \text{with probability } \phi_1, \\ X, & \text{with probability } \phi_2, \end{cases}
\end{aligned}
$$

where the symbol "$\stackrel{d}{=}$" means that the random variables in both sides of the equality have the same distribution. The justification of this fact is as follows. By noting that $Z_0 + Z_1 + Z_2 = 1$ and $\Pr(Z_i = 1) = \phi_i$ for $i = 0, 1, 2$, from (2.1), we have

(2.2)
$$
\begin{cases}
\Pr(Y = 0) = \Pr(Z_0 = 1) + \Pr(Z_2 = 1, X = 0) \\
\qquad\qquad = \phi_0 + \phi_2 \mathrm{e}^{-\lambda}, \\
\Pr(Y = 1) = \Pr(Z_1 = 1) + \Pr(Z_2 = 1, X = 1) \\
\qquad\qquad = \phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}, \\
\Pr(Y = y) = \Pr(Z_2 = 1, X = y) = \phi_2 \dfrac{\lambda^y \mathrm{e}^{-\lambda}}{y!}, \; y \geqslant 2,
\end{cases}
$$

which is identical to (1.1), implying that $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$. The SR (2.1) means that $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ is a mixture of three distributions: Degenerate(0), Degenerate(1) and Poisson(λ).

2.2 The second stochastic representation

Alternatively, we can develop the second SR of the zero-and-one inflated Poisson random variable. Let $Z \sim \text{Bernoulli}(1 - \phi)$, $\eta \sim \text{Bernoulli}(p)$, $X \sim \text{Poisson}(\lambda)$ and (Z, η, X) be mutually independent. Then

(2.3)
$$
Y \stackrel{d}{=} (1 - Z)\eta + ZX = \begin{cases} \eta, & \text{with probability } \phi, \\ X, & \text{with probability } 1 - \phi \end{cases}
$$

follows the distribution $\text{ZOIP}(\phi_0, \phi_1; \lambda)$ with $\phi_0 = \phi(1 - p)$ and $\phi_1 = \phi p$. This fact can be verified as follows:

(2.4)
$$
\begin{cases}
\Pr(Y = 0) = \Pr(Z = 0, Y = 0) + \Pr(Z = 1, Y = 0) \\
\qquad\qquad = \Pr(Z = 0, \eta = 0) + \Pr(Z = 1, X = 0) \\
\qquad\qquad = \phi(1 - p) + (1 - \phi)\mathrm{e}^{-\lambda}, \\
\Pr(Y = 1) = \Pr(Z = 0, Y = 1) + \Pr(Z = 1, Y = 1) \\
\qquad\qquad = \Pr(Z = 0, \eta = 1) + \Pr(Z = 1, X = 1) \\
\qquad\qquad = \phi p + (1 - \phi)\lambda\, \mathrm{e}^{-\lambda}, \\
\Pr(Y = y) = \Pr(Z = 1, X = y) \\
\qquad\qquad = (1 - \phi)\lambda^y \mathrm{e}^{-\lambda}/y!, \qquad y \geqslant 2.
\end{cases}
$$

By comparing (2.4) with (2.2), we obtain a one-to-one mapping between (ϕ, p) and (ϕ_0, ϕ_1):

$$\begin{cases}\phi(1-p) &= \phi_0,\\ \phi p &= \phi_1,\\ 1-\phi &= \phi_2,\end{cases} \iff \begin{cases}\phi &= \phi_0+\phi_1,\\ p &= \dfrac{\phi_1}{\phi_0+\phi_1}.\end{cases} \tag{2.5}$$

The SR (2.3) indicates that $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ is also a mixture of a Bernoulli$(\phi_1/(\phi_0+\phi_1))$ distribution and a Poisson(λ) distribution.

2.3 The third stochastic representation

Now we consider the third SR of $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$. Let $Z \sim \text{Bernoulli}(1-\phi)$, $\xi_1 \sim \text{Degenerate}(1)$, $Y^* \sim \text{ZIP}(\phi^*, \lambda)$ and (Z, ξ_1, Y^*) be mutually independent. Then

$$Y \stackrel{d}{=} (1-Z)\xi_1 + ZY^* = \begin{cases} 1, & \text{with probability } \phi\\ Y^*, & \text{with probability } 1-\phi\end{cases} \tag{2.6}$$

follows the distribution $\text{ZOIP}(\phi_0, \phi_1; \lambda)$ with $\phi_0 = (1-\phi)\phi^*$ and $\phi_1 = \phi$. This fact can be shown as follows:

$$\begin{cases}\Pr(Y=0) &= \Pr(Z=1, Y^*=0)\\ &= (1-\phi)[\phi^* + (1-\phi^*)\mathrm{e}^{-\lambda}],\\ \Pr(Y=1) &= \Pr(Z=0) + \Pr(Z=1, Y^*=1)\\ &= \phi + (1-\phi)(1-\phi^*)\lambda\,\mathrm{e}^{-\lambda},\\ \Pr(Y=y) &= \Pr(Z=1, Y^*=y)\\ &= (1-\phi)(1-\phi^*)\lambda^y \mathrm{e}^{-\lambda}/y!, \qquad y \geqslant 2.\end{cases} \tag{2.7}$$

By comparing (2.7) with (2.2), we obtain a one-to-one mapping between (ϕ, ϕ^*) and (ϕ_0, ϕ_1):

$$\begin{cases}(1-\phi)\phi^* &= \phi_0,\\ \phi &= \phi_1,\\ (1-\phi)(1-\phi^*) &= \phi_2,\end{cases} \iff \begin{cases}\phi &= \phi_1,\\ \phi^* &= \dfrac{\phi_0}{1-\phi_1}.\end{cases} \tag{2.8}$$

The SR (2.6) shows that $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ is also a mixture of a degenerate distribution with all mass at 1 and a $\text{ZIP}(\phi^*, \lambda)$ distribution.

2.4 The fourth stochastic representation

To derive the fourth SR of $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$, we first introduce the *zero-truncated Poisson* (ZTP) distribution (David and Johnson, [6]), denoted by $V_0 \sim \text{ZTP}(\lambda)$, whose pmf is defined as

$$\Pr(V_0 = v) = \frac{\lambda^v \mathrm{e}^{-\lambda}/v!}{1-\mathrm{e}^{-\lambda}}, \qquad v = 1, 2, \ldots, \infty. \tag{2.9}$$

Now let $\mathbf{z}^* = (Z_0^*, Z_1^*, Z_2^*)^\top \sim \text{Multinomial}\,(1; \phi_0^*, \phi_1^*, \phi_2^*)$, $V_0 \sim \text{ZTP}(\lambda)$ and $\mathbf{z}^* \perp\!\!\!\perp V_0$. Then

$$Y \stackrel{d}{=} Z_1^* + Z_2^* V_0 = \begin{cases}0, & \text{with probability } \phi_0^*,\\ 1, & \text{with probability } \phi_1^*,\\ V_0, & \text{with probability } \phi_2^*\end{cases} \tag{2.10}$$

follows the distribution $\text{ZOIP}(\phi_0, \phi_1; \lambda)$ with $\phi_0 = \phi_0^* - \phi_2^*/(\mathrm{e}^\lambda - 1)$ and $\phi_1 = \phi_1^*$. This fact can be verified as follows:

$$\begin{cases}\Pr(Y=0) &= \Pr(Z_0^*=1) = \phi_0^*,\\ \Pr(Y=1) &= \Pr(Z_1^*=1) + \Pr(Z_2^*=1, V_0=1)\\ &= \phi_1^* + \phi_2^*\lambda\,\mathrm{e}^{-\lambda}/(1-\mathrm{e}^{-\lambda}),\\ \Pr(Y=y) &= \Pr(Z_2^*=1, V_0=y)\\ &= \phi_2^*\lambda^y \mathrm{e}^{-\lambda}/[y!(1-\mathrm{e}^{-\lambda})], \qquad y \geqslant 2.\end{cases} \tag{2.11}$$

By comparing (2.11) with (2.2), we obtain a one-to-one mapping between $(\phi_0^*, \phi_1^*, \phi_2^*)$ and (ϕ_0, ϕ_1, ϕ_2):

$$\begin{cases}\phi_0^* &= \phi_0 + \phi_2\mathrm{e}^{-\lambda},\\ \phi_1^* &= \phi_1,\\ \phi_2^* &= \phi_2(1-\mathrm{e}^{-\lambda}),\end{cases} \iff \begin{cases}\phi_0 &= \phi_0^* - \dfrac{\phi_2^*}{\mathrm{e}^\lambda - 1},\\ \phi_1 &= \phi_1^*,\\ \phi_2 &= \phi_2^*\mathrm{e}^\lambda/(\mathrm{e}^\lambda - 1).\end{cases} \tag{2.12}$$

The SR (2.10) shows that $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ is also a mixture of three distributions: Degenerate(0), Degenerate(1) and ZTP(λ).

2.5 The fifth stochastic representation

First of all, we introduce the *one-truncated Poisson* (OTP) distribution (Cohen, [5]; Hilbe, [13], p. 388) with pmf defined by

$$\Pr(V_1 = v) = \frac{\lambda^v \mathrm{e}^{-\lambda}/v!}{1-\mathrm{e}^{-\lambda} - \lambda\,\mathrm{e}^{-\lambda}}, \qquad v = 2, 3, \ldots, \infty, \tag{2.13}$$

and denote it by $V_1 \sim \text{OTP}(\lambda)$. Now let $\mathbf{z}^* = (Z_0^*, Z_1^*, Z_2^*)^\top \sim \text{Multinomial}\,(1; \phi_0^*, \phi_1^*, \phi_2^*)$, $V_1 \sim \text{OTP}(\lambda)$ and $\mathbf{z}^* \perp\!\!\!\perp V_1$. Then

$$Y \stackrel{d}{=} Z_1^* + Z_2^* V_1 = \begin{cases}0, & \text{with probability } \phi_0^*,\\ 1, & \text{with probability } \phi_1^*,\\ V_1, & \text{with probability } \phi_2^*\end{cases} \tag{2.14}$$

follows the distribution $\text{ZOIP}(\phi_0, \phi_1; \lambda)$ with $\phi_0 = \phi_0^* - \phi_2^*/(\mathrm{e}^\lambda - 1 - \lambda)$ and $\phi_1 = \phi_1^* - \phi_2^*\lambda/(\mathrm{e}^\lambda - 1 - \lambda)$. This fact can be verified as follows:

(2.15)
$$\begin{cases} \Pr(Y=0) &= \Pr(Z_0^*=1)=\phi_0^*, \\ \Pr(Y=1) &= \Pr(Z_1^*=1)=\phi_1^*, \\ \Pr(Y=y) &= \Pr(Z_2^*=1, V_1=y) \\ &= \phi_2^*\lambda^y \mathrm{e}^{-\lambda}/[y!(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda})],\ y\geqslant 2. \end{cases}$$

By comparing (2.15) with (2.2), we obtain a one-to-one mapping between $(\phi_0^*, \phi_1^*, \phi_2^*)$ and (ϕ_0, ϕ_1, ϕ_2):

$$\text{(2.16)}\qquad \begin{cases} \phi_0^* &= \phi_0+\phi_2\mathrm{e}^{-\lambda}, \\ \phi_1^* &= \phi_1+\phi_2\lambda\,\mathrm{e}^{-\lambda}, \\ \phi_2^* &= \phi_2(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}), \end{cases}$$
$$\Longleftrightarrow \begin{cases} \phi_0 &= \phi_0^*-\phi_2^*/(\mathrm{e}^{\lambda}-1-\lambda), \\ \phi_1 &= \phi_1^*-\phi_2^*\lambda/(\mathrm{e}^{\lambda}-1-\lambda), \\ \phi_2 &= \phi_2^*\mathrm{e}^{\lambda}/(\mathrm{e}^{\lambda}-1-\lambda). \end{cases}$$

The SR (2.14) shows that $Y\sim \mathrm{ZOIP}(\phi_0,\phi_1;\lambda)$ is also a mixture of three distributions: Degenerate(0), Degenerate(1) and OTP(λ).

3. DISTRIBUTIONAL PROPERTIES

3.1 The cumulative distribution function

Let $Y\sim \mathrm{ZOIP}(\phi_0,\phi_1;\lambda)$. For any non-negative real number y, the cumulative distribution function of Y is given by

$$\begin{aligned} \text{(3.1)}\quad & \Pr(Y\leqslant y) \\ =\ & (\phi_0+\phi_2\mathrm{e}^{-\lambda})I(0\leqslant y<1) \\ & + \left(\phi_0+\phi_1+\phi_2\sum_{i=0}^{\lfloor y\rfloor}\frac{\lambda^i\mathrm{e}^{-\lambda}}{i!}\right)I(y\geqslant 1) \\ =\ & (\phi_0+\phi_2\mathrm{e}^{-\lambda})I(0\leqslant y<1) \\ & + \left[\phi_0+\phi_1+\phi_2\frac{\Gamma(\lfloor y+1\rfloor,\lambda)}{\lfloor y\rfloor!}\right]I(y\geqslant 1), \end{aligned}$$

where $\lfloor k\rfloor$ denotes the largest integer not greater than k, and

$$\text{(3.2)}\qquad \Gamma(k,\lambda)\,\hat{=}\int_{\lambda}^{\infty} t^{k-1}\mathrm{e}^{-t}\,\mathrm{d}t$$

is the upper incomplete gamma function.

3.2 Moments

The SR (2.1) is a useful tool to derive the moments of the ZOIP random variable. Since $(Z_0,Z_1,Z_2)^{\top}\sim$ Multinomial$\,(1;\phi_0,\phi_1,\phi_2)$, we have $E(Z_i)=\phi_i$, $\mathrm{Var}(Z_i)=\phi_i(1-\phi_i)$, $\mathrm{Cov}(Z_i,Z_j)=-\phi_i\phi_j$ and $E(Z_iZ_j)=0$ for $i\neq j$. From (2.1), we immediately obtain

$$\text{(3.3)}\qquad \begin{aligned} E(Y) &= \phi_1+\phi_2\lambda\ \hat{=}\ \mu, \\ E(Y^2) &= \phi_1+\phi_2(\lambda+\lambda^2), \\ \mathrm{Var}(Y) &= \mu-\mu^2+(\mu-\phi_1)^2/\phi_2. \end{aligned}$$

In general, for any positive integers r and s, we have $Z_i^rZ_j^s\sim$ Degenerate(0) for $i\neq j$. Let n be an arbitrary positive integer, we have

$$\text{(3.4)}\qquad \begin{aligned} E(Y^n) &= \sum_{k=0}^{n}\binom{n}{k}E(Z_1^kZ_2^{n-k})E(X^{n-k}) \\ &= \phi_1+\phi_2E(X^n). \end{aligned}$$

Alternatively, the SR (2.3) can be applied to calculate $E(Y^n)$, and we can obtain the same result as (3.4) by noting that $(1-Z)^rZ^s\sim$ Degenerate(0) for any Bernoulli random variable Z, where r and s are two arbitrary positive integers.

3.3 Moment generating function

By using the formula of $E(W_1)=E[E(W_1|W_2)]$, the moment generating function of $Y\sim \mathrm{ZOIP}(\phi_0,\phi_1;\lambda)$ is given by

$$\begin{aligned} \text{(3.5)}\qquad & M_Y(t)=E[\exp(tY)] \\ \overset{(2.3)}{=}\ & E\Big\{\exp[t(1-Z)\eta+tZX]\Big\} \\ =\ & E\Big\{E\{\exp[t(1-Z)\eta+tZX]\,|\,Z\}\Big\} \\ =\ & E[M_{\eta}(t(1-Z))\cdot M_X(tZ)] \\ =\ & E\Big\{[p\,\mathrm{e}^{t(1-Z)}+1-p]\cdot\exp[\lambda(\mathrm{e}^{tZ}-1)]\Big\} \\ =\ & \phi(p\,\mathrm{e}^{t}+1-p)+(1-\phi)\exp[\lambda(\mathrm{e}^{t}-1)]. \end{aligned}$$

3.4 Conditional distributions based on the first SR

In Section 2.1, we assumed that $\mathbf{z}=(Z_0,Z_1,Z_2)^{\top}\sim$ Multinomial$\,(1;\phi_0,\phi_1,\phi_2)$ and $Y\sim \mathrm{ZOIP}(\phi_0,\phi_1;\lambda)$. It is clear that $\mathbf{z}$ only takes one of the three base-vectors $(1,0,0)^{\top}$, $(0,1,0)^{\top}$ and $(0,0,1)^{\top}$. We first consider the joint conditional distribution of $\mathbf{z}|Y$, which is stated as follows.

