Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Likelihood-based methods for the zero-one-two inflated Poisson model with applications to biomedicine

Yuan Sun, Shishun Zhao, Guo-Liang Tian, Man-Lai Tang & Tao Li

To cite this article: Yuan Sun, Shishun Zhao, Guo-Liang Tian, Man-Lai Tang & Tao Li (2023) Likelihood-based methods for the zero-one-two inflated Poisson model with applications to biomedicine, Journal of Statistical Computation and Simulation, 93:6, 956-982, DOI: 10.1080/00949655.2021.1970162

To link to this article: https://doi.org/10.1080/00949655.2021.1970162

Published online: 05 Sep 2021.

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2023, VOL. 93, NO. 6, 956–982
https://doi.org/10.1080/00949655.2021.1970162

Taylor & Francis
Taylor & Francis Group

Check for updates

Likelihood-based methods for the zero-one-two inflated Poisson model with applications to biomedicine

Yuan Sun [a], Shishun Zhao [b], Guo-Liang Tian [c], Man-Lai Tang [d] and Tao Li [c]

[a]Department of Mathematics, Harbin Institute of Technology, Harbin, Heilongjiang, People's Republic of China; [b]School of Mathematics, Jilin University, Changchun, Jilin, People's Republic of China; [c]Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong, People's Republic of China; [d]Department of Mathematics, College of Engineering, Design & Physical Sciences, Brunel University London, Uxbridge, UK

ABSTRACT

To model count data with excess zeros, ones and twos, for the first time we introduce a so-called *zero-one-two-inflated Poisson* (ZOTIP) distribution, including the *zero-inflated Poisson* (ZIP) and the *zero-and-one-inflated Poisson* (ZOIP) distributions as two special cases. We establish three equivalent stochastic representations for the ZOTIP random variable to develop important distributional properties of the ZOTIP distribution. The Fisher scoring and *expectation–maximization* (EM) algorithms are derived to obtain the maximum likelihood estimates of parameters of interest. Bootstrap confidence intervals are also provided. Testing hypotheses are considered, simulation studies are conducted, and two real data sets are used to illustrate the proposed methods.

ARTICLE HISTORY
Received 30 May 2020
Accepted 16 August 2021

KEYWORDS
Bootstrap confidence intervals; EM algorithm; Fisher scoring algorithm; zero-and-one-inflated Poisson model; zero-one-two-inflated Poisson distribution.

1. Introduction

Lambert [1] proposed a so-called *zero-inflated Poisson* (ZIP) regression distribution to model count data with extra zeros when the traditional Poisson distribution could not fit the data with inflation at the zero points. The ZIP model and its variants such as zero-inflated negative binomial model [2] and zero-inflated generalized Poisson model [3] provide a set of tools to analysis such count data with excess zeros. When the count observations contain a high proportion of both zeros and ones, a so-called *zero-and-one-inflated Poisson* (ZOIP) distribution was proposed by Melkersson and Olsson [4]. Saito and Rodrigues [5] employed the Bayesian approach for the ZOIP model without considering any covariates by the data augmentation algorithm of Tanner and Wong [6]. Zhang et al. [7] extensively explored important distributional properties of the ZOIP and introduced likelihood-based inference methods. Later, Alshkaki [8] and Tang et al. [9] also studied the ZOIP distribution from structural properties and estimations of its parameters. With the increase of count data with excess zeros, the related research is more sufficient in recent

CONTACT Yuan Sun 11849457@mail.sustech.edu.cn Department of Mathematics, Harbin Institute of Technology, Harbin, Heilongjiang 150001, People's Republic of China

years. Rahayu et al. [10], Pittman et al. [11] and Hayati et al. [12] develop studies of overdispersion for Poisson, ZIP regression and Conway–Maxwell–Poisson models. A mediation analysis for count and zero-inflated count data are developed by Guo et al. [13] and Cheng et al. [14]. In addition, Sakthivel and Rajitha [15] propose an algorithm for selecting the better model among a set of models and computed the misclassification rates for a zero-inflated count data set using different classifiers. Lee et al. [16] develop a Bayesian variable selection model for multivariate count data with excess zeros.

Although the known models are more flexible than the traditional distributions when there are excess zeros, there are real count data sets, where there exist not only excess zeros and ones but also excess twos. Count data with excess values occur in biomedicine, economics, and social sciences [17–23]. For example, Morgan et al. [24] reported a data set (see Table 5), showing the distribution of stillbirths in 402 litters of New Zealand white rabbits. The frequencies for the counts zero, one and two are 314, 48 and 20, respectively; i.e. the corresponding proportions are approximately 78%, 12% and 5%. Neither the Poisson, ZIP nor the ZOIP is significant when they are utilized to fit this data set, in terms of the p-values of the Pearson chi-square goodness-of-fit test (see Table 7). As the second example, Eriksson and Åberg [25] reported a two-year panel data from Swedish Level of Living Surveys in 1974 and 1981. Table 8 only displayed the data in 1981, where the visits to a dentist have higher proportions of zeros (17.5%), ones (41%), twos (19.5%), and two-visit observations are even more frequent than zero-visit. It is common to visit a dentist for a routine control, e.g. school-children go to a dentist once/twice a year almost as a rule.