Theorem 1. (Joint conditional distribution of $\mathbf{z}|Y$). Let $\mathbf{z}\sim$ Multinomial$\,(1;\phi_0,\phi_1,\phi_2)$ and $Y\sim \mathrm{ZOIP}(\phi_0,\phi_1;\lambda)$. Based on the SR (2.1), we have

(3.6)
$$\mathbf{z}|(Y=y)\sim \begin{cases} \text{Multinomial}\,(1;\psi_1,0,1-\psi_1), & \text{if } y=0, \\ \text{Multinomial}\,(1;0,\psi_2,1-\psi_2), & \text{if } y=1, \\ \text{Multinomial}\,(1;0,0,1), & \text{if } y\geqslant 2, \end{cases}$$

where

$$\text{(3.7)}\qquad \psi_1\,\hat{=}\,\frac{\phi_0}{\phi_0+\phi_2\mathrm{e}^{-\lambda}}\quad\text{and}\quad \psi_2\,\hat{=}\,\frac{\phi_1}{\phi_1+\phi_2\lambda\,\mathrm{e}^{-\lambda}}.$$

Proof. The joint conditional pmf of $\mathbf{z}|(Y=y)$ is given by

$$\Pr(\mathbf{z}=\boldsymbol{z}|Y=y)=\frac{\Pr(Z_0=z_0,Z_1=z_1,Z_2=z_2,Y=y)}{f(y|\phi_0,\phi_1;\lambda)}.$$

If $y = 0$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0)^\top \,|\, Y = 0\} \overset{(1.1)}{=} \dfrac{\phi_0}{\phi_0 + \phi_2 \mathrm{e}^{-\lambda}} \overset{(3.7)}{=} \psi_1, \\ \Pr\{\mathbf{z} = (0,1,0)^\top \,|\, Y = 0\} = 0, \\ \Pr\{\mathbf{z} = (0,0,1)^\top \,|\, Y = 0\} = \dfrac{\phi_2 \mathrm{e}^{-\lambda}}{\phi_0 + \phi_2 \mathrm{e}^{-\lambda}} = 1 - \psi_1, \end{cases}$$

which imply the first assertion of (3.6).

If $y = 1$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0)^\top \,|\, Y = 1\} = 0, \\ \Pr\{\mathbf{z} = (0,1,0)^\top \,|\, Y = 1\} \overset{(1.1)}{=} \dfrac{\phi_1}{\phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}} \\ \qquad \overset{(3.7)}{=} \psi_2, \\ \Pr\{\mathbf{z} = (0,0,1)^\top \,|\, Y = 1\} = \dfrac{\phi_2 \lambda\, \mathrm{e}^{-\lambda}}{\phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}} \\ \qquad = 1 - \psi_2, \end{cases}$$

which imply the second assertion of (3.6).

If $y \geqslant 2$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0)^\top \,|\, Y = y\} = 0, \\ \Pr\{\mathbf{z} = (0,1,0)^\top \,|\, Y = y\} = 0, \\ \Pr\{\mathbf{z} = (0,0,1)^\top \,|\, Y = y\} \overset{(1.1)}{=} \dfrac{\phi_2 \lambda^y \mathrm{e}^{-\lambda}/y!}{\phi_2 \lambda^y \mathrm{e}^{-\lambda}/y!} = 1, \end{cases}$$

implying the last assertion of (3.6). □

Next, we consider the marginal conditional distributions of $Z_i|Y$ for $i = 0, 1, 2$. As a corollary of Theorem 1, we summarize these results as follows.

Corollary 1. (Marginal conditional distributions of $Z_i|Y$). Let $\mathbf{z} \sim \text{Multinomial}(1; \phi_0, \phi_1, \phi_2)$ and $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$. Based on the SR (2.1), we have

$$(3.8) \quad Z_0|(Y = y) \sim \begin{cases} \text{Bernoulli}(\psi_1), & \text{if } y = 0, \\ \text{Degenerate}(0), & \text{if } y \neq 0, \end{cases}$$

$$(3.9) \quad Z_1|(Y = y) \sim \begin{cases} \text{Bernoulli}(\psi_2), & \text{if } y = 1, \\ \text{Degenerate}(0), & \text{if } y \neq 1, \end{cases}$$

$$(3.10) \quad Z_2|(Y = y) \sim \begin{cases} \text{Bernoulli}(1-\psi_1), & \text{if } y = 0, \\ \text{Bernoulli}(1-\psi_2), & \text{if } y = 1, \\ \text{Degenerate}(1), & \text{if } y \geqslant 2, \end{cases}$$

where ψ_1 and ψ_2 are given by (3.7).

Finally, based on the SR (2.1), we discuss the conditional distribution of $X|Y$, which is stated in the following theorem.

Theorem 2. (Conditional distribution of $X|Y$). Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$. Based on the SR (2.1), we have

$$(3.11) \qquad X|(Y = y) \sim \begin{cases} \text{ZIP}(1-\psi_1, \lambda), & \text{if } y = 0, \\ \text{OIP}(1-\psi_2, \lambda), & \text{if } y = 1, \\ \text{Degenerate}(y), & \text{if } y \geqslant 2, \end{cases}$$

where ψ_1 and ψ_2 are given by (3.7).

Proof. If $y = 0$, we have

$$\begin{aligned} \Pr(X = x|Y = 0) &= \frac{\Pr(X = x, Y = 0)}{\Pr(Y = 0)} \\ &= \frac{\Pr(X = 0, Z_1 = 0)}{f(0|\phi_0, \phi_1; \lambda)} I(x = 0) \\ &\quad + \frac{\Pr(X = x, Z_0 = 1)}{f(0|\phi_0, \phi_1; \lambda)} I(x \neq 0) \\ &\overset{(1.1)}{=} \frac{(1-\phi_1)\mathrm{e}^{-\lambda}}{\phi_0 + \phi_2 \mathrm{e}^{-\lambda}} I(x = 0) + \frac{\phi_0 \lambda^x \mathrm{e}^{-\lambda}/x!}{\phi_0 + \phi_2 \mathrm{e}^{-\lambda}} I(x \neq 0) \\ &\overset{(3.7)}{=} (1 - \psi_1 + \psi_1 \mathrm{e}^{-\lambda}) I(x = 0) + \left(\psi_1 \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!}\right) I(x \neq 0), \end{aligned}$$

implying $X|(Y = 0) \sim \text{ZIP}(1 - \psi_1, \lambda)$.

If $y = 1$, we have

$$\begin{aligned} \Pr(X = x|Y = 1) &= \frac{\Pr(X = x, Y = 1)}{\Pr(Y = 1)} \\ &= \frac{\Pr(X = 1, Z_0 = 0)}{f(1|\phi_0, \phi_1; \lambda)} I(x = 1) \\ &\quad + \frac{\Pr(X = x, Z_1 = 1)}{f(1|\phi_0, \phi_1; \lambda)} I(x \neq 1) \\ &\overset{(1.1)}{=} \frac{(1-\phi_0)\lambda\,\mathrm{e}^{-\lambda}}{\phi_1 + \phi_2 \lambda\,\mathrm{e}^{-\lambda}} I(x = 1) + \frac{\phi_1 \lambda^x \mathrm{e}^{-\lambda}/x!}{\phi_1 + \phi_2 \lambda\,\mathrm{e}^{-\lambda}} I(x \neq 1) \\ &\overset{(3.7)}{=} (1 - \psi_2 + \psi_2 \lambda\,\mathrm{e}^{-\lambda}) I(x = 1) + \left(\psi_2 \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!}\right) I(x \neq 1), \end{aligned}$$

implying $X|(Y = 1) \sim \text{OIP}(1 - \psi_2, \lambda)$.

If $y \geqslant 2$, we have

$$\begin{aligned} \Pr(X = x|Y = y) &= \frac{\Pr(X = x, Y = y)}{f(y|\phi_0, \phi_1; \lambda)} \\ &\overset{(1.1)}{=} \frac{\Pr(X = y, Z_2 = 1)}{\phi_2 \lambda^y \mathrm{e}^{-\lambda}/y!} = 1, \end{aligned}$$

implying $X|(Y = y \geqslant 2) \sim \text{Degenerate}(y)$. □

3.5 Conditional distributions based on the second SR

In Section 2.2, we assumed that $Z \sim \text{Bernoulli}(1 - \phi)$, $\eta \sim \text{Bernoulli}(p)$, $X \sim \text{Poisson}(\lambda)$ and (Z, η, X) are mutually independent. According to (2.4), the pmf of $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$ can be rewritten as

$$(3.12) \quad f(y|\phi, p; \lambda) = \left[\phi(1-p) + (1-\phi)\mathrm{e}^{-\lambda}\right] I(y = 0)$$

$$+ \left[\phi p + (1-\phi)\lambda\, \mathrm{e}^{-\lambda}\right] I(y=1)$$
$$+ \left[(1-\phi)\lambda^y \mathrm{e}^{-\lambda}/y!\right] I(y \geqslant 2).$$

Based on the SR (2.3), we will derive the conditional distributions of $Z|Y$, $\eta|Y$ and $X|Y$, which are given in Theorems 3, 4 and 5, respectively.

Theorem 3. (Conditional distribution of $Z|Y$). Let $Z \sim$ Bernoulli$(1-\phi)$ and the pmf of Y be specified by (3.12). Based on the SR (2.3), we have

$$Z|(Y=y) \sim \begin{cases} \text{Bernoulli}(\varphi_1), & \text{if } y=0, \\ \text{Bernoulli}(\varphi_2), & \text{if } y=1, \\ \text{Degenerate}(1), & \text{if } y \geqslant 2, \end{cases} \tag{3.13}$$

where

$$\begin{cases} \varphi_1 \;\hat{=}\; \dfrac{(1-\phi)\mathrm{e}^{-\lambda}}{\phi(1-p)+(1-\phi)\mathrm{e}^{-\lambda}} = 1-\psi_1, \\ \varphi_2 \;\hat{=}\; \dfrac{(1-\phi)\lambda\, \mathrm{e}^{-\lambda}}{\phi p+(1-\phi)\lambda\, \mathrm{e}^{-\lambda}} = 1-\psi_2, \end{cases} \tag{3.14}$$

and (ψ_1, ψ_2) are given by (3.7).

Proof. Since $Z \sim$ Bernoulli$(1-\phi)$, Z only takes the value 0 or 1. By the SR (2.3), we have

$$\begin{aligned} & \Pr(Z=1|Y=y) \\ = \; & \frac{\Pr(Z=1, X=y)}{f(y|\phi,p;\lambda)} \\ = \; & \frac{(1-\phi)\lambda^y \mathrm{e}^{-\lambda}/y!}{f(y|\phi,p;\lambda)} \\ \overset{(3.12)}{=} & \begin{cases} \dfrac{(1-\phi)\mathrm{e}^{-\lambda}}{\phi(1-p)+(1-\phi)\mathrm{e}^{-\lambda}} \overset{(3.14)}{=} \varphi_1, & \text{if } y=0, \\ \dfrac{(1-\phi)\lambda\, \mathrm{e}^{-\lambda}}{\phi p+(1-\phi)\lambda\, \mathrm{e}^{-\lambda}} \overset{(3.14)}{=} \varphi_2, & \text{if } y=1, \\ 1, & \text{if } y \geqslant 2, \end{cases} \end{aligned}$$

which implies (3.13). In addition, by applying (2.5), we can immediately verify that $\varphi_1 = 1-\psi_1$ and $\varphi_2 = 1-\psi_2$. □

Theorem 4. (Conditional distribution of $\eta|Y$). Let $\eta \sim$ Bernoulli(p) and the pmf of Y be specified by (3.12). Based on the SR (2.3), we have

$$\eta|(Y=y) \sim \begin{cases} \text{Bernoulli}(p\varphi_1), & \text{if } y=0, \\ \text{Bernoulli}(1-\varphi_2+p\varphi_2), & \text{if } y=1, \\ \text{Bernoulli}(p), & \text{if } y \geqslant 2, \end{cases} \tag{3.15}$$

where φ_1 and φ_2 are given by (3.14).

Proof. Since $\eta \sim$ Bernoulli(p), η only takes the value 0 or 1. By the SR (2.3), we have

$$\begin{aligned} & \Pr(\eta=1|Y=y) \\ = \; & \frac{\Pr(\eta=1, 1-Z+ZX=y)}{f(y|\phi,p;\lambda)} \\ \overset{(3.12)}{=} & \begin{cases} \dfrac{p\cdot(1-\phi)\mathrm{e}^{-\lambda}}{\phi(1-p)+(1-\phi)\mathrm{e}^{-\lambda}} \overset{(3.14)}{=} p\varphi_1, \\ \qquad \text{if } y=0, \\ \dfrac{p[\phi+(1-\phi)\lambda\, \mathrm{e}^{-\lambda}]}{\phi p+(1-\phi)\lambda\, \mathrm{e}^{-\lambda}} \overset{(3.14)}{=} 1-\varphi_2+p\varphi_2, \\ \qquad \text{if } y=1, \\ \dfrac{p\cdot(1-\phi)\lambda^y\mathrm{e}^{-\lambda}/y!}{(1-\phi)\lambda^y\mathrm{e}^{-\lambda}/y!}, = p \\ \qquad \text{if } y \geqslant 2, \end{cases} \end{aligned}$$

which implies (3.15). □

By combining Theorem 2 with (3.14), we immediately obtain the following results.

Theorem 5. (Conditional distribution of $X|Y$). Let $X \sim$ Poisson(λ) and the pmf of Y be specified by (3.12). Based on the SR (2.3), we have

$$X|(Y=y) \sim \begin{cases} \text{ZIP}(\varphi_1, \lambda), & \text{if } y=0, \\ \text{OIP}(\varphi_2, \lambda), & \text{if } y=1, \\ \text{Degenerate}(y), & \text{if } y \geqslant 2, \end{cases} \tag{3.16}$$

where φ_1 and φ_2 are given by (3.14).

4. LIKELIHOOD-BASED INFERENCES

Assume that $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{ZOIP}(\phi_0, \phi_1; \lambda)$ and $y_1, \ldots, y_n$ denote their realizations. Let$Y_{\text{obs}} = \{y_i\}_{i=1}^n$ denote the observed data. Moreover, let $\mathbb{I}_0 \,\hat{=}\, \{i|y_i=0,\ 1 \leqslant i \leqslant n\}$ and $m_0 = \sum_{i=1}^n I(y_i=0)$ denote the number of elements in $\mathbb{I}_0$; $\mathbb{I}_1 \,\hat{=}\, \{i|y_i=1,\ 1 \leqslant i \leqslant n\}$ and $m_1 = \sum_{i=1}^n I(y_i=1)$ denote the number of elements in $\mathbb{I}_1$. The observed-data likelihood function for $(\phi_0, \phi_1, \lambda)$ is then given by

$$\begin{aligned} & L(\phi_0, \phi_1, \lambda|Y_{\text{obs}}) \\ = \; & (\phi_0+\phi_2\mathrm{e}^{-\lambda})^{m_0} \times (\phi_1+\phi_2\lambda\,\mathrm{e}^{-\lambda})^{m_1} \\ & \times \phi_2^{n-m_0-m_1} \prod_{i \notin \mathbb{I}_0 \cup \mathbb{I}_1} \frac{\lambda^{y_i}\mathrm{e}^{-\lambda}}{y_i!}, \end{aligned} \tag{4.1}$$

so that the log-likelihood function is

$$\begin{aligned} \ell \;\hat{=}\; & \ell(\phi_0, \phi_1, \lambda|Y_{\text{obs}}) \\ = \; & m_0 \log(\phi_0+\phi_2\mathrm{e}^{-\lambda}) + m_1\log(\phi_1+\phi_2\lambda\,\mathrm{e}^{-\lambda}) \\ & + (n-m_0-m_1)(\log\phi_2 - \lambda) + N\log\lambda, \end{aligned}$$

where $\phi_2 = 1-\phi_0-\phi_1$ and $N = \sum_{i \notin \mathbb{I}_0 \cup \mathbb{I}_1} y_i$. To calculate the Fisher information matrix, we need the following results.

Theorem 6. (Expectations). The expectations of m_0, m_1 and N are given by

$$E(m_0) = n(\phi_0+\phi_2\mathrm{e}^{-\lambda}), \tag{4.2}$$

$$E(m_1) = n(\phi_1+\phi_2\lambda\,e^{-\lambda}),$$
$$E(N) = n\phi_2\lambda(1-e^{-\lambda}).$$

Proof. It is easy to show the expressions of $E(m_0)$ and $E(m_1)$, and we have two methods to verify the last one. The first way is to directly calculate $E(N)$ as follows. Since

$$E(Y_i|Y_i \geqslant 2) = \frac{\lambda(1-e^{-\lambda})}{1-e^{-\lambda}-\lambda\,e^{-\lambda}}, \qquad i \notin \mathbb{I}_0 \cup \mathbb{I}_1,$$

we have

$$\begin{aligned} E(N) &= \big[E(N|m_0,m_1)\big] \\ &= E\left[\frac{(n-m_0-m_1)\lambda(1-e^{-\lambda})}{1-e^{-\lambda}-\lambda\,e^{-\lambda}}\right] \\ &= \frac{[n-E(m_0)-E(m_1)]\lambda(1-e^{-\lambda})}{1-e^{-\lambda}-\lambda\,e^{-\lambda}} \\ &\overset{(4.2)}{=} n\phi_2\lambda(1-e^{-\lambda}). \end{aligned}$$

Alternatively, we noted that

$$N = \sum_{i\notin\mathbb{I}_0\cup\mathbb{I}_1} y_i = \sum_{i=1}^{n} y_i - m_1. \tag{4.3}$$

Thus, $E(N) = nE(Y_1) - E(m_1) \overset{(3.3)}{=} n(\phi_1+\phi_2\lambda) - n(\phi_1+\phi_2\lambda\,e^{-\lambda}) = n\phi_2\lambda(1-e^{-\lambda})$. □

4.1 MLEs via the Fisher scoring algorithm

In this section, we use the Fisher scoring algorithm to calculate the MLEs of ϕ_0, ϕ_1 and λ. The score vector $\nabla\ell$ and the Hessian matrix $\nabla^2\ell$ are given by

$$\nabla\ell(\phi_0,\phi_1,\lambda|Y_{\rm obs}) = \left(\frac{\partial\ell}{\partial\phi_0}, \frac{\partial\ell}{\partial\phi_1}, \frac{\partial\ell}{\partial\lambda}\right)^{\top} \quad \text{and}$$

$$\nabla^2\ell(\phi_0,\phi_1,\lambda|Y_{\rm obs}) = \begin{pmatrix} \frac{\partial^2\ell}{\partial\phi_0^2} & \frac{\partial^2\ell}{\partial\phi_0\partial\phi_1} & \frac{\partial^2\ell}{\partial\phi_0\partial\lambda} \\ \frac{\partial^2\ell}{\partial\phi_1\partial\phi_0} & \frac{\partial^2\ell}{\partial\phi_1^2} & \frac{\partial^2\ell}{\partial\phi_1\partial\lambda} \\ \frac{\partial^2\ell}{\partial\lambda\partial\phi_0} & \frac{\partial^2\ell}{\partial\lambda\partial\phi_1} & \frac{\partial^2\ell}{\partial\lambda^2} \end{pmatrix},$$

respectively, where

$$\begin{aligned} \frac{\partial\ell}{\partial\phi_0} &= \frac{m_0(1-e^{-\lambda})}{\phi_0+\phi_2 e^{-\lambda}} - \frac{m_1\lambda\,e^{-\lambda}}{\phi_1+\phi_2\lambda\,e^{-\lambda}} - \frac{n-m_0-m_1}{\phi_2}, \\ \frac{\partial\ell}{\partial\phi_1} &= -\frac{m_0 e^{-\lambda}}{\phi_0+\phi_2 e^{-\lambda}} + \frac{m_1(1-\lambda\,e^{-\lambda})}{\phi_1+\phi_2\lambda\,e^{-\lambda}} - \frac{n-m_0-m_1}{\phi_2}, \\ \frac{\partial\ell}{\partial\lambda} &= -\frac{m_0\phi_2 e^{-\lambda}}{\phi_0+\phi_2 e^{-\lambda}} + \frac{m_1\phi_2(1-\lambda)e^{-\lambda}}{\phi_1+\phi_2\lambda\,e^{-\lambda}} \\ &\quad -(n-m_0-m_1) + \frac{N}{\lambda}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2\ell}{\partial\phi_0^2} &= -\frac{m_0(1-e^{-\lambda})^2}{[\phi_0+\phi_2 e^{-\lambda}]^2} - \frac{m_1\lambda^2 e^{-2\lambda}}{[\phi_1+\phi_2\lambda\,e^{-\lambda}]^2} \\ &\quad - \frac{n-m_0-m_1}{\phi_2^2}, \\ \frac{\partial^2\ell}{\partial\phi_1^2} &= -\frac{m_0 e^{-2\lambda}}{[\phi_0+\phi_2 e^{-\lambda}]^2} - \frac{m_1(1-\lambda\,e^{-\lambda})^2}{[\phi_1+\phi_2\lambda\,e^{-\lambda}]^2} \\ &\quad - \frac{n-m_0-m_1}{\phi_2^2}, \\ \frac{\partial^2\ell}{\partial\lambda^2} &= \frac{m_0\phi_0\phi_2 e^{-\lambda}}{[\phi_0+\phi_2 e^{-\lambda}]^2} \\ &\quad + \frac{m_1\phi_2 e^{-\lambda}\,[\phi_1(\lambda-2)-\phi_2 e^{-\lambda}]}{[\phi_1+\phi_2\lambda\,e^{-\lambda}]^2} - \frac{N}{\lambda^2}, \\ \frac{\partial^2\ell}{\partial\phi_0\partial\phi_1} &= \frac{m_0 e^{-\lambda}(1-e^{-\lambda})}{[\phi_0+\phi_2 e^{-\lambda}]^2} + \frac{m_1\lambda\,e^{-\lambda}(1-\lambda\,e^{-\lambda})}{[\phi_1+\phi_2\lambda\,e^{-\lambda}]^2} \\ &\quad - \frac{n-m_0-m_1}{\phi_2^2}, \\ \frac{\partial^2\ell}{\partial\phi_0\partial\lambda} &= \frac{m_0(1-\phi_1)e^{-\lambda}}{[\phi_0+\phi_2 e^{-\lambda}]^2} - \frac{m_1\phi_1(1-\lambda)e^{-\lambda}}{[\phi_1+\phi_2\lambda\,e^{-\lambda}]^2}, \\ \frac{\partial^2\ell}{\partial\phi_1\partial\lambda} &= \frac{m_0\phi_0 e^{-\lambda}}{[\phi_0+\phi_2 e^{-\lambda}]^2} - \frac{m_1(1-\phi_0)(1-\lambda)e^{-\lambda}}{[\phi_1+\phi_2\lambda\,e^{-\lambda}]^2}. \end{aligned}$$

Thus, by utilizing (4.2), we can calculate the Fisher information as follows:

$$J(\phi_0,\phi_1,\lambda) = E\Big[-\nabla^2\ell(\phi_0,\phi_1,\lambda|Y_{\rm obs})\Big]. \tag{4.4}$$

Let $(\phi_0^{(0)}, \phi_1^{(0)}, \lambda^{(0)})$ be initial values of the MLEs $(\hat\phi_0, \hat\phi_1, \hat\lambda)$. If $(\phi_0^{(t)}, \phi_1^{(t)}, \lambda^{(t)})$ denote the t-th approximations of $(\hat\phi_0, \hat\phi_1, \hat\lambda)$, then their $(t+1)$-th approximation can be obtained by the following Fisher scoring algorithm:

$$\begin{pmatrix} \phi_0^{(t+1)} \\ \phi_1^{(t+1)} \\ \lambda^{(t+1)} \end{pmatrix} = \begin{pmatrix} \phi_0^{(t)} \\ \phi_1^{(t)} \\ \lambda^{(t)} \end{pmatrix} + J^{-1}(\phi_0^{(t)}, \phi_1^{(t)}, \lambda^{(t)})\nabla\ell(\phi_0^{(t)}, \phi_1^{(t)}, \lambda^{(t)}|Y_{\rm obs}). \tag{4.5}$$

As a by-product, the standard errors of the MLEs $(\hat\phi_0, \hat\phi_1, \hat\lambda)$ are the square roots of the diagonal elements J^{kk} of the inverse Fisher information matrix $J^{-1}(\hat\phi_0, \hat\phi_1, \hat\lambda)$. Thus, the $(1-\alpha)100\%$ asymptotic Wald *confidence intervals* (CIs) of ϕ_0, ϕ_1 and λ are given by

$$\begin{aligned} &[\hat\phi_{k-1} - z_{\alpha/2}\sqrt{J^{kk}}, \quad \hat\phi_{k-1} + z_{\alpha/2}\sqrt{J^{kk}}], \quad k=1,2, \\ &[\hat\lambda - z_{\alpha/2}\sqrt{J^{33}}, \quad \hat\lambda + z_{\alpha/2}\sqrt{J^{33}}], \end{aligned} \tag{4.6}$$

respectively, where z_α denotes the α-th upper quantile of the standard normal distribution.

4.2 MLEs via the EM algorithm

The zero observations from a ZOIP distribution can be categorized into the *extra zeros* resulted from the degenerate distribution at zero because of population variability at the point zero and the *structural zeros* came from an ordinary Poisson distribution. The one observations can similarly be classified into the *extra ones* resulted from the degenerate distribution at one because of population variability at the point one and *structural ones* came from an ordinary Poisson distribution. Thus, we partition

$$\mathbb{I}_0 = \mathbb{I}_0^{\text{extra}} \cup \mathbb{I}_0^{\text{structral}} \quad \text{and} \quad \mathbb{I}_1 = \mathbb{I}_1^{\text{extra}} \cup \mathbb{I}_1^{\text{structral}}.$$

Note that a major obstacle from obtaining explicit solutions of MLEs of parameters from (4.1) is the first and second terms in the right-hand-side of (4.1). To overcome this difficulty, we augment Y_{obs} with two latent variables W_0 and W_1, where W_0 denotes the number of $\mathbb{I}_0^{\text{extra}}$ to split m_0 into W_0 and $m_0 - W_0$ and W_1 denotes the number of $\mathbb{I}_1^{\text{extra}}$ to split m_1 into W_1 and $m_1 - W_1$. Thus, the resulting conditional predictive distributions of W_0 and W_1 given $(Y_{\text{obs}}, \phi_0, \phi_1, \lambda)$ are given by

(4.7)

$$W_0|(Y_{\text{obs}}, \phi_0, \phi_1, \lambda) \sim \text{Binomial}\left(m_0, \frac{\phi_0}{\phi_0 + \phi_2 \mathrm{e}^{-\lambda}}\right) \quad \text{and}$$
$$W_1|(Y_{\text{obs}}, \phi_0, \phi_1, \lambda) \sim \text{Binomial}\left(m_1, \frac{\phi_1}{\phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}}\right),$$

respectively. On the other hand, the complete-data likelihood function is proportional to

(4.8)

$$\begin{aligned} &L(\phi_0, \phi_1, \lambda|Y_{\text{com}}) \\ &\propto \phi_0^{w_0}[(1-\phi_0-\phi_1)\mathrm{e}^{-\lambda}]^{m_0-w_0} \\ &\quad \times\ \phi_1^{w_1}[(1-\phi_0-\phi_1)\lambda\,\mathrm{e}^{-\lambda}]^{m_1-w_1} \\ &\quad \times\ (1-\phi_0-\phi_1)^{n-m_0-m_1}\mathrm{e}^{-(n-m_0-m_1)\lambda}\lambda^N \\ &= \phi_0^{w_0}\phi_1^{w_1}(1-\phi_0-\phi_1)^{n-w_0-w_1}\mathrm{e}^{-(n-w_0-w_1)\lambda}\lambda^{m_1-w_1+N}. \end{aligned}$$

Thus the M-step is to find the complete-data MLEs

$$\text{(4.9)} \qquad \begin{aligned} \hat\phi_0 &= \frac{w_0}{n}, \quad \hat\phi_1 = \frac{w_1}{n} \quad \text{and} \\ \hat\lambda &= \frac{N+m_1-w_1}{n-w_0-w_1} = \frac{N+m_1-n\hat\phi_1}{n(1-\hat\phi_0-\hat\phi_1)}. \end{aligned}$$

The E-step is to replace w_0 and w_1 in (4.9) by their conditional expectations:

(4.10)

$$E(W_0|Y_{\text{obs}}, \phi_0, \phi_1, \lambda) = \frac{m_0\phi_0}{\phi_0 + (1-\phi_0-\phi_1)\mathrm{e}^{-\lambda}} \quad \text{and}$$
$$E(W_1|Y_{\text{obs}}, \phi_0, \phi_1, \lambda) = \frac{m_1\phi_1}{\phi_1 + (1-\phi_0-\phi_1)\lambda\,\mathrm{e}^{-\lambda}}.$$

4.3 Bootstrap confidence intervals

The Wald CIs of ϕ_0 and ϕ_1 specified by (4.6) may fall outside the unit interval [0,1). The Wald CI of λ given by (4.6) is reliable only for large sample sizes and its lower bound may be less than 0. For small sample sizes, the bootstrap method is a useful tool to find a bootstrap CI for an arbitrary function of ϕ_0, ϕ_1 and λ, say, $\vartheta = h(\phi_0, \phi_1, \lambda)$. Let $\hat\vartheta = h(\hat\phi_0, \hat\phi_1, \hat\lambda)$ denote the MLE of ϑ, where $(\hat\phi_0, \hat\phi_1, \hat\lambda)$ represent the MLEs of $(\phi_0, \phi_1, \lambda)$ calculated by either the Fisher scoring algorithm (4.5) or the EM algorithm (4.9)–(4.10). Based on the obtained MLEs $(\hat\phi_0, \hat\phi_1, \hat\lambda)$, by using the SR (2.1) we can generate $Y_1^*, \ldots, Y_n^* \overset{\text{iid}}{\sim} \text{ZOIP}(\hat\phi_0, \hat\phi_1, \hat\lambda)$. Having obtained $Y_{\text{obs}}^* = \{y_1^*, \ldots, y_n^*\}$, we can calculate the bootstrap replications $(\hat\phi_0^*, \hat\phi_1^*, \hat\lambda^*)$ and get $\hat\vartheta^* = h(\hat\phi_0^*, \hat\phi_1^*, \hat\lambda^*)$. Independently repeating this process G times, we obtain G bootstrap replications $\{\hat\vartheta_g^*\}_{g=1}^G$. Consequently, the standard error, $\text{se}(\hat\vartheta)$, of $\hat\vartheta$ can be estimated by the sample standard deviation of the G replications, i.e.,

(4.11)

$$\widehat{\text{se}}(\hat\vartheta) = \left\{\frac{1}{G-1}\sum_{g=1}^G [\hat\vartheta_g^* - (\hat\vartheta_1^* + \cdots + \hat\vartheta_G^*)/G]^2\right\}^{1/2}.$$

If $\{\hat\vartheta_g^*\}_{g=1}^G$ is approximately normally distributed, the first $(1-\alpha)100\%$ bootstrap CI for ϑ is

$$\text{(4.12)} \qquad [\hat\vartheta - z_{\alpha/2}\cdot\widehat{\text{se}}(\hat\vartheta), \hat\vartheta + z_{\alpha/2}\cdot\widehat{\text{se}}(\hat\vartheta)].$$

Alternatively, if $\{\hat\vartheta_g^*\}_{g=1}^G$ is non-normally distributed, the second $(1-\alpha)100\%$ bootstrap CI of ϑ is given by

$$\text{(4.13)} \qquad [\hat\vartheta_{\text{L}},\ \hat\vartheta_{\text{U}}],$$

where $\hat\vartheta_{\text{L}}$ and $\hat\vartheta_{\text{U}}$ are the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of $\{\hat\vartheta_g^*\}_{g=1}^G$, respectively.