Therefore, neither the Poisson, ZIP nor the ZOIP is appropriate to model the count data with excess zeros, ones and twos. Motivated by count data with excess twos, in this paper, we aim to introduce a new distribution, called *zero-one-two inflated Poisson* (ZOTIP) distribution.

Let ξ be a random variable with *probability mass function* (pmf) $\Pr(\xi = c) = 1$ with c being a constant; i.e. ξ follows a degenerate distribution at a single point c, we denote it by $\xi \sim \text{Degenerate}(c)$.

For $i = 0, 1, 2$, let $\xi_i \sim \text{Degenerate}(i)$, $X \sim \text{Poisson}(\lambda)$ and they are independent. A discrete random variable Y is said to follow a ZOTIP distribution, denoted by $Y \sim \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$, if its pmf is

$$\begin{aligned}
f(y|\phi_0, \phi_1, \phi_2; \lambda) &= \sum_{i=0}^{2} \phi_i \Pr(\xi_i = y) + \phi_3 \Pr(X = y) \\
&= \begin{cases} \phi_0 + \phi_3 \mathrm{e}^{-\lambda}, & \text{if } y = 0, \\ \phi_1 + \phi_3 \lambda \, \mathrm{e}^{-\lambda}, & \text{if } y = 1, \\ \phi_2 + \phi_3 \lambda^2 \mathrm{e}^{-\lambda}/2, & \text{if } y = 2, \\ \phi_3 \lambda^y \mathrm{e}^{-\lambda}/y!, & \text{if } y = 3, 4, \ldots, \end{cases} \\
&= (\phi_0 + \phi_3 \mathrm{e}^{-\lambda}) I(y = 0) + (\phi_1 + \phi_3 \lambda \, \mathrm{e}^{-\lambda}) I(y = 1) \\
&\quad + (\phi_2 + \phi_3 \lambda^2 \mathrm{e}^{-\lambda}/2) I(y = 2) + (\phi_3 \lambda^y \mathrm{e}^{-\lambda}/y!) I(y \geqslant 3), \qquad (1)
\end{aligned}$$

where the parameters $\phi_0 \in [0,1)$, $\phi_1 \in [0,1)$ and $\phi_2 \in [0,1)$ represent the corresponding proportions for incorporating extra zeros, extra ones and extra twos, and $\phi_3 \hat{=} 1-\phi_0-\phi_1-\phi_2 \in (0,1]$. It is easy to see that the ZOTIP is a mixture of three degenerate distributions with all mass at zero, at one and at two, respectively, and a Poisson(λ) distribution. In particular, when $\phi_2=0$, the ZOTIP is reduced to the ZOIP distribution (denoted by ZOIP($\phi_0, \phi_1; \lambda$)); when $\phi_1=\phi_2=0$, the ZOTIP distribution is reduced to ZIP distribution (denoted by ZIP(ϕ_0, λ)); when $\phi_0=\phi_1=\phi_2=0$, the ZOTIP distribution becomes the traditional Poisson distribution.

The rest of this paper is organized as follows. Section 2 develops important distributional properties including one stochastic representation for the ZOTIP random variable, the cumulative distribution function, moments, moment generating function, and some useful conditional distributions. In Section 3, we derive the Fisher scoring algorithm and the *expectation–maximization* (EM) algorithm for calculating the *maximum likelihood estimates* (MLEs) of parameters of interest. Bootstrap confidence intervals are also presented. The likelihood ratio test and the score test for two-inflation and for simultaneous zero-one-two inflation are provided in Section 4. Simulation studies and comparisons studies are conducted in Section 5. In Section 6, two real data sets are used to illustrate the proposed methods. A discussion is given in Section 7. Some technical details are put in Appendix.

2. Some distributional properties

In this section, we first introduce a *stochastic representation* (SR) for the random variable $Y \sim$ ZOTIP($\phi_0, \phi_1, \phi_2; \lambda$). Next, we derive the cumulative distribution function, moments, moment generating function, and some useful conditional distributions.

2.1. A stochastic representation

Let the random vector $\mathbf{z}=(Z_0, Z_1, Z_2, Z_3)^\top \sim$ Multinomial$_4(1; \phi_0, \phi_1, \phi_2, \phi_3)$, the random variable $X \sim$ Poisson(λ), and $\mathbf{z}$ and X be mutually independent (denoted by $\mathbf{z} \perp\!\!\!\perp X$). It is easy to show that an SR of the random variable $Y \sim$ ZOTIP($\phi_0, \phi_1, \phi_2; \lambda$) is given by

$$Y \stackrel{\mathrm{d}}{=} Z_0 \cdot 0 + Z_1 \cdot 1 + Z_2 \cdot 2 + Z_3 X = \begin{cases} 0, & \text{with probability } \phi_0, \\ 1, & \text{with probability } \phi_1, \\ 2, & \text{with probability } \phi_2, \\ X, & \text{with probability } \phi_3, \end{cases} \tag{2}$$