5. TESTING HYPOTHESES WITH LARGE SAMPLE SIZES

In this section, we consider the testing hypotheses for (i) H_0: $(\phi_0, \phi_1) = (0, 0)$; (ii) H_0: $\phi_1 = 0$; (iii) H_0: $\phi_0 = 0$; (iv) H_0: $\lambda = \lambda_0$. For (i), when H_0 is true, since the parameter values are located at the vertex boundary of the bounded parameter space, the traditional asymptotic property of LRT is not applicable. Therefore, we only consider the score test. For (ii)–(iii), although the parameter value is still on the boundary of the parameter space, we can provide appropriate reference distribution. Finally for (iv), it is a standard two-sided test.

5.1 Score test for simultaneous zero-and-one inflation

First, we develop a score test to examine whether there exist excessive zeros and excessive ones in the observations simultaneously, i.e., zero-and-one inflation in the ZOIP model. The null and alternative hypotheses are as follows

(5.1)
$$H_0\colon (\phi_0, \phi_1) = (0,0) \quad \text{against} \quad H_1\colon (\phi_0, \phi_1) \neq (0,0).$$

By reparametrization, three new parameters are introduced as follows:

(5.2)
$$\theta_0 = \frac{\phi_0}{1-\phi_0-\phi_1}, \quad \theta_1 = \frac{\phi_1}{1-\phi_0-\phi_1} \quad \text{and} \quad \beta = \log\lambda.$$

Then, testing H_0 specified in (5.1) is equivalent to testing H_0^*: $(\theta_0, \theta_1) = (0,0)$. The observed-data log-likelihood function now becomes

$$\begin{aligned}\ell_1 \;\hat{=}\; & \ell_1(\theta_0,\theta_1,\beta) = \sum_{i=1}^n \Big\{ -\log(1+\theta_0+\theta_1) \\ & + [\log(\theta_0 + \mathrm{e}^{-\lambda})] I(y_i = 0) \\ & + [\log(\theta_1 + \lambda\,\mathrm{e}^{-\lambda})] I(y_i = 1) \\ & + [y_i \log\lambda - \lambda - \log(y_i!)] I(y_i \geqslant 2)\Big\}.\end{aligned}$$

The score vector is now denoted by

$$U(\theta_0,\theta_1,\beta) = \left(\frac{\partial \ell_1}{\partial\theta_0}, \frac{\partial \ell_1}{\partial\theta_1}, \frac{\partial \ell_1}{\partial\beta}\right)^{\top},$$

where

$$\begin{aligned}\frac{\partial \ell_1}{\partial\theta_0} &= \sum_{i=1}^n \left[-\frac{1}{1+\theta_0+\theta_1} + \frac{1}{\theta_0+\mathrm{e}^{-\lambda}} I(y_i=0)\right], \\ \frac{\partial \ell_1}{\partial\theta_1} &= \sum_{i=1}^n \left[-\frac{1}{1+\theta_0+\theta_1} + \frac{1}{\theta_1+\lambda\,\mathrm{e}^{-\lambda}} I(y_i=1)\right], \\ \frac{\partial \ell_1}{\partial\beta} &= \sum_{i=1}^n \Bigg[-\frac{\lambda\,\mathrm{e}^{-\lambda}}{\theta_0+\mathrm{e}^{-\lambda}} I(y_i=0) \\ & \quad - \frac{(\lambda^2-\lambda)\mathrm{e}^{-\lambda}}{\theta_1+\lambda\,\mathrm{e}^{-\lambda}} I(y_i=1) + (y_i-\lambda) I(y_i \geqslant 2)\Bigg].\end{aligned}$$

The second derivatives are given by

$$\begin{aligned}\frac{\partial^2 \ell_1}{\partial\theta_0^2} &= \sum_{i=1}^n \left[\frac{1}{(1+\theta_0+\theta_1)^2} - \frac{1}{(\theta_0+\mathrm{e}^{-\lambda})^2} I(y_i=0)\right], \\ \frac{\partial^2 \ell_1}{\partial\theta_1^2} &= \sum_{i=1}^n \Bigg[\frac{1}{(1+\theta_0+\theta_1)^2} \\ & \quad - \frac{1}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})^2} I(y_i=1)\Bigg], \\ \frac{\partial^2 \ell_1}{\partial\beta^2} &= \sum_{i=1}^n \Bigg\{-\frac{\lambda\,\mathrm{e}^{-\lambda}[\theta_0(1-\lambda)+\mathrm{e}^{-\lambda}]}{(\theta_0+\mathrm{e}^{-\lambda})^2} I(y_i=0) \\ & \quad - \frac{[\theta_1(3\lambda^2-\lambda-\lambda^3)\mathrm{e}^{-\lambda}+\lambda^3\mathrm{e}^{-2\lambda}]}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})^2} I(y_i=1) \\ & \quad -\lambda I(y_i \geqslant 2)\Bigg\}, \\ \frac{\partial^2 \ell_1}{\partial\theta_0\partial\theta_1} &= \sum_{i=1}^n \frac{1}{(1+\theta_0+\theta_1)^2}, \\ \frac{\partial^2 \ell_1}{\partial\theta_0\partial\beta} &= \sum_{i=1}^n \left[\frac{\lambda\,\mathrm{e}^{-\lambda}}{(\theta_0+\mathrm{e}^{-\lambda})^2} I(y_i=0)\right], \\ \frac{\partial^2 \ell_1}{\partial\theta_1\partial\beta} &= \sum_{i=1}^n \left[\frac{(\lambda^2-\lambda)\mathrm{e}^{-\lambda}}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})^2} I(y_i=1)\right].\end{aligned}$$

Since

$$\begin{aligned}E[I(y_i=0)] &= \frac{\theta_0+\mathrm{e}^{-\lambda}}{1+\theta_0+\theta_1}, \\ E[I(y_i=1)] &= \frac{\theta_1+\lambda\,\mathrm{e}^{-\lambda}}{1+\theta_0+\theta_1} \quad \text{and} \\ E[I(y_i \geqslant 2)] &= \frac{1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}}{1+\theta_0+\theta_1},\end{aligned}$$

the Fisher information matrix can be calculated as $J(\theta_0,\theta_1,\beta) = (J_{jj'})$, where

$$\begin{aligned}J_{11} &= -E\left(\frac{\partial^2\ell_1}{\partial\theta_0^2}\right) = \frac{n(1+\theta_1-\mathrm{e}^{-\lambda})}{(1+\theta_0+\theta_1)^2(\theta_0+\mathrm{e}^{-\lambda})}, \\ J_{22} &= -E\left(\frac{\partial^2\ell_1}{\partial\theta_1^2}\right) = \frac{n(1+\theta_0-\lambda\,\mathrm{e}^{-\lambda})}{(1+\theta_0+\theta_1)^2(\theta_1+\lambda\,\mathrm{e}^{-\lambda})}, \\ J_{33} &= -E\left(\frac{\partial^2\ell_1}{\partial\beta^2}\right) = \frac{n\lambda\,\mathrm{e}^{-\lambda}[\theta_0(1-\lambda)+\mathrm{e}^{-\lambda}]}{(\theta_0+\mathrm{e}^{-\lambda})(1+\theta_0+\theta_1)} \\ & \quad + \frac{n[\theta_1(3\lambda^2-\lambda-\lambda^3)\mathrm{e}^{-\lambda}+\lambda^3\mathrm{e}^{-2\lambda}]}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})(1+\theta_0+\theta_1)} \\ & \quad + \frac{n\lambda(1-\mathrm{e}^{\lambda}-\lambda\,\mathrm{e}^{\lambda})}{1+\theta_0+\theta_1}, \\ J_{12} &= -E\left(\frac{\partial^2\ell_1}{\partial\theta_0\partial\theta_1}\right) = -\frac{n}{(1+\theta_0+\theta_1)^2}, \\ J_{13} &= -E\left(\frac{\partial^2\ell_1}{\partial\theta_0\partial\beta}\right) = -\frac{n\lambda\,\mathrm{e}^{-\lambda}}{(1+\theta_0+\theta_1)(\theta_0+\mathrm{e}^{-\lambda})}, \\ J_{23} &= -E\left(\frac{\partial^2\ell_1}{\partial\theta_1\partial\beta}\right) = -\frac{n(\lambda^2-\lambda)\mathrm{e}^{-\lambda}}{(1+\theta_0+\theta_1)(\theta_1+\lambda\,\mathrm{e}^{-\lambda})}.\end{aligned}$$

Under H_0^*, the score test statistic

(5.3)
$$T_1 = U^{\top}(0,0,\hat\beta)\, J^{-1}(0,0,\hat\beta)\, U(0,0,\hat\beta) \,\dot\sim\, \chi^2(2),$$

where $\hat\beta = \log(\bar y)$ and $U(0,0,\hat\beta) = (m_0\mathrm{e}^{\bar y} - n,\ m_1\mathrm{e}^{\bar y}/\bar y - n,\ 0)^{\top}$. The p-value is

(5.4)
$$p_{\mathrm{v}1} = \Pr(T_1 > t_1 | H_0^*) = \Pr(\chi^2(2) > t_1),$$

where t_1 is the observed value of T_1.

5.2 Likelihood ratio test for one inflation

If the null hypothesis H_0 specified by (5.1) is rejected at the α level of significance (i.e, at least one of the ϕ_0 and ϕ_1 are positive), we could next test whether there exist extra ones in the observations, i.e., one-inflation in the ZOIP model. We consider the following null and alternative hypotheses

$$H_0\colon \phi_1 = 0 \quad \text{against} \quad H_1\colon \phi_1 > 0. \tag{5.5}$$

Under H_0, the *likelihood ratio* (LR) test statistic is given by

$$T_2 = -2\{\ell(\hat\phi_{0,H_0}, 0, \hat\lambda_{H_0}|Y_{\text{obs}}) - \ell(\hat\phi_0, \hat\phi_1, \hat\lambda|Y_{\text{obs}})\}, \tag{5.6}$$

where $(\hat\phi_{0,H_0}, \hat\lambda_{H_0})$ denote the constrained MLEs of (ϕ_0, λ) under H_0 and $(\hat\phi_0, \hat\phi_1, \hat\lambda)$ denote the unconstrained MLEs of $(\phi_0, \phi_1, \lambda)$, which are obtained by either the Fisher scoring algorithm (4.5) or the EM algorithm (4.9)–(4.10). Note that under H_0, the ZOIP$(\phi_0, \phi_1; \lambda)$ distribution reduces to the zero-inflated Poisson distribution ZIP(ϕ_0, λ). Thus, the MLEs of (ϕ_0, λ) can be calculated by the following EM iteration:

$$\begin{aligned} \phi_{0,H_0}^{(t+1)} &= \frac{m_0 \phi_{0,H_0}^{(t)}}{n\Big[\phi_{0,H_0}^{(t)} + \Big(1 - \phi_{0,H_0}^{(t)}\Big)\mathrm{e}^{-\lambda_{H_0}^{(t+1)}}\Big]}, \\ \lambda_{H_0}^{(t+1)} &= \frac{\bar y}{1 - \phi_{0,H_0}^{(t)}}, \end{aligned} \tag{5.7}$$

where $\bar y = (1/n)\sum_{i=1}^n y_i$.

Standard large-sample theory suggests that the asymptotic null distribution of T_2 is $\chi^2(1)$. However, the null hypothesis in (5.5) corresponds to ϕ_1 being on the boundary of the parameter space and the appropriate null distribution is a 50:50 mixture of $\chi^2(0)$ (i.e., Degenerate(0)) and $\chi^2(1)$, see Self and Liang [28] and Feng and McCulloch [9]. Hence, the corresponding p-value (Jansakul and Hinde, [14], p. 78; Joe and Zhu, [15], p. 225) is

$$p_{\text{v}2} = \Pr(T_2 > t_2|H_0) = \frac{1}{2}\Pr(\chi^2(1) > t_2). \tag{5.8}$$

5.3 Score test for one inflation

Alternatively, the score test can be used for testing H_0 specified in (5.5), which is equivalent to testing $H_0^*\colon \theta_1 = 0$. Let $(\theta_0, \theta_1, \beta)$ be defined by (5.2). Under H_0^*, the score test statistic

$$T_3 = U^\top(\hat\theta_0, 0, \hat\beta)\, J^{-1}(\hat\theta_0, 0, \hat\beta)\, U(\hat\theta_0, 0, \hat\beta) \,\dot\sim\, \chi^2(1), \tag{5.9}$$

where $\hat\theta_0 = \hat\phi_{0,H_0}/(1 - \hat\phi_{0,H_0})$ and $\hat\beta = \log(\hat\lambda_{H_0})$ denote the MLEs of θ_0 and β under H_0^*, and $(\hat\phi_{0,H_0}, \hat\lambda_{H_0})$ are determined by (5.7). Note that the score vector $U(\theta_0, \theta_1, \beta)$ evaluated at $(\theta_0, \theta_1, \beta) = (\hat\theta_0, 0, \hat\beta)$ is given by

$$U(\hat\theta_0, 0, \hat\beta) = \left(0,\ -\frac{n}{1+\hat\theta_0} + \frac{m_1}{\hat\lambda_{H_0}\mathrm{e}^{-\hat\lambda_{H_0}}},\ 0\right)^\top,$$

where $m_1 = \sum_{i=1}^n I(y_i = 1)$. The corresponding p-value is

$$p_{\text{v}3} = \Pr(T_3 > t_3|H_0^*) = \Pr(\chi^2(1) > t_3). \tag{5.10}$$

5.4 LR test for zero inflation

To test whether there exist extra zeros in the observations, i.e., zero-inflation in the ZOIP model, we consider the following null and alternative hypotheses

$$H_0\colon \phi_0 = 0 \quad \text{against} \quad H_1\colon \phi_0 > 0. \tag{5.11}$$

Under H_0, the LR test statistic (Jansakul and Hinde, [14], p. 78; Joe and Zhu, [15], p. 225)

$$\begin{aligned} T_4 &= -2\{\ell(0, \hat\phi_{1,H_0}, \hat\lambda_{H_0}|Y_{\text{obs}}) - \ell(\hat\phi_0, \hat\phi_1, \hat\lambda|Y_{\text{obs}})\} \\ &\dot\sim\, 0.5\chi^2(0) + 0.5\chi^2(1), \end{aligned} \tag{5.12}$$

where $(\hat\phi_{1,H_0}, \hat\lambda_{H_0})$ denote the constrained MLEs of (ϕ_1, λ) under H_0 and $(\hat\phi_0, \hat\phi_1, \hat\lambda)$ denote the unconstrained MLEs of $(\phi_0, \phi_1, \lambda)$, which are obtained by either the Fisher scoring algorithm (4.5) or the EM algorithm (4.9)–(4.10). Note that under H_0, the ZOIP$(\phi_0, \phi_1; \lambda)$ distribution reduces to the one-inflated Poisson distribution OIP(ϕ_1, λ). Thus, the MLEs of (ϕ_1, λ) can be calculated by the following EM iteration:

$$\begin{aligned} \phi_{1,H_0}^{(t+1)} &= \frac{m_1 \phi_{1,H_0}^{(t)}}{n\Big[\phi_{1,H_0}^{(t)} + \Big(1 - \phi_{1,H_0}^{(t)}\Big)\lambda_{H_0}^{(t+1)}\mathrm{e}^{-\lambda_{H_0}^{(t+1)}}\Big]}, \\ \lambda_{H_0}^{(t+1)} &= \frac{\bar y - \phi_{1,H_0}^{(t)}}{1 - \phi_{1,H_0}^{(t)}}, \end{aligned} \tag{5.13}$$

where $\bar y = (1/n)\sum_{i=1}^n y_i$. Hence the corresponding p-value is

$$p_{\text{v}4} = \Pr(T_4 > t_4|H_0) = \frac{1}{2}\Pr(\chi^2(1) > t_4). \tag{5.14}$$

5.5 Score test for zero inflation

Let $(\theta_0, \theta_1, \beta)$ be defined in (5.2), testing H_0 specified by (5.11) is equivalent to testing $H_0^*\colon \theta_0 = 0$. Under H_0^*, the score test statistic

$$T_5 = U^\top(0, \hat\theta_1, \hat\beta)\, J^{-1}(0, \hat\theta_1, \hat\beta)\, U(0, \hat\theta_1, \hat\beta) \,\dot\sim\, \chi^2(1), \tag{5.15}$$

where $\hat\theta_1 = \hat\phi_{1,H_0}/(1 - \hat\phi_{1,H_0})$ and $\hat\beta = \log(\hat\lambda_{H_0})$ denote the MLEs of θ_1 and β under H_0^*, and $(\hat\phi_{1,H_0}, \hat\lambda_{H_0})$ are determined by (5.13). Note that the score vector $U(\theta_0, \theta_1, \beta)$ evaluated at $(\theta_0, \theta_1, \beta) = (0, \hat\theta_1, \hat\beta)$ is given by

$$U(0, \hat\theta_1, \hat\beta) = \left(-\frac{n}{1+\hat\theta_1} + m_0\mathrm{e}^{\hat\lambda_{H_0}},\ 0,\ 0\right)^\top,$$

where $m_0 = \sum_{i=1}^n I(y_i = 0)$. The corresponding p-value is

$$p_{\text{v}5} = \Pr(T_5 > t_5|H_0^*) = \Pr(\chi^2(1) > t_5). \tag{5.16}$$

5.6 LR test for testing $\lambda = \lambda_0$

Suppose that we want to test the null hypothesis

$$H_0\colon \lambda = \lambda_0 \quad \text{against} \quad H_1\colon \lambda \neq \lambda_0. \tag{5.17}$$

Under H_0, the LR test statistic

$$T_6 = -2\{\ell(\hat{\phi}_{0,H_0}, \hat{\phi}_{1,H_0}, \lambda_0|Y_{\rm obs}) - \ell(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda}|Y_{\rm obs})\} \dot{\sim} \chi^2(1), \tag{5.18}$$

where $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ are the unconstrained MLEs of $(\phi_0, \phi_1, \lambda)$ and $(\hat{\phi}_{0,H_0}, \hat{\phi}_{1,H_0})$ are the restricted MLEs of (ϕ_0, ϕ_1) under H_0, which can be obtained by the following EM iteration:
(5.19)

$$\phi_{0,H_0}^{(t+1)} = \frac{m_0\phi_{0,H_0}^{(t)}}{n\left[\phi_{0,H_0}^{(t)} + (1-\phi_{0,H_0}^{(t)} - \phi_{1,H_0}^{(t)})\mathrm{e}^{-\lambda_0}\right]},$$

$$\phi_{1,H_0}^{(t+1)} = \frac{m_1\phi_{1,H_0}^{(t)}}{n\left[\phi_{1,H_0}^{(t)} + (1-\phi_{0,H_0}^{(t)} - \phi_{1,H_0}^{(t)})\lambda_0\mathrm{e}^{-\lambda_0}\right]}.$$

The corresponding p-value is

$$p_{\rm v6} = \Pr(T_6 > t_6|H_0) = \Pr(\chi^2(1) > t_6). \tag{5.20}$$

5.7 Score test for testing $\lambda = \lambda_0$

Let $(\theta_0, \theta_1, \beta)$ be defined in (5.2), then, testing H_0 specified by (5.17) is equivalent to testing $H_0^*\colon \beta = \beta_0 = \log\lambda_0$. Under H_0^*, the score test statistic
(5.21)

$$T_7 = U^{\top}(\hat{\theta}_0, \hat{\theta}_1, \beta_0)\, J^{-1}(\hat{\theta}_0, \hat{\theta}_1, \beta_0)\, U(\hat{\theta}_0, \hat{\theta}_1, \beta_0) \dot{\sim} \chi^2(1),$$

where $(\hat{\theta}_0, \hat{\theta}_1)$ are the MLEs of (θ_0, θ_1) under H_0^*. Hence the p-value is

$$p_{\rm v7} = \Pr(T_7 > t_7|H_0^*) = \Pr(\chi^2(1) > t_7). \tag{5.22}$$

6. SIMULATION STUDIES

To investigate the performance of the *likelihood ratio test* (LRT) and the score test, we compare the type I error rate and the power of the two tests for (1) $\phi_1 = 0$, (2) $\phi_0 = 0$ and (3) $\lambda = \lambda_0$. And the sample sizes are set to be $n = 50(50)500$.

6.1 Tests for one inflation

In this subsection, we compare the type I error rates (with $H_0\colon \phi_1 = 0$) and powers (with $H_1\colon \phi_1 > 0$) between the LRT and the score test for various sample sizes via simulations, where the values of ϕ_1 in H_1 are chosen to be $0.01, 0.03, 0.05, 0.07, 0.10, 0.15$. For a given combination of $(n, \phi_0 = 0.5, \phi_1, \lambda = 2)$, we first independently draw $\mathbf{z}_1^{(l)}, \ldots, \mathbf{z}_n^{(l)} \overset{\rm iid}{\sim} \text{Multinomial}\,(1; \phi_0, \phi_1, \phi_2)$ for $l = 1, \ldots, L$, where $\mathbf{z}_i^{(l)} = (Z_{0i}^{(l)}, Z_{1i}^{(l)}, Z_{2i}^{(l)})^{\top}$, $i = 1, \ldots, n$. And then we independently generate $X_1^{(l)}, \ldots, X_n^{(l)} \overset{\rm iid}{\sim} \text{Poisson}(\lambda)$. Finally, we set

$$Y_i^{(l)} = Z_{1i}^{(l)} + Z_{2i}^{(l)} \cdot X_i^{(l)}, \quad i = 1, \ldots, n; \quad l = 1, \ldots, L\,(L = 1{,}000). \tag{6.1}$$

All hypothesis testings are conducted at significant level of $\alpha = 0.05$. Let r_k denote the number of rejecting the null hypothesis $H_0\colon \phi_1 = 0$ by the test statistics T_k $(k = 2, 3)$ given by (5.6) and (5.9), respectively. Hence, the actual significance level can be estimated by r_k/L with $\phi_1 = 0$ and the power of the test statistic T_k can be estimated by r_k/L with $\phi_1 > 0$. We repeat the process of estimating each significance level and each power for 1,000 times, respectively, to obtain $\{\hat{\alpha}_k\}_{k=1}^{1000}$ and $\{1-\hat{\beta}_k\}_{k=1}^{1000}$. Then we use the mean, $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of these values to estimate the empirical level/power and the corresponding $(1-\alpha)100\%$ upper and lower bounds.

Figure 1 shows the comparison of type I error rates between the LRT and the score test for various sample sizes, and the 95% CIs for the empirical significance level associated with the two tests. We can see that the score test has the correct size around $\alpha = 0.05$, while the LRT can control its type I error rates in a lower level (around 0.045). As far as we know, the lower the type I error rate the better the test's performance, thus the LRT has a better performance in controlling its type I error rates around the pre-chosen nominal level than the score test. An interpretation is as follows: For testing $H_0\colon \phi_1 = 0$ and $H_0\colon \phi_0 = 0$ (corresponding to Figure 3), since the LRT considers both the null and alternative situations, it is more sensitive in distinguishing two models, which results in a lower type I error rate.

Figure 2 gives the comparison of powers between the LRT and the score test for different values of $\phi_1 > 0$. It is not difficult to find that the LRT is always more powerful than the score test. A possible interpretation is as follows: For testing $H_0\colon \phi_1 = 0$ and $H_0\colon \phi_0 = 0$ (corresponding to Figure 4) which are one-sided tests, the LRT should be more powerful than the score test in the sense that the LRT evaluates things under both the null and alternative while the score test only evaluates things under the null, thus it is blind to the sign of the ϕ. Then the score test is more likely to make type II error rates and is less powerful.

The empirical levels/powers of the LRT statistic T_2 and the score test statistic T_3 are summarized in Tables 1 and 2, respectively, for six scenarios: $\phi_1 = 0.01, 0.03, 0.05, 0.07, 0.10, 0.15$.

6.2 Tests for zero inflation

In this subsection, we compare the type I error rates (with $H_0\colon \phi_0 = 0$) and powers (with $H_1\colon \phi_0 > 0$) between the LRT and the score test for various sample sizes via simulations, where the values of ϕ_0 in H_1 are set to be $0.01, 0.03, 0.05, 0.07, 0.10, 0.15$. For a given combination

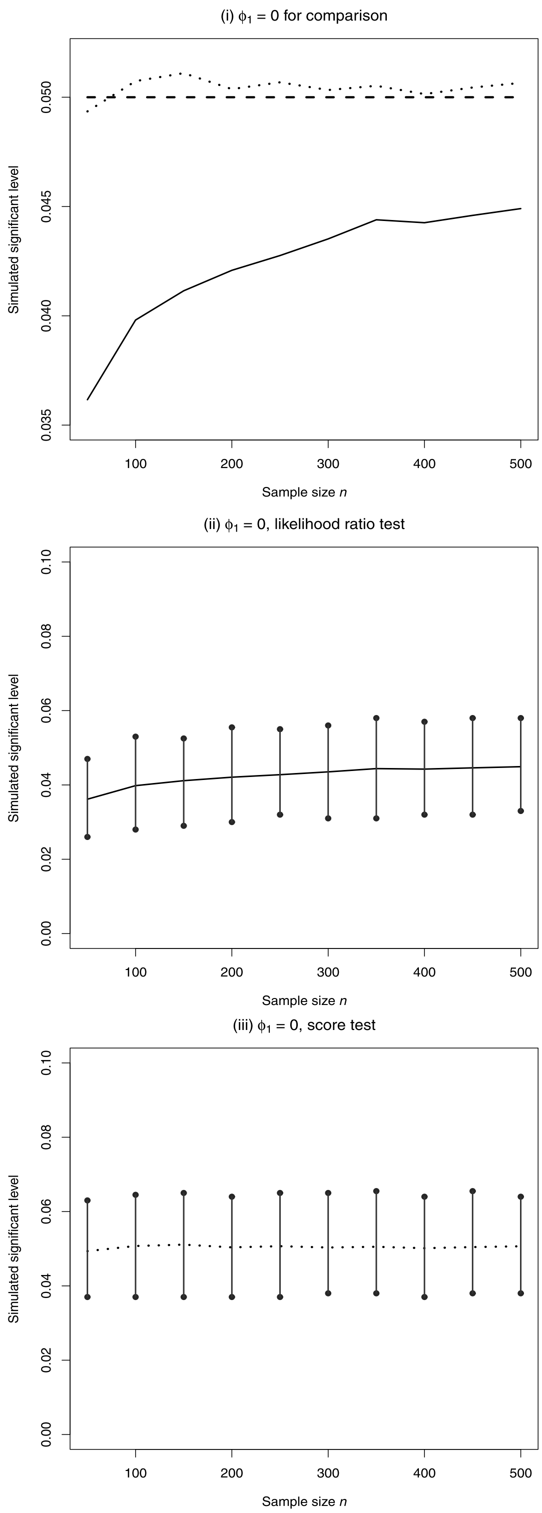

Figure 1. (i) Comparison of type I error rates between the LRT (solid line) and the score test (dotted line) for testing one inflation in the ZOIP model with H_0: $\phi_1 = 0$ against H_1: $\phi_1 > 0$. The dashed line is set as the predetermined significance level of $\alpha = 0.05$; (ii) 95% CIs for the empirical level of significance of the LRT; (iii) 95% CIs for the empirical level of significance of the score test.

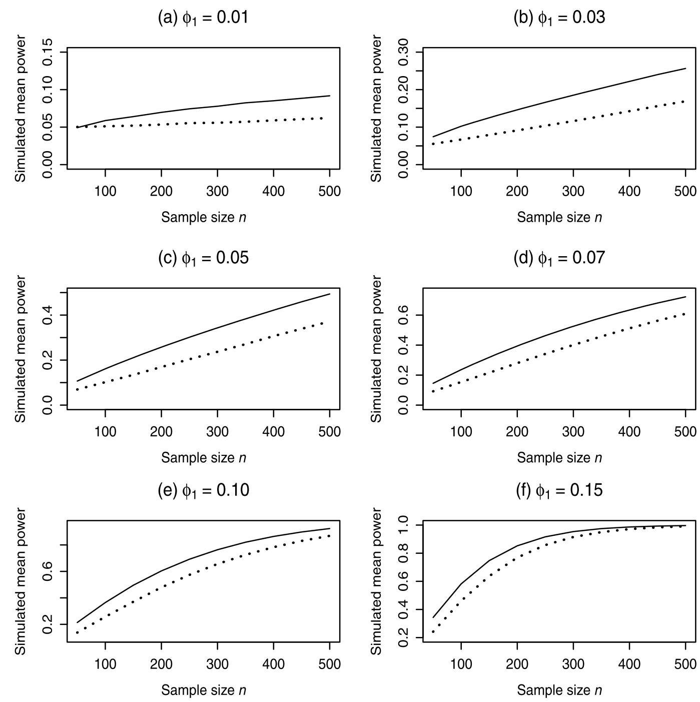

Figure 2. Comparison of powers between the LRT (solid line) and the score test (dotted line) for testing one inflation in the ZOIP model with H_0: $\phi_1 = 0$ against H_1: $\phi_1 > 0$ for different values of ϕ_1.

Table 1. Empirical levels/powers of the LRT statistic T_2 based on 1,000 replications for $\phi_0 = 0.5$ and $\lambda = 2$

Sample size (n)	Empirical level	Empirical power ϕ_1					
		0.010	0.030	0.050	0.070	0.100	0.150
50	0.036	0.049	0.074	0.107	0.145	0.214	0.343
100	0.040	0.059	0.102	0.161	0.236	0.365	0.582
150	0.041	0.064	0.125	0.211	0.319	0.496	0.747
200	0.042	0.070	0.147	0.257	0.394	0.604	0.852
250	0.043	0.074	0.166	0.302	0.463	0.692	0.917
300	0.044	0.078	0.184	0.343	0.525	0.764	0.954
350	0.044	0.082	0.204	0.384	0.582	0.821	0.975
400	0.044	0.085	0.222	0.422	0.634	0.864	0.986
450	0.045	0.088	0.240	0.459	0.680	0.898	0.993
500	0.045	0.092	0.256	0.494	0.722	0.924	0.996

of $(n, \phi_0, \phi_1 = 0.5, \lambda = 2)$, we generate $Y_1^{(l)}, \ldots, Y_n^{(l)} \stackrel{\text{iid}}{\sim}$ ZOIP$(\phi_0, 0.5; 2)$ with $L = 1{,}000$ replications. All hypothesis testings are conducted at significant level of $\alpha = 0.05$.

Let r_k denote the number of rejecting the null hypothesis H_0: $\phi_0 = 0$ by the test statistics T_k $(k = 4, 5)$ given by (5.12) and (5.15), respectively. Hence, the actual significance level can be estimated by r_k/L with $\phi_0 = 0$ and the power of the test statistic T_k can be estimated by r_k/L with $\phi_0 > 0$. The estimated level/power and 95% CIs are calculated in a similar way as in Section 6.1.

Table 2. Empirical levels/powers of the score statistic T_3 based on 1,000 replications for $\phi_0 = 0.5$ and $\lambda = 2$

Sample size (n)	Empirical level	Empirical power					
		ϕ_1					
		0.010	0.030	0.050	0.070	0.100	0.150
50	0.050	0.050	0.055	0.070	0.092	0.138	0.242
100	0.051	0.051	0.067	0.102	0.154	0.257	0.461
150	0.051	0.052	0.079	0.135	0.216	0.371	0.637
200	0.050	0.053	0.091	0.169	0.280	0.479	0.769
250	0.051	0.055	0.104	0.204	0.341	0.574	0.858
300	0.050	0.056	0.116	0.237	0.401	0.656	0.915
350	0.051	0.057	0.129	0.271	0.458	0.725	0.950
400	0.050	0.059	0.143	0.307	0.512	0.783	0.972
450	0.050	0.061	0.156	0.339	0.563	0.830	0.984
500	0.051	0.062	0.169	0.372	0.609	0.869	0.991

Figure 3 shows the comparison of type I error rates between the LRT and the score test, and the 95% CIs for the empirical level. Both the LRT and the score test have the correct size around $\alpha = 0.05$. However, the LRT test performs relatively better in controlling its type I error rates around the pre-chosen nominal level than the score test.