where the notation '$\stackrel{\mathrm{d}}{=}$' indicates that the random variables in both sides of the equality have the same distribution. In fact, noting that $Z_0+Z_1+Z_2+Z_3=1$ and $\Pr(Z_i=1)=\phi_i$ for $i=0, 1, 2, 3$, we obtain from (2) that

$$\begin{cases} \Pr(Y=0) & = \Pr(Z_0=1)+\Pr(Z_3=1, X=0) = \phi_0+\phi_3 \mathrm{e}^{-\lambda}, \\ \Pr(Y=1) & = \Pr(Z_1=1)+\Pr(Z_3=1, X=1) = \phi_1+\phi_3 \lambda\, \mathrm{e}^{-\lambda}, \\ \Pr(Y=2) & = \Pr(Z_2=1)+\Pr(Z_3=1, X=2) = \phi_2+\phi_3 \lambda^2 \mathrm{e}^{-\lambda}/2, \\ \Pr(Y=y) & = \Pr(Z_3=1, X=y) = \phi_3 \lambda^y \mathrm{e}^{-\lambda}/y!, \quad y \geqslant 3. \end{cases} \tag{3}$$

Since (3) is identical to (1), we know that the SR (2) is true. The SR (2) means that $\text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ is a mixture of four distributions: Degenerate(0), Degenerate(1), Degenerate(2) and Poisson(λ). The other two different but equivalent SRs are presented in Appendix 1.

2.2. The cumulative distribution function

Assume that $Y \sim \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$, it is easy to show that for any non-negative real number y, the cumulative distribution function of Y is

$$\begin{aligned}
\Pr(Y \leqslant y) &= (\phi_0 + \phi_3 \mathrm{e}^{-\lambda}) I(0 \leqslant y < 1) + (\phi_0 + \phi_1 + \phi_3 \mathrm{e}^{-\lambda} + \phi_3 \lambda \mathrm{e}^{-\lambda}) I(1 \leqslant y < 2) \\
&\quad + \left(\phi_0 + \phi_1 + \phi_2 + \phi_3 \sum_{i=0}^{\lfloor y \rfloor} \frac{\lambda^i \mathrm{e}^{-\lambda}}{i!} \right) I(y \geqslant 2) \\
&= (\phi_0 + \phi_3 \mathrm{e}^{-\lambda}) I(0 \leqslant y < 1) + (\phi_0 + \phi_1 + \phi_3 \mathrm{e}^{-\lambda} + \phi_3 \lambda \mathrm{e}^{-\lambda}) I(1 \leqslant y < 2) \\
&\quad + \left[\phi_0 + \phi_1 + \phi_2 + \phi_3 \frac{\Gamma(\lfloor y + 1 \rfloor, \lambda)}{\lfloor y \rfloor !} \right] I(y \geqslant 2),
\end{aligned}$$

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

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

is the upper incomplete gamma function.

2.3. Moments and moment generating function

Based on the SR (2), we know that $(Z_0, Z_1, Z_2, Z_3)^\top \sim \text{Multinomial}_4(1; \phi_0, \phi_1, \phi_2, \phi_3)$, it is easy to see that $E(Z_i) = \phi_i$, $\text{Var}(Z_i) = \phi_i(1 - \phi_i)$, $\text{Cov}(Z_i, Z_j) = -\phi_i \phi_j$ and $E(Z_i Z_j) = 0$ for $i \neq j$, where $i, j = 0, 1, 2, 3$. We can obtain

$$\begin{aligned}
E(Y) &= \phi_1 + 2\phi_2 + \phi_3 \lambda \;\hat{=}\; \mu, \\
E(Y^2) &= \phi_1 + 4\phi_2 + \phi_3(\lambda + \lambda^2), \\
\text{Var}(Y) &= \mu - \mu^2 + 2\phi_2 + \frac{(\mu - \phi_1 - 2\phi_2)^2}{\phi_3}.
\end{aligned} \tag{4}$$

Let r and s be two arbitrary positive integers, we have $Z_i^r Z_j^s \sim \text{Degenerate}(0)$ for $i \neq j$, where $i, j = 0, 1, 2, 3$. For any positive integer n, we have

$$E(Y^n) \overset{(2)}{=} \sum_{k=0}^{n} \binom{n}{k} E\Big[(Z_1 + 2Z_2)^k Z_3^{n-k} \Big] E(X^{n-k}) = \phi_1 + 2^n \phi_2 + \phi_3 E(X^n),$$

which can be derived alternatively based on (A3) by noting that $(1 - Z)^r Z^s \sim \text{Degenerate}(0)$ for any Bernoulli random variable Z, where r and s are two arbitrary positive integers.

Using the formula of $E(W_1) = E[E(W_1|W_2)]$, we obtain the moment generating function of $Y \sim \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ as

$$\begin{aligned}
M_Y(t) &= E[\exp(tY)] \\
&\overset{\text{(A3)}}{=} E\Big\{\exp[t(1-Z)V + tZX]\Big\} \\
&= E\Big\{E\{\exp[t(1-Z)V + tZX] \,|\, Z\}\Big\} \\
&= E[M_V(t(1-Z)) \cdot M_X(tZ)] \\
&= E\left\{\Big[p_0 + p_1\, \mathrm{e}^{t(1-Z)} + (1-p_0-p_1)\mathrm{e}^{2t(1-Z)}\Big] \cdot \exp[\lambda(\mathrm{e}^{tZ}-1)]\right\} \\
&= \phi[p_0 + p_1\, \mathrm{e}^t + (1-p_0-p_1)\mathrm{e}^{2t}] + (1-\phi)\exp[\lambda\,(\mathrm{e}^t-1)],
\end{aligned}$$

where $V \sim \text{TP}(p_0, p_1)$ with parameters p_0 and p_1 defined in (A3).

2.4. Conditional distributions

Based on the results in Section 2.1, we can derive the conditional distributions of $\mathbf{z}|Y$, $Z_i|Y$ and $X|Y$, and the corresponding proofs are presented in Appendix 2.

Proposition 2.1 (Joint conditional distribution of z|*Y*): *Let $Y \sim ZOTIP(\phi_0, \phi_1, \phi_2; \lambda)$ and $\mathbf{z} = (Z_0, Z_1, Z_2, Z_3)^\top \sim Multinomial_4(1; \phi_0, \phi_1, \phi_2, \phi_3)$. Based on the SR (2), we have*

$$\mathbf{z}|(Y=y) \sim \begin{cases} Multinomial_4(1; \psi_1, 0, 0, 1-\psi_1), & \text{if } y = 0, \\ Multinomial_4(1; 0, \psi_2, 0, 1-\psi_2), & \text{if } y = 1, \\ Multinomial_4(1; 0, 0, \psi_3, 1-\psi_3), & \text{if } y = 2, \\ Multinomial_4(1; 0, 0, 0, 1), & \text{if } y \geqslant 3, \end{cases} \tag{5}$$

where

$$\psi_1 \mathrel{\hat{=}} \frac{\phi_0}{\phi_0 + \phi_3 e^{-\lambda}}, \quad \psi_2 \mathrel{\hat{=}} \frac{\phi_1}{\phi_1 + \phi_3 \lambda\, e^{-\lambda}} \quad \text{and} \quad \psi_3 \mathrel{\hat{=}} \frac{\phi_2}{\phi_2 + \phi_3 \lambda^2\, e^{-\lambda}/2}. \tag{6}$$

Corollary 2.1 (Marginal conditional distributions of Z_i|*Y*): *Let $Y \sim ZOTIP(\phi_0, \phi_1, \phi_2; \lambda)$ and $\mathbf{z} = (Z_0, Z_1, Z_2, Z_3)^\top \sim Multinomial_4(1; \phi_0, \phi_1, \phi_2, \phi_3)$. Based on the SR (2), we have*

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

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

$$Z_2|(Y=y) \sim \begin{cases} Bernoulli(\psi_3), & \text{if } y = 2, \\ Degenerate(0), & \text{if } y \neq 2, \end{cases}$$

$$Z_3|(Y=y) \sim \begin{cases} \mathit{Bernoulli}(1-\psi_1), & \text{if } y=0, \\ \mathit{Bernoulli}(1-\psi_2), & \text{if } y=1, \\ \mathit{Bernoulli}(1-\psi_3), & \text{if } y=2, \\ \mathit{Degenerate}(1), & \text{if } y \geqslant 3, \end{cases}$$

where ψ_1, ψ_2 *and* ψ_3 *are given by (6).*

Proposition 2.2 (Conditional distribution of $X|Y$): *Let* $Y \sim \mathrm{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ *and* $X \sim \mathit{Poisson}(\lambda)$. *Based on the SR (2), we have*

$$X|(Y=y) \sim \begin{cases} \mathit{ZIP}(1-\psi_1, \lambda), & \text{if } y=0, \\ \mathit{OIP}(1-\psi_2, \lambda), & \text{if } y=1, \\ \mathit{TIP}(1-\psi_3, \lambda), & \text{if } y=2, \\ \mathit{Degenerate}(y), & \text{if } y \geqslant 3, \end{cases}$$

where ψ_1, ψ_2 *and* ψ_3 *are given by (6).*

3. Likelihood-based estimation

Assume that $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ and $y_1, \ldots, y_n$ denote their realizations. Let $Y_{\text{obs}} = \{y_i\}_{i=1}^n$ be the observed data, and define $\mathbb{I}_t \hat{=} \{i: y_i = t,\ 1 \leqslant i \leqslant n\}$ and $m_t = \sum_{i=1}^n I(y_i = t)$ is the number of elements in $\mathbb{I}_t$, $t = 0, 1, 2$. The observed-data likelihood function for $\boldsymbol{\theta} \hat{=} (\phi_0, \phi_1, \phi_2, \lambda)^\top$ is given by