Figure 4 gives the comparison of powers between the LRT and the score test for different values of ϕ_0. We found that the score test is always less powerful than the LRT.

The empirical levels/powers of the LRT statistic T_4 and the score test statistic T_5 are summarized in Tables 3 and 4, respectively, for six scenarios: $\phi_0 = 0.01, 0.03, 0.05, 0.07, 0.10, 0.15$.

6.3 Tests for $\lambda = \lambda_0$

In this subsection, we compare the type I error rates (with H_0: $\lambda = \lambda_0$) and powers (with H_1: $\lambda \neq \lambda_0$) between the LRT and the score test for various sample sizes, different values of λ_0 (set to be 2, 3, 5, 7 for the comparison of empirical levels) and $\lambda = 1.5$ for the comparison of empirical powers. For a given combination of $(n, \phi_0 = 0.3, \phi_1 = 0.1, \lambda)$, we generate $Y_1^{(l)}, \ldots, Y_n^{(l)} \stackrel{\text{iid}}{\sim} \text{ZOIP}(\phi_0, \phi_1; \lambda)$ with $L = 1{,}000$ replications.

Let r_k denote the number of rejecting the null hypothesis H_0: $\lambda = \lambda_0$ by the test statistics T_k $(k = 6, 7)$ given by (5.18) and (5.21), respectively. Hence, the actual significance level can be estimated by r_k/L with $\lambda = \lambda_0$ and the power of the test statistic T_k can be estimated by r_k/L with $\lambda \neq \lambda_0$. All hypothesis testings are conducted at significant level of $\alpha = 0.05$. The estimated significance level/power and 95% CIs are calculated in a similar way as in Section 6.1.

Figure 5 shows that some comparison of type I error rates between the LRT and the score test. When $\lambda_0 = 2$ and the sample size is smaller than 200, the score test has a better performance in controlling its type I error rates around the pre-chosen nominal level than the LRT, while they are interlaced as the sample size becomes larger. When $\lambda_0 \geqslant 3$, it

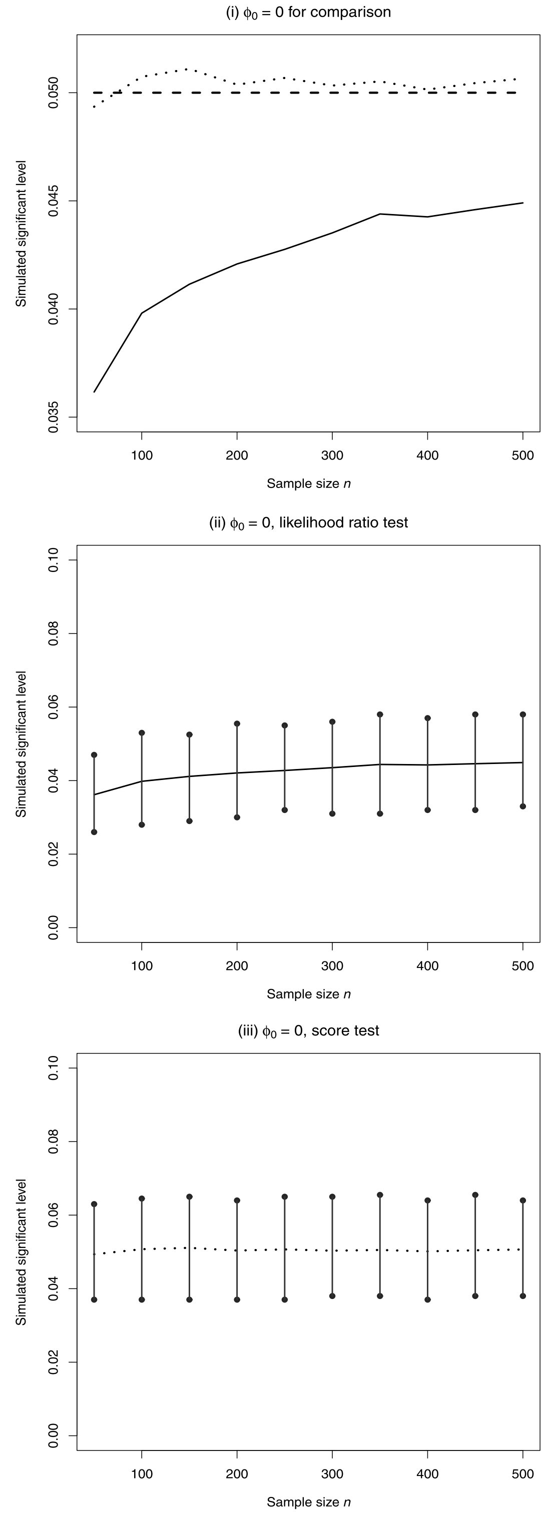

Figure 3. (i) Comparison of type I error rates between the LRT (solid line) and the score test (dotted line) for testing zero inflation in the ZOIP model with H_0: $\phi_0 = 0$ against H_1: $\phi_0 > 0$. The dashed line is set as the predetermined significance level of $\alpha = 0.05$; (ii) 95% CIs for the empirical level of significance of the LRT; (iii) 95% CIs for the empirical level of significance of the score test.

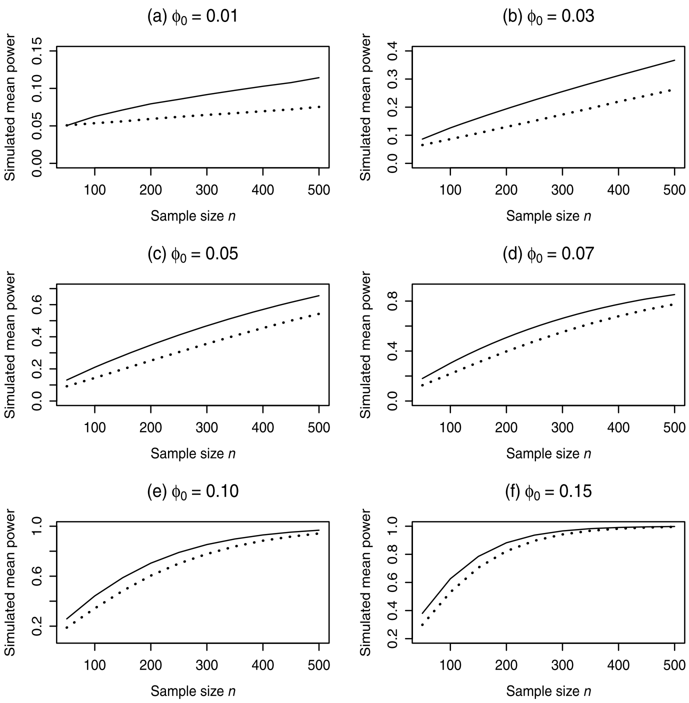

Figure 4. Comparison of powers between the LRT (solid line) and the score test (dotted line) for testing zero inflation in the ZOIP model with H_0: $\phi_0 = 0$ against H_1: $\phi_0 > 0$ for different values of ϕ_0.

Table 3. Empirical levels/powers of the LRT statistic T_4 based on 1,000 replications for $\phi_1 = 0.5$ and $\lambda = 2$

Sample size (n)	Empirical level	Empirical power					
		ϕ_0					
		0.010	0.030	0.050	0.070	0.100	0.150
50	0.036	0.050	0.086	0.131	0.180	0.258	0.380
100	0.039	0.063	0.126	0.210	0.303	0.442	0.626
150	0.041	0.071	0.161	0.281	0.412	0.588	0.785
200	0.043	0.079	0.194	0.349	0.507	0.704	0.882
250	0.043	0.085	0.224	0.410	0.591	0.790	0.937
300	0.044	0.092	0.256	0.468	0.662	0.853	0.966
350	0.044	0.097	0.284	0.521	0.723	0.898	0.983
400	0.045	0.103	0.313	0.570	0.774	0.930	0.991
450	0.044	0.108	0.340	0.615	0.818	0.953	0.996
500	0.045	0.114	0.367	0.656	0.852	0.968	0.998

can be seen that the two lines fluctuate with several points and sections, so we cannot say which one is more powerful. Thus when comparing the type I error rates between the two tests, there is no absolutely better one.

Figure 6 gives the comparison of powers between the LRT and the score test for the case with $\lambda = 1.5 \neq \lambda_0 = 1$. It is not difficult to find that the LRT is slightly less powerful than the score test when sample size is less than 200 and the two tests almost have the same powers when the sample size is larger than 200.

Table 4. Empirical levels/powers of the score statistic T_5 based on 1,000 replications for $\phi_1 = 0.5$ and $\lambda = 2$

Sample size (n)	Empirical level	Empirical power					
		ϕ_0					
		0.010	0.030	0.050	0.070	0.100	0.150
50	0.049	0.051	0.065	0.092	0.126	0.188	0.298
100	0.051	0.054	0.086	0.144	0.218	0.343	0.531
150	0.050	0.056	0.107	0.198	0.310	0.483	0.706
200	0.050	0.059	0.130	0.252	0.397	0.605	0.823
250	0.050	0.062	0.151	0.305	0.478	0.702	0.897
300	0.050	0.065	0.174	0.356	0.553	0.778	0.942
350	0.051	0.067	0.196	0.407	0.620	0.838	0.968
400	0.050	0.070	0.220	0.454	0.679	0.884	0.983
450	0.050	0.072	0.241	0.501	0.731	0.917	0.991
500	0.050	0.075	0.263	0.543	0.775	0.942	0.995

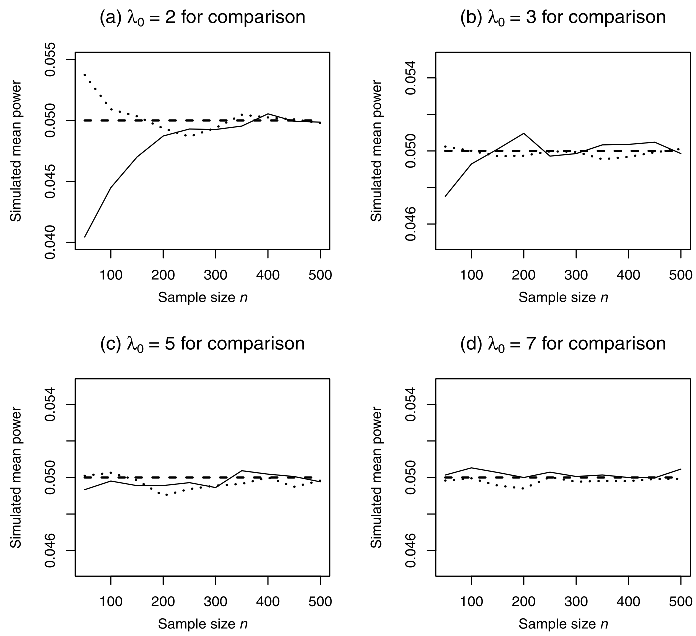

Figure 5. Comparison of type I error rates between the LRT (solid line) and the score test (dotted line) for testing H_0: $\lambda = \lambda_0$ against H_1: $\lambda \neq \lambda_0$ with different values of λ_0. The dashed line is set as the predetermined significance level of $\alpha = 0.05$.

The empirical levels/powers of the LRT statistic T_6 and the score test statistic T_7 are summarized in Tables 5 and 6, respectively.

7. APPLICATIONS

In this section, five real data sets are used to illustrate the proposed methods, where the Newton–Raphson algorithm for finding the MLEs of parameters is not available in all examples because the corresponding observed information matrices are nearly singular, while the Fisher scoring and EM algorithms work well in all examples.

Table 5. *Empirical levels/powers of the LRT statistic T_6 based on 1,000 replications for $\phi_0 = 0.3$ and $\phi_1 = 0.1$*

Sample size	Empirical level				Empirical power
(n)	$\lambda_0 = 2$	$\lambda_0 = 3$	$\lambda_0 = 5$	$\lambda_0 = 7$	$\lambda = 1.5\,(H_0\colon \lambda = \lambda_0 = 1)$
50	0.040	0.048	0.049	0.050	0.877
100	0.044	0.049	0.050	0.051	0.987
150	0.047	0.050	0.050	0.050	0.999
200	0.049	0.051	0.050	0.050	1
250	0.049	0.050	0.050	0.050	1
300	0.049	0.050	0.049	0.050	1
350	0.050	0.050	0.050	0.050	1
400	0.051	0.050	0.050	0.050	1
450	0.050	0.050	0.050	0.050	1
500	0.050	0.050	0.050	0.050	1

Table 6. *Empirical levels/powers of the score statistic T_7 based on 1,000 replications for $\phi_0 = 0.3$ and $\phi_1 = 0.1$*

Sample size	Empirical level				Empirical power
(n)	$\lambda_0 = 2$	$\lambda_0 = 3$	$\lambda_0 = 5$	$\lambda_0 = 7$	$\lambda = 1.5\,(H_0\colon \lambda = \lambda_0 = 1)$
50	0.054	0.050	0.050	0.050	0.968
100	0.051	0.050	0.050	0.050	0.998
150	0.050	0.050	0.050	0.050	1
200	0.049	0.050	0.049	0.049	1
250	0.049	0.050	0.049	0.050	1
300	0.049	0.050	0.050	0.050	1
350	0.050	0.050	0.050	0.050	1
400	0.050	0.050	0.050	0.050	1
450	0.050	0.050	0.049	0.050	1
500	0.050	0.050	0.050	0.050	1

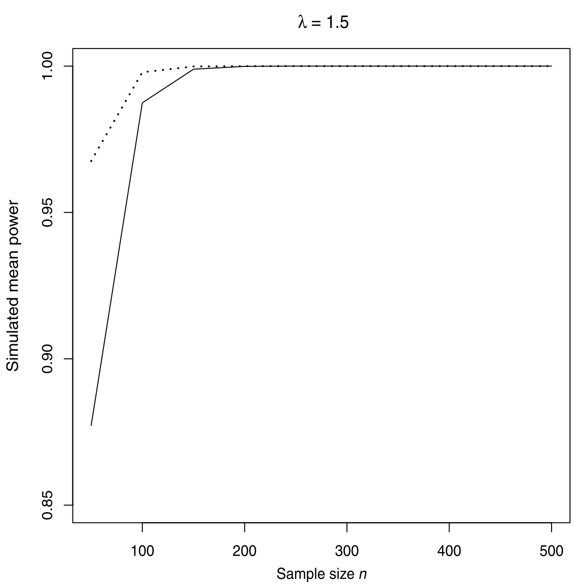

Figure 6. Comparison of powers between the LRT (solid line) and the score test (dotted line) for testing H_0: $\lambda = \lambda_0 = 1$ against H_1: $\lambda = 1.5 \neq \lambda_0$.

7.1 Dentist visiting data in Sweden

Eriksson and Åberg [8] reported a two-year panel data from Swedish Level of Living Surveys in 1974 and 1991. To investigate the long term effect of the regular dentist visits during childhood and adolescence, the questions were set to know about people's habit of visiting dentist. The questions in the surveys are retrospective and in 1974 (1991) the questions refer to the individuals' situation in 1973 (1990). The panel includes 766 individuals who were between 15 and 29 years old in 1973. The number of visits to a dentist during the previous twelve month is only available for 1991, and whether the individual visits a dentist regularly is only available for 1974. Since we focus on modeling the distribution of visits to a dentist, we only gives the sample with different visit frequencies to a dentist in 1990 in Table 7.