$$L(\boldsymbol{\theta}|Y_{\text{obs}}) = (\phi_0 + \phi_3 \mathrm{e}^{-\lambda})^{m_0} \times (\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda})^{m_1} \times (\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2)^{m_2} \times \phi_3^{n-m_0-m_1-m_2} \prod_{i \notin \mathbb{I}_0 \cup \mathbb{I}_1 \cup \mathbb{I}_2} \frac{\lambda^{y_i} \mathrm{e}^{-\lambda}}{y_i!}, \tag{7}$$

so we obtain

$$\begin{aligned} \ell = \ell(\boldsymbol{\theta}|Y_{\text{obs}}) &\hat{=} \log[L(\boldsymbol{\theta}|Y_{\text{obs}})] \\ &= m_0 \log(\phi_0 + \phi_3 \mathrm{e}^{-\lambda}) + m_1 \log(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda}) + m_2 \log(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2) \\ &\quad + (n - m_0 - m_1 - m_2)[\log(\phi_3) - \lambda] + N \log(\lambda) - \sum_{i \notin \mathbb{I}_0 \cup \mathbb{I}_1 \cup \mathbb{I}_2} \log(y_i!), \end{aligned}$$

where $\phi_3 = 1 - \phi_0 - \phi_1 - \phi_2$ and $N = \sum_{i \notin \mathbb{I}_0 \cup \mathbb{I}_1 \cup \mathbb{I}_2} y_i$. The following results will be used in the calculation of the Fisher information matrix and the derivation is given in Appendix 2.

Proposition 3.1 (Expectations): *The expectations of* m_0, m_1, m_2 *and* N *are given by*

$$\begin{aligned} &E(m_0) = n(\phi_0 + \phi_3 e^{-\lambda}), \quad E(m_1) = n(\phi_1 + \phi_3 \lambda\, e^{-\lambda}), \\ &E(m_2) = n(\phi_2 + \phi_3 \lambda^2\, e^{-\lambda}/2), \quad E(N) = n\phi_3 \lambda (1 - e^{-\lambda} - \lambda\, e^{-\lambda}). \end{aligned} \tag{8}$$

3.1. MLEs via the Fisher scoring algorithm

In this section, we use the *Fisher scoring* (FS) algorithm to calculate the MLE of $\boldsymbol{\theta}$. The score vector $\nabla\ell$ and the Hessian matrix $\nabla^2\ell$ are given by

$$\nabla\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = \left(\frac{\partial\ell}{\partial\phi_0}, \frac{\partial\ell}{\partial\phi_1}, \frac{\partial\ell}{\partial\phi_2}, \frac{\partial\ell}{\partial\lambda}\right)^\top \quad \text{and}$$

$$\nabla^2\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = \begin{pmatrix} \frac{\partial^2\ell}{\partial\phi_0^2} & \frac{\partial^2\ell}{\partial\phi_0\partial\phi_1} & \frac{\partial^2\ell}{\partial\phi_0\partial\phi_2} & \frac{\partial^2\ell}{\partial\phi_0\partial\lambda} \\ \frac{\partial^2\ell}{\partial\phi_1\partial\phi_0} & \frac{\partial^2\ell}{\partial\phi_1^2} & \frac{\partial^2\ell}{\partial\phi_1\partial\phi_2} & \frac{\partial^2\ell}{\partial\phi_1\partial\lambda} \\ \frac{\partial^2\ell}{\partial\phi_2\partial\phi_0} & \frac{\partial^2\ell}{\partial\phi_2\partial\phi_1} & \frac{\partial^2\ell}{\partial\phi_2^2} & \frac{\partial^2\ell}{\partial\phi_2\partial\lambda} \\ \frac{\partial^2\ell}{\partial\lambda\partial\phi_0} & \frac{\partial^2\ell}{\partial\lambda\partial\phi_1} & \frac{\partial^2\ell}{\partial\lambda\partial\phi_2} & \frac{\partial^2\ell}{\partial\lambda^2} \end{pmatrix},$$

where

$$\frac{\partial\ell}{\partial\phi_0} = \frac{m_0(1-\mathrm{e}^{-\lambda})}{\phi_0+\phi_3\mathrm{e}^{-\lambda}} - \frac{m_1\lambda\,\mathrm{e}^{-\lambda}}{\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}} - \frac{m_2\lambda^2\,\mathrm{e}^{-\lambda}/2}{\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} - \frac{n-m_0-m_1-m_2}{\phi_3},$$

$$\frac{\partial\ell}{\partial\phi_1} = -\frac{m_0\mathrm{e}^{-\lambda}}{\phi_0+\phi_3\mathrm{e}^{-\lambda}} + \frac{m_1(1-\lambda\,\mathrm{e}^{-\lambda})}{\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}} - \frac{m_2\lambda^2\,\mathrm{e}^{-\lambda}/2}{\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} - \frac{n-m_0-m_1-m_2}{\phi_3},$$

$$\frac{\partial\ell}{\partial\phi_2} = -\frac{m_0\mathrm{e}^{-\lambda}}{\phi_0+\phi_3\mathrm{e}^{-\lambda}} - \frac{m_1\lambda\,\mathrm{e}^{-\lambda}}{\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}} - \frac{m_2(1-\lambda^2\,\mathrm{e}^{-\lambda}/2)}{\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} - \frac{n-m_0-m_1-m_2}{\phi_3},$$