7.1.1 Likelihood-based inferences

We noted that the data are characterized by both large proportions of zero visits and one visits to a dentist which are 17.5 and 41 percent, respectively. Therefore, the ZOIP distribution can be considered to capture the data. Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{ZOIP}(\phi_0, \phi_1; \lambda)$. To find the MLEs of parameters, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \lambda^{(0)}) = (0.3, 0.3, 3)$ as their initial values. The MLEs of $(\phi_0, \phi_1, \lambda)$ converged to $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ as shown in the second column of Table 8 in 3 iterations for

Table 7. The dentist visiting data from Swedish Level of Living Surveys (Eriksson and Åberg, 1987; Melkersson and Olsson, 1999)

Count	0	1	2	3	4	5	6	7	8	9	10	12	15	20
Frequency	134	314	149	69	32	26	14	6	1	0	11	3	3	4

Table 8. MLEs and CIs of parameters for the dentist visiting data in Sweden

Parameter	MLE	std$^{\mathrm{F}}$	95% asymptotic Wald CI	std$^{\mathrm{B}}$	95% bootstrap CI†	95% bootstrap CI‡
ϕ_0	0.1535	0.0144	[0.1253, 0.1817]	0.0146	[0.1250, 0.1821]	[0.1255, 0.1828]
ϕ_1	0.3422	0.0210	[0.3010, 0.3834]	0.0212	[0.3000, 0.3831]	[0.3009, 0.3828]
λ	3.1580	0.1169	[2.9289, 3.3870]	0.1176	[2.9272, 3.3881]	[2.9340, 3.3854]

std$^{\mathrm{F}}$: Square roots of the diagonal elements of inverse Fisher information matrix $J^{-1}(\phi_0, \phi_1, \lambda)$. std$^{\mathrm{B}}$: Sample standard deviation of the bootstrap samples, cf. (4.11). CI†: Normal-based bootstrap CI, cf. (4.12). CI‡: Non-normal-based bootstrap CI, cf. (4.13).

the Fisher scoring algorithm (4.5) and in 24 iterations for the EM algorithm (4.9)–(4.10), while the Newton–Raphson method is not available because the observed information matrix is nearly singular. The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ are given in the third column and 95% asymptotic Wald CIs (specified by (4.6)) of the three parameters are listed in the fourth column of Table 8. With $G = 6{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \lambda)$ are presented in the last two columns of Table 8.

Suppose we want to test H_0: $(\phi_0, \phi_1) = (0, 0)$ against H_1: $(\phi_0, \phi_1) \neq (0, 0)$. According to (5.3), we calculate the value of the score test statistic which is given by $t_1 = 217.3718$ and from (5.4), we have $p_{\mathrm{v}1} = 0 \ll 0.05$. Thus, the H_0 should be rejected at the significance level of $\alpha = 0.05$.

If we want to test the null hypothesis H_0: $\phi_1 = 0$ against H_1: $\phi_1 > 0$ at $\alpha = 0.05$. According to (5.6) and (5.9), we calculate the values of the likelihood ratio test statistic and score test statistic, which are given by $t_2 = 214.6707$ and $t_3 = 214.0573$, respectively. Then, from (5.8) and (5.10), we have $p_{\mathrm{v}2} = p_{\mathrm{v}3} = 0 \ll \alpha = 0.05$, resulting in a rejection of H_0.

If we want to test H_0: $\phi_0 = 0$ against H_1: $\phi_0 > 0$. According to (5.12) and (5.15), we calculate the values of the likelihood ratio test statistic and score test statistic which are given by $t_4 = 146.3721$ and $t_5 = 161.5884$, respectively. Then, from (5.14) and (5.16), we have $p_{\mathrm{v}4} = p_{\mathrm{v}5} = 0 \ll 0.05$. As a result, the H_0 should be rejected at the level of $\alpha = 0.05$.

7.1.2 Model comparison

We assess the goodness-of-fit by Pearson's chi-squared test (Pearson, [23]) via the predicted counts and we also use the Akaike information criterion (AIC; Akaike, [1]) and Bayesian information criterion (BIC; Schwarz, [27]) to compare models. A comparison of the fitted Poisson, ZIP and ZOIP distributions is shown in Table 9. According to Pearson's goodness-of-fit criterion, although the values of Pearson's chi-squared statistics for the three fitted distributions are sharply decreasing from the highest 792.9705 to 638.0514 to the lowest 131.1833, all of the three distributions exhibit lack of fit to the data. However, the tests for testing three hypotheses H_0: $(\phi_0, \phi_1) = (0, 0)$, H_0: $\phi_1 = 0$ and H_0: $\phi_0 = 0$ reject all three H_0's at the significant level $\alpha = 0.05$, indicating the existence of excess zeros and ones in the distribution. Moreover, both AIC and BIC criteria favor the ZOIP distribution.

Table 9. Comparison of the fitted frequencies from Poisson, ZIP and ZOIP distribution for the dentist visiting data

Count	Observed frequency	Poisson	ZIP	ZOIP
0	134	110.66	134.00	134.00
1	314	214.10	192.69	314.00
2	149	207.11	196.55	81.88
3	69	133.57	133.66	86.20
4	32	64.60	68.17	68.05
5	26	25.00	27.81	42.98
6	14	8.06	9.46	22.62
7	6	2.23	2.76	10.21
8-20	22	0.68	0.90	6.06
ϕ_0			0.0516	0.1535
ϕ_1				0.3422
λ		1.9347	2.0400	3.1580
Pearson's χ^2		792.97	638.05	131.18
d.f.		7	6	5
p-value		<0.001	<0.001	<0.001
AIC		3182.05	3175.78	**2963.11**
BIC		3186.70	3185.06	**2977.03**

Table 10. Criminal acts data (Dieckmann, 1981)

Count	0	1	2	3	4	5
Frequency	4037	219	29	9	5	2

7.2 Criminal acts data

Dieckmann [7] provided a data set from crime sociology consisting a sample of people with deviating behavior. Table 10 lists the distribution of number of criminal

Table 11. MLEs and CIs of parameters for the criminal acts data

Parameter	MLE	std^{F}	95% asymptotic Wald CI	std^{B}	95% bootstrap CI†	95% bootstrap CI‡
ϕ_0	0.9316	0.0053	[0.9212, 0.9420]	0.0064	[0.9176, 0.9428]	[0.9157, 0.9405]
ϕ_1	0.0415	0.0045	[0.0326, 0.0504]	0.0048	[0.0314, 0.0504]	[0.0309, 0.0498]
λ	1.3431	0.2447	[0.8635, 1.8227]	0.2466	[0.8537, 1.8202]	[0.8501, 1.8198]

std^{F}: Square roots of the diagonal elements of inverse Fisher information matrix $J^{-1}(\phi_0, \phi_1, \lambda)$. std^{B}: Sample standard deviation of the bootstrap samples, cf. (4.11). CI†: Normal-based bootstrap CI, cf. (4.12). CI‡: Non-normal-based bootstrap CI, cf. (4.13).

acts of 4301 persons with deviating behavior. Dieckmann [7] pointed out that the Poisson model does not fit the data well. Böhning [2] showed that the fit of the ZIP model is still not good in the upper classes (i.e., number of criminal acts 3, 4, and 5). By observing that the data have relatively higher proportions of zeros and ones, while the frequencies for other categories are quite low, we apply the ZOIP model to fit the data.

7.2.1 Likelihood-based inferences

Data in Table 10 exhibit excess zeros and ones, while the traditional Poisson model does not adequately capture such characteristics, so we consider the ZOIP model. Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{ZOIP}(\phi_0, \phi_1; \lambda)$. To find the MLEs of $(\phi_0, \phi_1, \lambda)$, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \lambda^{(0)}) = (0.3, 0.3, 3)$ as the initial values. The MLEs of $(\phi_0, \phi_1, \lambda)$ converged to $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ as shown in the second column of Table 11 in 6 iterations for the Fisher scoring algorithm (4.5) and in 249 iterations for the EM algorithm (4.9)–(4.10). The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ are given in the third column and 95% asymptotic Wald CIs (i.e. (4.6)) of the three parameters are listed in the fourth column of Table 11. With $G = 6{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \lambda)$ are shown in the last two columns of Table 11.

Let $\alpha = 0.05$. Suppose we want to test H_0: $(\phi_0, \phi_1) = (0, 0)$ against H_1: $(\phi_0, \phi_1) \neq (0, 0)$. According to (5.3), we calculate the value of the score test statistic which is given by $t_1 = 1848.2450$ and from (5.4), we have $p_{\text{v}1} = 0 \ll 0.05$. Thus, the H_0 should be rejected at the significance level of $\alpha = 0.05$.

Suppose we want to test the null hypothesis H_0: $\phi_1 = 0$ against the alternative hypothesis H_1: $\phi_1 > 0$. According to (5.6) and (5.9), we calculate the values of the likelihood ratio test and score test statistics, which are given by $t_2 = 25.5011$ and $t_3 = 30.0044$, respectively. Then from (5.8) and (5.10), we have $p_{\text{v}2} = p_{\text{v}3} = 0 \ll \alpha$. Thus, we should reject H_0.

7.2.2 Model comparison

We assess the model fitting by Pearson's chi-squared test, AIC and BIC. We list the fitted frequencies from Poisson, ZIP and ZOIP models in Table 12 and compute the corresponding p-values and MLEs of parameters. According to Pearson's goodness-of-fit criterion, the Poisson and ZIP models exhibit lack of fit to the data because of very small p-values while the ZOIP model shows a satisfactory fit. Furthermore, both the likelihood ratio test and the score test support the ZOIP model. Finally, both AIC and BIC favor the ZOIP model, so the ZOIP model is more favorable than the other two models.

Table 12. Comparison of the fitted frequencies from Poisson, ZIP and ZOIP distributions for the criminal acts data

Count	Observed frequency	Poisson	ZIP	ZOIP
0	4037	3979.64	4037.00	4037.00
1	219	309.04	204.54	219.00
2	29	12.00	50.15	27.28
3	9	0.31	8.20	12.21
4	5	0.01	1.01	4.10
5	2	0.00	0.12	1.40
Pearson's χ^2		46582.53	59.31	1.40
d.f.		4	3	2
p-value		<0.001	<0.001	0.4958
4+	7	0.01	1.11	5.51
Pearson's χ^2		8279.97	41.19	1.36
d.f.		3	2	1
p-value		<0.001	<0.001	0.2436
ϕ_0			0.8416	0.9316
ϕ_1				0.0415
λ		0.0777	0.4904	1.3431
AIC		2500.43	2346.80	**2323.30**
BIC		2506.80	2359.54	**2342.40**

7.3 Fetal lamb movement data

Leroux and Puterman [17] analyzed one particular sequence of counts in a study of breathing and body movements in fetal lambs designed to examine the possible changes in the amount and pattern of fetal activity during the last two-thirds of gestation. Table 13 lists the number of movements made by a fetal lamb observed through ultrasound in 240 consecutive 5-second intervals. Gupta *et al.* [11] further analyzed this data set by the zero-adjusted generalized Poisson distribution which results in a better fit than both the traditional Poisson distribution and a mixture of two Poisson distributions. Alternatively, we can apply the ZOIP distribution to fit this data set since there are extra zeros and ones.

Table 13. Number of movements made by a fetal lamb (Leroux and Puterman, 1992) and fitted frequencies from ZIP and ZOIP distributions

Number of movements	0	1	2	3	4	5	6	7
Number of intervals	182	41	12	2	2	0	0	1
Fitted frequencies (ZIP)	182.00	36.86	15.61	4.41	0.93	0.16	0.02	0.00
Fitted frequencies (ZOIP)	182.00	41.00	9.56	4.85	1.85	0.56	0.14	0.04

Table 14. MLEs and CIs of parameters for the fetal lamb movement data

Parameter	MLE	std^F	95% asymptotic Wald CI	std^B	95% bootstrap $\text{CI}^\dagger$	95% bootstrap $\text{CI}^\ddagger$
ϕ_0	0.7240	0.0407	[0.6442, 0.8038]	0.0570	[0.5965, 0.8198]	[0.5539, 0.7891]
ϕ_1	0.1185	0.0369	[0.0461, 0.1909]	0.0408	[0.0297, 0.1897]	[0.0099, 0.1820]
λ	1.5224	0.4142	[0.7106, 2.3342]	0.4150	[0.6955, 2.3224]	[0.7289, 2.3517]

std^F: Square roots of the diagonal elements of inverse Fisher information matrix $J^{-1}(\phi_0, \phi_1, \lambda)$. std^B: Sample standard deviation of the bootstrap samples, cf. (4.11). $\text{CI}^\dagger$: Normal-based bootstrap CI, cf. (4.12). $\text{CI}^\ddagger$: Non-normal-based bootstrap CI, cf. (4.13).

7.3.1 Likelihood-based inferences

To find the MLEs of $(\phi_0, \phi_1, \lambda)$, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \lambda^{(0)}) = (0.3, 0.3, 3)$ as the initial values. The MLEs of $(\phi_0, \phi_1, \lambda)$ converged to $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ as shown in the second column of Table 14 in 6 iterations for the Fisher scoring algorithm (4.5) and in 131 iterations for EM algorithm (4.9)–(4.10). The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ are given in the third column and 95% asymptotic Wald CIs (i.e. (4.6)) of the three parameters are listed in the fourth column of Table 14. With $G = 6{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \lambda)$ are reported in the last two columns of Table 14.

Let $\alpha = 0.05$. Suppose we want to test H_0: $(\phi_0, \phi_1) = (0, 0)$ against H_1: $(\phi_0, \phi_1) \neq (0, 0)$. According to (5.3), we calculate the value of the score test statistic which is given by $t_1 = 57.0687$ and from (5.4), we have $p_{\text{v}1} = 0 \ll 0.05$. Thus, the H_0 should be rejected at the significance level of $\alpha = 0.05$.

Suppose we want to test the null hypothesis H_0: $\phi_1 = 0$ against the alternative hypothesis H_1: $\phi_1 > 0$. According to (5.6) and (5.9), we calculate the values of the likelihood ratio test statistic and score test statistic, which are given by $t_2 = 4.9434$ and $t_3 = 5.1433$, respectively. Then from (5.8) and (5.10), we have $p_{\text{v}2} = 0.0131$ and $p_{\text{v}3} = 0.0233$, both are less than 0.05. Thus, we should reject H_0.

7.3.2 Model comparison

The model fitting is conducted by Pearson's chi-squared test, AIC and BIC. The corresponding results are listed in Table 15. According to Pearson's chi-squared test statistics, the ZIP model does not exhibit lack of fit, but the ZOIP model substantially improves the fit. In Section 7.3.1, we have shown that the tests for testing H_0: $(\phi_0, \phi_1) = (0, 0)$ and H_0: $\phi_1 = 0$ reject the H_0 at the significant level of $\alpha = 0.05$, indicating the existence of excess zeros and ones in the model. The AIC favors the ZOIP model, while the BIC favors the ZIP model. From Table 13, we can see that the number of zero counts and the number of one counts predicted by the ZOIP model are close to the corresponding observed counts, and the ZOIP model fits better than the ZIP model for most of other count categories. Therefore, the ZOIP model is a more reasonable alternative model.

Table 15. Results of the model fitting for the fetal lamb movement data

	ZIP	ZOIP
4+		
Pearson's χ^2	5.79	2.36
d.f.	2	1
p-value	0.0553	0.1242
5+		
Pearson's χ^2	7.46	2.40
d.f.	3	2
p-value	0.0585	0.3011
ϕ_0	0.5771	0.7240
ϕ_1		0.1185
λ	0.8473	1.5224
AIC	384.87	**381.93**
BIC	**391.84**	392.37

7.4 Death notice data of London Times

The data are the numbers of death notices of women 80 years of age and over, appearing in the London Times on each day for three consecutive years (Schilling, [26]). The data set was fitted by Hasselblad [12] with a mixture of two Poisson distributions. Subsequently, Gupta *et al.* [11] analyzed the data by the zero-adjusted generalized Poisson distribution with smaller variances of estimates when comparing with the mixture model. The counts are given in Table 16. Considering the relatively higher proportions of zero and one counts, we apply the ZOIP distribution to fit the data.