$$\frac{\partial\ell}{\partial\lambda} = -\frac{m_0\phi_3\mathrm{e}^{-\lambda}}{\phi_0+\phi_3\mathrm{e}^{-\lambda}} + \frac{m_1\phi_3(1-\lambda)\mathrm{e}^{-\lambda}}{\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}} + \frac{m_2\phi_3(\lambda-\lambda^2/2)\mathrm{e}^{-\lambda}}{\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2}$$
$$- (n-m_0-m_1-m_2) + \frac{N}{\lambda},$$

$$\frac{\partial^2\ell}{\partial\phi_0^2} = -\frac{m_0(1-\mathrm{e}^{-\lambda})^2}{\left(\phi_0+\phi_3\mathrm{e}^{-\lambda}\right)^2} - \frac{m_1\lambda^2\mathrm{e}^{-2\lambda}}{\left(\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}\right)^2} - \frac{m_2\lambda^4\mathrm{e}^{-2\lambda}}{4\left(\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2\right)^2}$$
$$- \frac{n-m_0-m_1-m_2}{\phi_3^2},$$

$$\frac{\partial^2\ell}{\partial\phi_1^2} = -\frac{m_0\mathrm{e}^{-2\lambda}}{\left(\phi_0+\phi_3\mathrm{e}^{-\lambda}\right)^2} - \frac{m_1(1-\lambda\mathrm{e}^{-\lambda})^2}{\left(\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}\right)^2} - \frac{m_2\lambda^4\mathrm{e}^{-2\lambda}}{4\left(\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2\right)^2}$$
$$- \frac{n-m_0-m_1-m_2}{\phi_3^2},$$

$$\frac{\partial^2\ell}{\partial\phi_2^2} = -\frac{m_0\mathrm{e}^{-2\lambda}}{\left(\phi_0+\phi_3\mathrm{e}^{-\lambda}\right)^2} - \frac{m_1\lambda^2\mathrm{e}^{-2\lambda}}{\left(\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda}\right)^2} - \frac{m_2(1-\lambda^2\mathrm{e}^{-\lambda}/2)^2}{\left(\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2\right)^2}$$

$$- \frac{n - m_0 - m_1 - m_2}{\phi_3^2},$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = \frac{m_0 \phi_0 \phi_3 \mathrm{e}^{-\lambda}}{\left(\phi_0 + \phi_3 \mathrm{e}^{-\lambda}\right)^2} + \frac{m_1 \phi_3 \mathrm{e}^{-\lambda}\left[\phi_1(\lambda - 2) - \phi_3 \mathrm{e}^{-\lambda}\right]}{\left(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda}\right)^2}$$

$$+ \frac{m_2 \phi_3 \mathrm{e}^{-\lambda}\left[\phi_2(2 - 4\lambda + \lambda^2) - \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}\right]}{2\left[\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2\right]^2} - \frac{N}{\lambda^2},$$

$$\frac{\partial^2 \ell}{\partial \phi_0 \partial \phi_1} = \frac{m_0 \mathrm{e}^{-\lambda}(1 - \mathrm{e}^{-\lambda})}{\left(\phi_0 + \phi_3 \mathrm{e}^{-\lambda}\right)^2} + \frac{m_1 \lambda\, \mathrm{e}^{-\lambda}(1 - \lambda\, \mathrm{e}^{-\lambda})}{\left(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda}\right)^2} - \frac{m_2 \lambda^4\, \mathrm{e}^{-2\lambda}}{4\left(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2\right)^2}$$

$$- \frac{n - m_0 - m_1 - m_2}{\phi_3^2},$$

$$\frac{\partial^2 \ell}{\partial \phi_0 \partial \phi_2} = \frac{m_0 \mathrm{e}^{-\lambda}(1 - \mathrm{e}^{-\lambda})}{\left(\phi_0 + \phi_3 \mathrm{e}^{-\lambda}\right)^2} - \frac{m_1 \lambda^2\, \mathrm{e}^{-2\lambda}}{\left(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda}\right)^2} + \frac{m_2 \lambda^2\, \mathrm{e}^{-\lambda}(1 - \lambda^2\, \mathrm{e}^{-\lambda}/2)}{2\left(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2\right)^2}$$

$$- \frac{n - m_0 - m_1 - m_2}{\phi_3^2},$$

$$\frac{\partial^2 \ell}{\partial \phi_0 \partial \lambda} = \frac{m_0(\phi_0 + \phi_3)\mathrm{e}^{-\lambda}}{(\phi_0 + \phi_3 \mathrm{e}^{-\lambda})^2} + \frac{m_1 \phi_1(\lambda - 1)\mathrm{e}^{-\lambda}}{(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda})^2} + \frac{m_2 \phi_2 \lambda\,(\lambda/2 - 1)\mathrm{e}^{-\lambda}}{(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2)^2},$$