7.4.1 Likelihood-based inferences

To find the MLEs of parameters, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \lambda^{(0)}) = (0.3, 0.3, 3)$ as the initial values. The MLEs of

Table 16. Death notice data of London Times (Schilling, 1947)

Observed death count	0	1	2	3	4	5	6	7	8	9
Frequency	162	267	271	185	111	61	27	8	3	1

Table 17. MLEs and CIs of parameters for the death notice data of London Times

Parameter	MLE	std^{F}	95% asymptotic Wald CI	std^{B}	95% bootstrap $\text{CI}^{\dagger}$	95% bootstrap $\text{CI}^{\ddagger}$
ϕ_0	0.0660	0.0144	[0.0379, 0.0942]	0.0143	[0.0375, 0.0936]	[0.0371, 0.0931]
ϕ_1	0.0488	0.0212	[0.0072, 0.0904]	0.0207	[0.0074, 0.0885]	[0.0062, 0.0874]
λ	2.3816	0.0751	[2.2345, 2.5287]	0.0739	[2.2356, 2.5252]	[2.2354, 2.5231]

std^{F}: Square roots of the diagonal elements of inverse Fisher information matrix $J^{-1}(\phi_0, \phi_1, \lambda)$. std^{B}: Sample standard deviation of the bootstrap samples, cf. (4.11). $\text{CI}^{\dagger}$: Normal-based bootstrap CI, cf. (4.12). $\text{CI}^{\ddagger}$: Non-normal-based bootstrap CI, cf. (4.13).

$(\phi_0, \phi_1, \lambda)$ converged to $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ as shown in the second column of Table 17 in 5 iterations for the Fisher scoring algorithm (4.5) and in 200 iterations for the EM algorithm (4.9)–(4.10). The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ are given in the third column and 95% asymptotic Wald CIs (i.e. (4.6)) of the three parameters are listed in the fourth column of Table 17. With $G = 6{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \lambda)$ are provided in the last two columns of Table 17.

Let $\alpha = 0.05$. Suppose we want to test H_0: $(\phi_0, \phi_1) = (0, 0)$ against H_1: $(\phi_0, \phi_1) \neq (0, 0)$. According to (5.3), we calculate the value of the score test statistic which is given by $t_1 = 20.6166$ and from (5.4), we have $p_{\text{v}1} = 0 \ll 0.05$. Thus, the H_0 should be rejected at the significance level of $\alpha = 0.05$.

Suppose we want to test H_0: $\phi_1 = 0$ against H_1: $\phi_1 > 0$. According to (5.6) and (5.9), we calculate the values of the LRT statistic and score test statistic, which are given by $t_2 = 5.0760$ and $t_3 = 5.1068$, respectively. Then, from (5.8) and (5.10), we have $p_{\text{v}2} = 0.0121$ and $p_{\text{v}3} = 0.0238$, both are less than 0.05. Thus we should reject H_0.

7.4.2 Model comparison

The model fitting is conducted by Pearson's chi-squared test, AIC and BIC. We listed the predicted frequencies in Table 18 to compare Poisson, ZIP and ZOIP models. The computed Pearson's chi-square statistics indicate that the Poisson model exhibits lack of fit, the ZIP model does not exhibit lack of fit, but the ZOIP model substantially improves the fit. Moreover, both the LRT and the score test favor the ZOIP model. However, the AIC favors the ZOIP model, while the BIC favors the ZIP model.

Table 18. Comparison of the fitted frequencies from Poisson, ZIP and ZOIP distribution for the death notice data of London Times

Count	Observed frequency	Poisson	ZIP	ZOIP
0	162	126.78	162.00	162.00
1	267	273.47	244.38	267.00
2	271	294.92	277.29	254.22
3	185	212.04	209.76	201.82
4	111	114.34	119.01	120.17
5	61	49.33	54.02	57.24
6	27	17.73	20.43	22.72
7	8	5.46	6.62	7.73
8	3	1.47	1.88	2.30
9	1	0.45	0.61	0.79
Pearson's χ^2		26.47	9.92	4.54
d.f.		8	7	6
p-value		0.0009	0.1931	0.6044
7+	12	7.38	9.11	10.82
Pearson's χ^2		25.91	9.63	4.39
d.f.		6	5	4
p-value		0.0002	0.0864	0.3558
ϕ_0			0.0496	0.0660
ϕ_1				0.0488
λ		2.1569	2.2694	2.3816
AIC		4004.80	3992.10	**3989.03**
BIC		4009.80	**4002.10**	4004.03

7.5 Ammunition factory accidents data

Table 19 lists the number of accidents of 647 female workers in an ammunition factory (Greenwood and Yule, [10]). Since the simple Poisson model gave a bad fit to the data, Böhning [2] fitted the data by the ZIP model. From the result of Pearson's chi-squared test from Böhning [2], it seems that the ZIP model still has a lack of fit to the data, so we will apply the ZOIP model.

Table 19. Ammunition factory accidents data (Greenwood and Yule, 1920)

Count	0	1	2	3	4	5
Frequency	447	132	42	21	3	2

7.5.1 Likelihood-based inferences

To find the MLEs of parameters, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \lambda^{(0)}) = (0.3, 0.3, 3)$ as the initial values. The MLEs of $(\phi_0, \phi_1, \lambda)$ converged to $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ as shown in the second column of Table 20 in 6 iterations for the Fisher scoring

Table 20. MLEs and CIs of parameters for the ammunition factory accidents data

Parameter	MLE	std^F	95% asymptotic Wald CI	std^B	95% bootstrap CI†	95% bootstrap CI‡
ϕ_0	0.5969	0.0452	[0.5084, 0.6855]	0.0515	[0.4844, 0.6862]	[0.4639, 0.6647]
ϕ_1	0.0913	0.0347	[0.0233, 0.1594]	0.0359	[0.0156, 0.1565]	[0.0036, 0.1501]
λ	1.1994	0.1918	[0.8236, 1.5752]	0.1915	[0.8196, 1.5704]	[0.0036, 0.1501]

std^F: Square roots of the diagonal elements of inverse Fisher information matrix $J^{-1}(\phi_0, \phi_1, \lambda)$. std^B: Sample standard deviation of the bootstrap samples, cf. (4.11). CI†: Normal-based bootstrap CI, cf. (4.12). CI‡: Non-normal-based bootstrap CI, cf. (4.13).

algorithm (4.5) and in 412 iterations for the EM algorithm (4.9)–(4.10). The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\lambda})$ are given in the third column and 95% asymptotic Wald CIs (i.e. (4.6)) of the three parameters are listed in the fourth column of Table 20. With $G = 6{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \lambda)$ are presented in the last two columns of Table 20.

Let $\alpha = 0.05$. Suppose we want to test H_0: $(\phi_0, \phi_1) = (0, 0)$ against H_1: $(\phi_0, \phi_1) \neq (0, 0)$. According to (5.3), we calculate the value of the score test statistic which is given by $t_1 = 76.6301$ and from (5.4), we have $p_{v1} = 0 \ll 0.05$. Thus, the H_0 should be rejected at the significance level of $\alpha = 0.05$.

Suppose we want to test H_0: $\phi_1 = 0$ against H_1: $\phi_1 > 0$. According to (5.6) and (5.9), we calculate the values of the LRT statistic and score test statistic, which are given by $t_2 = 4.4298$ and $t_3 = 5.1068$, respectively. Then, from (5.8) and (5.10), we have $p_{v2} = 0.0177$ and $p_{v3} = 0.0238$, both are less than 0.05. Thus we should reject H_0.

7.5.2 Model comparison

The model fitting is conducted by Pearson's chi-squared test, AIC and BIC. We listed the predicted frequencies in Table 21 to compare Poisson, ZIP and ZOIP models. The computed Pearson's chi-square statistics indicate that the Poisson model exhibits a strong lack of fit and the ZIP model does not exhibit lack of fit, but the ZOIP model substantially improves the fit. Moreover, both the LRT and the score test favor the ZOIP model. However, the AIC favors the ZOIP model, while the BIC favors the ZIP model.

Table 21. Comparison of the fitted frequencies from Poisson, ZIP and ZOIP distribution for the ammunition factory accidents data

Count	Observed frequency	Poisson	ZIP	ZOIP
0	447	406.31	447.00	447.00
1	132	189.03	124.60	132.00
2	42	43.97	54.95	43.72
3	21	6.82	16.15	17.48
4	3	0.79	3.56	5.24
5	2	0.08	0.73	1.56
Pearson's χ^2		103.14	7.22	1.86
d.f.		4	3	2
p-value		<0.001	0.0653	0.3946
4+	5	0.87	4.30	6.80
Pearson's χ^2		70.37	5.06	1.25
d.f.		3	2	1
p-value		<0.001	0.0797	0.2629
ϕ_0			0.4725	0.5969
ϕ_1				0.0913
λ		0.4652	0.8820	1.1994
AIC		1236.37	1190.54	**1188.12**
BIC		1240.84	**1199.49**	1201.53

8. DISCUSSION

In this paper, we extensively study the zero-and-one inflated Poisson distribution by first establishing different but equivalent stochastic representations for the ZOIP random variable and then developing some important distributional properties. As we have seen, these stochastic representations play key roles in obtaining explicit expressions of the moments, the moment generating function, and in deriving an EM algorithm and so on. Useful results related to the conditional distributions based on the first two stochastic representations are also presented. Although the ZOIP model involves one more parameter indicating the existence of extra ones when comparing with the ZIP model, the MLEs and confidence intervals of parameters of interest can be easily obtained by the Fisher scoring algorithm or the EM algorithm, and the bootstrap methods developed in this paper. We also noted that the Pearson chi-squared goodness-of-fit test is more useful than both the AIC and BIC in model selection because the former can provide a p-value.

In Section 2.5, the SR (2.14) in fact defines a zero-and-one adjusted (or modified, or altered) Poisson distribution that can incorporate both inflation and deflation. Therefore, it is a further extension of the ZOIP model.

In Section 4.2, the MLEs of parameters in ZOIP model were obtained through the EM algorithm specified by (4.9) and (4.10). Once the EM converged, the iteration will arrive at an equilibrium, i.e.,

$$\hat{\phi}_0 = \frac{m_0 \hat{\phi}_0}{n[\hat{\phi}_0 + (1 - \hat{\phi}_0 - \hat{\phi}_1)\mathrm{e}^{-\hat{\lambda}}]} \quad \text{and}$$

$$\hat{\phi}_1 = \frac{m_1 \hat{\phi}_1}{n[\hat{\phi}_1 + (1 - \hat{\phi}_0 - \hat{\phi}_1)\hat{\lambda}\,\mathrm{e}^{-\hat{\lambda}}]}.$$

Therefore, we obtain

$$(8.1)\quad m_0 = n[\hat{\phi}_0 + (1-\hat{\phi}_0-\hat{\phi}_1)\mathrm{e}^{-\hat{\lambda}}] = n[\hat{\phi}_0 + \hat{\phi}_2\mathrm{e}^{-\hat{\lambda}}],$$

$$(8.2)\quad m_1 = n[\hat{\phi}_1 + (1-\hat{\phi}_0-\hat{\phi}_1)\hat{\lambda}\,\mathrm{e}^{-\hat{\lambda}}] = n[\hat{\phi}_1 + \hat{\phi}_2\hat{\lambda}\,\mathrm{e}^{-\hat{\lambda}}].$$

Recall that in the beginning of Section 4, $m_0 = \sum_{i=1}^n I(y_i = 0)$ denotes the number of zero observations in the observed data $Y_{\text{obs}} = \{y_i\}_{i=1}^n$, while $m_1 = \sum_{i=1}^n I(y_i = 1)$ denotes the number of 1 observations in Y_{obs}. On the other hand, from (1.1), we know that the probability of $\{Y = 0\}$ is $\phi_0 + \phi_2\mathrm{e}^{-\lambda}$ so that the estimated frequency of zero category should be $n[\hat{\phi}_0 + \hat{\phi}_2\mathrm{e}^{-\hat{\lambda}}]$, which must be equal to m_0 according to (8.1). Similar interpretation can be applied to m_1. For example, in Table 9, the observed frequencies for 0 and 1 are $m_0 = 134$ and $m_1 = 314$, while the fitted frequencies for 0 and 1 based on ZOIP are exactly 134 and 314. In addition, in a ZIP model (a special case of ZOIP), since we also apply the EM algorithm, it is not surprising that the fitted frequency for zero category is exactly equal to the observed frequency for the zero category.

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, an Associate Editor and one referee for their helpful comments and useful suggestions.

Received 27 March 2014

REFERENCES

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6) 716–723. MR0423716

[2] Böhning, D. (1998). Zero-inflated Poisson models and C.A.MAN: A tutorial collection of evidence. *Biometrical Journal* **40**(7) 833–843.

[3] Carrivick, P. J. W., Lee, A. H., and Yau, K. K. W. (2003). Zero-inflated Poisson models and C.A.MAN: A tutorial collection of evidence. *Safety Science* **41**(1) 53–63.

[4] Cohen Jr., A. C. (1960a). Estimating the parameters of a modified Poisson distribution. *Journal of the American Statistical Association* **55** 139–143. MR0111065

[5] Cohen Jr., A. C. (1960b). Estimation in the truncated Poisson distribution when zeros and some ones are missing. *Journal of the American Statistical Association* **55** 342–348. MR0121891

[6] David, F. N. and Johnson, N. L. (1952). The truncated Poisson. *Biometrics* **8**(4) 275–285. MR0052729

[7] Dieckmann, A. (1981). Ein einfaches stochastisches modell zur analyse von häufigkeitsverteilungen abweichenden verhaltens. *Zeitschrift für Soziologie* **10**(3) 319–325.

[8] Eriksson, R. and Åberg, R. (eds) (1987). *Welfare in Transition: A Survey of Living Conditions in Sweden 1968–1981*. Clarendon Press, Oxford.

[9] Feng, Z. D. and McCulloch, C. E. (1992). Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. *Statistics & Probability Letters* **13**(4) 325–332. MR1160755

[10] Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the Occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society* **83**(2) 255–279.

[11] Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1996). Analysis of zero-adjusted count data. *Computational Statistics & Data Analysis* **23**(2) 207–218.

[12] Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association* **64** 1459–1471.

[13] Hilbe, J. M. (2011). *Negative Binomial Regression*, 2nd ed. Cambridge University Press, New York. MR2797563

[14] Jansakul, N. and Hinde, J. P. (2002). Score tests for zero-inflated Poisson models. *Computational Statistics & Data Analysis* **40** 75–96. MR1921123

[15] Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal* **47**(2) 219–229. MR2137236

[16] Kemp, A. W. (1986). Weighted discrepancies and maximum likelihood estimation for discrete distributions. *Communications in Statistics – Theory and Methods* **15**(3) 783–803. MR0832079

[17] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1) 1–14.

[18] Leroux, B. G. and Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics* **48** 545–558.

[19] Martin, D. C. and Katti, S. K. (1965). Fitting of certain contagious distributions to some available data by the maximum likelihood method. *Biometrics* **21**(1) 34–48. MR0182101

[20] Melkersson, M. and Olsson, C. (1999). Is visiting the dentist a good habit? Analyzing count data with excess zeros and excess ones. Umeå Economic Studies, No. 492, Umeå University, Umeå, Sweden.

[21] Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics* **33**(3) 341–365. MR0867980

[22] Neyman, J. (1939). On a new class of "contagious" distributions, applicable in entomology and bacteriology. *The Annals of Mathematical Statistics* **10**(1) 35–57.

[23] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5* **50**(302) 157–175.

[24] Ridout, M., Demétrio, C. G. B., and Hinde, J. (1998). Models for count data with many zeros. Invited Paper in the XIX-th International Biometric Conference, Cape Town.

[25] Saito, M. Y. and Rodrigues, J. (2005). A Bayesian analysis of zero and one inflated distributions. *Rev. Mat. Estat.* **23**(1) 47–57. MR2304490

[26] Schilling, W. (1947). A frequency distribution represented as the sum of two Poisson distributions. *Journal of the American Statistical Association* **42** 407–424. MR0021280

[27] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**(2) 461–464. MR0468014

[28] Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of the maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82** 605–610. MR0898365

[29] Singh, S. N. (1963). A note on inflated Poisson distribution. *Journal of the Indian Statistical Association* **1**(3) 140–144. MR0162309

[30] Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**(398) 528–540. MR0898357

Chi Zhang
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road
Hong Kong
P.R. China
E-mail address: zhangchi@hku.hk

Guo-Liang Tian
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road
Hong Kong
P.R. China
E-mail address: gltian@hku.hk

Kai-Wang Ng
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road
Hong Kong
P.R. China
E-mail address: kaing@hku.hk