$$\frac{\partial^2 \ell}{\partial \phi_1 \partial \phi_2} = \frac{m_0 \mathrm{e}^{-2\lambda}}{\left(\phi_0 + \phi_3 \mathrm{e}^{-\lambda}\right)^2} - \frac{m_1 \lambda\,(1 - \lambda\, \mathrm{e}^{-\lambda})\mathrm{e}^{-\lambda}}{\left(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda}\right)^2} + \frac{m_2 \lambda^2\, \mathrm{e}^{-\lambda}(1 - \lambda^2\, \mathrm{e}^{-\lambda}/2)}{2\left(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2\right)^2}$$

$$- \frac{n - m_0 - m_1 - m_2}{\phi_3^2},$$

$$\frac{\partial^2 \ell}{\partial \phi_1 \partial \lambda} = \frac{m_0 \phi_0 \mathrm{e}^{-\lambda}}{(\phi_0 + \phi_3 \mathrm{e}^{-\lambda})^2} + \frac{m_1(\phi_1 + \phi_3)(\lambda - 1)\mathrm{e}^{-\lambda}}{(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda})^2} + \frac{m_2 \phi_2 \lambda\,(\lambda/2 - 1)\mathrm{e}^{-\lambda}}{(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2)^2},$$

$$\frac{\partial^2 \ell}{\partial \phi_2 \partial \lambda} = \frac{m_0 \phi_0 \mathrm{e}^{-\lambda}}{(\phi_0 + \phi_3 \mathrm{e}^{-\lambda})^2} + \frac{m_1 \phi_1(\lambda - 1)\mathrm{e}^{-\lambda}}{(\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda})^2} + \frac{m_2(\phi_2 + \phi_3)\lambda\,(\lambda/2 - 1)\mathrm{e}^{-\lambda}}{(\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2)^2}.$$

With the expectations in (8), we can calculate the Fisher information as follows:

$$\mathbf{J}(\boldsymbol{\theta}) = E\Big[- \nabla^2 \ell(\boldsymbol{\theta} | Y_{\text{obs}})\Big].$$

Let $\boldsymbol{\theta}^{(0)} = (\phi_0^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \lambda^{(0)})$ be the initial values of the MLEs $\hat{\boldsymbol{\theta}} = (\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})^\top$ and $\boldsymbol{\theta}^{(t)} = (\phi_0^{(t)}, \phi_1^{(t)}, \phi_2^{(t)}, \lambda^{(t)})^\top$ be the t-th approximations of $\hat{\boldsymbol{\theta}}$, then the $(t+1)$-th approximation can be obtained by the FS algorithm

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \mathbf{J}^{-1}(\boldsymbol{\theta}^{(t)}) \nabla \ell(\boldsymbol{\theta}^{(t)} | Y_{\text{obs}}). \tag{9}$$

The estimated standard errors of the MLEs $\hat{\boldsymbol{\theta}}$ are the square root of the diagonal elements J^{kk} of the inverse Fisher information matrix $\mathbf{J}^{-1}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The $(1-\alpha)100\%$ asymptotic Wald *confidence intervals* (CIs) of ϕ_0, ϕ_1, ϕ_2 and λ are given by

$$\left[\hat{\phi}_{k-1} - z_{\alpha/2}\sqrt{J^{kk}},\ \hat{\phi}_{k-1} + z_{\alpha/2}\sqrt{J^{kk}}\right],\ k = 1,2,3 \quad \text{and} \quad \left[\hat{\lambda} - z_{\alpha/2}\sqrt{J^{44}}, \hat{\lambda} + z_{\alpha/2}\sqrt{J^{44}}\right], \tag{10}$$

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

3.2. MLEs via the EM algorithm

The zero (one or two) observations from any ZOTIP distribution can be classed into the *extra* zeros (ones or twos) which are from the degenerate distribution at the point zero (one or two) because of population variability and the *structural* zeros (ones or twos) resulted from an ordinary Poisson distribution. Thus, we obtain

$$\mathbb{I}_0 = \mathbb{I}_0^{\text{extra}} \cup \mathbb{I}_0^{\text{structral}}, \quad \mathbb{I}_1 = \mathbb{I}_1^{\text{extra}} \cup \mathbb{I}_1^{\text{structral}} \quad \text{and} \quad \mathbb{I}_2 = \mathbb{I}_2^{\text{extra}} \cup \mathbb{I}_2^{\text{structral}}.$$

The main difficulty in obtaining closed-form solutions of MLEs of parameters in (7) is the first three terms on the right hand side of (7). To tackle this question, we introduce three latent variables W_0, W_1 and W_2, where W_t denotes the number of $\mathbb{I}_t^{\text{extra}}$ to split m_t into W_t and $m_t - W_t$ for $t = 0, 1, 2$. Thus, the resulting conditional predictive distributions of W_0, W_1 and W_2 given $(Y_{\text{obs}}, \boldsymbol{\theta})$ are given by

$$W_0|(Y_{\text{obs}}, \boldsymbol{\theta}) \sim \text{Binomial}\left(m_0, \frac{\phi_0}{\phi_0 + \phi_3 \mathrm{e}^{-\lambda}}\right),$$
$$W_1|(Y_{\text{obs}}, \boldsymbol{\theta}) \sim \text{Binomial}\left(m_1, \frac{\phi_1}{\phi_1 + \phi_3 \lambda\, \mathrm{e}^{-\lambda}}\right),$$
$$W_2|(Y_{\text{obs}}, \boldsymbol{\theta}) \sim \text{Binomial}\left(m_2, \frac{\phi_2}{\phi_2 + \phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2}\right).$$

The complete-data likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}|Y_{\text{com}}) &\propto \phi_0^{w_0}(\phi_3 \mathrm{e}^{-\lambda})^{m_0 - w_0} \times \phi_1^{w_1}(\phi_3 \lambda\, \mathrm{e}^{-\lambda})^{m_1 - w_1} \\ &\quad \times\ \phi_2^{w_2}(\phi_3 \lambda^2\, \mathrm{e}^{-\lambda}/2)^{m_2 - w_2} \times \phi_3^{n - m_0 - m_1 - m_2} \mathrm{e}^{-(n - m_0 - m_1 - m_2)\lambda} \lambda^N \\ &= \phi_0^{w_0} \phi_1^{w_1} \phi_2^{w_2} \phi_3^{n - w_0 - w_1 - w_2} \times \mathrm{e}^{-(n - w_0 - w_1 - w_2)\lambda} \lambda^{m_1 - w_1 + 2m_2 - 2w_2 + N}, \end{aligned}$$

where $\phi_3 = 1 - \phi_0 - \phi_1 - \phi_2$. The M-step is to find the complete-data MLEs:

$$\hat{\phi}_0 = \frac{w_0}{n}, \quad \hat{\phi}_1 = \frac{w_1}{n}, \quad \hat{\phi}_2 = \frac{w_2}{n} \quad \text{and} \quad \hat{\lambda} = \frac{N + m_1 - w_1 + 2m_2 - 2w_2}{n - w_0 - w_1 - w_2}. \tag{11}$$

The E-step is to replace w_0, w_1 and w_2 in (11) by their conditional expectations:

$$\begin{aligned} E(W_0|Y_{\text{obs}}, \phi_0, \phi_1, \phi_2, \lambda) &= \frac{m_0\phi_0}{\phi_0 + (1-\phi_0-\phi_1-\phi_2)\mathrm{e}^{-\lambda}}, \\ E(W_1|Y_{\text{obs}}, \phi_0, \phi_1, \phi_2, \lambda) &= \frac{m_1\phi_1}{\phi_1 + (1-\phi_0-\phi_1-\phi_2)\lambda\,\mathrm{e}^{-\lambda}} \quad \text{and} \\ E(W_2|Y_{\text{obs}}, \phi_0, \phi_1, \phi_2, \lambda) &= \frac{m_2\phi_2}{\phi_2 + (1-\phi_0-\phi_1-\phi_2)\lambda^2\,\mathrm{e}^{-\lambda}/2}. \end{aligned} \tag{12}$$

3.3. Bootstrap confidence intervals

For small sample sizes, the bootstrap method is useful to calculate a bootstrap CI for an arbitrary function of ϕ_0, ϕ_1, ϕ_2 and λ. Let $\vartheta = h(\phi_0, \phi_1, \phi_2, \lambda)$, and $\hat{\vartheta} = h(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$ denote the MLE of ϑ calculating by either the FS algorithm (9) or the EM algorithm (11)–(12). Then using the SR (2) we can generate $Y_1^*, \ldots, Y_n^* \stackrel{\text{iid}}{\sim} \text{ZOTIP}(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$. Having obtained $Y_{\text{obs}}^* = \{y_1^*, \ldots, y_n^*\}$, we can calculate the bootstrap replications $(\hat{\phi}_0^*, \hat{\phi}_1^*, \hat{\phi}_2^*, \hat{\lambda}^*)$ and get $\hat{\vartheta}^* = h(\hat{\phi}_0^*, \hat{\phi}_1^*, \hat{\phi}_2^*, \hat{\lambda}^*)$. 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.

$$\widehat{\text{se}}(\hat{\vartheta}) = \left\{ \frac{1}{G-1} \sum_{g=1}^{G} [\hat{\vartheta}_g^* - (\hat{\vartheta}_1^* + \cdots + \hat{\vartheta}_G^*)/G]^2 \right\}^{1/2}. \tag{13}$$

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

$$[\hat{\vartheta} - z_{\alpha/2} \cdot \widehat{\text{se}}(\hat{\vartheta}), \hat{\vartheta} + z_{\alpha/2} \cdot \widehat{\text{se}}(\hat{\vartheta})]. \tag{14}$$

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

$$[\hat{\vartheta}_{\text{L}}, \; \hat{\vartheta}_{\text{U}}], \tag{15}$$

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.

4. Testing hypotheses with large sample sizes

In this section we only consider two hypotheses (i) H_0: $(\phi_0, \phi_1, \phi_2) = (0, 0, 0)$; and (ii) H_0: $\phi_2 = 0$. For (i), when H_0 is true, the traditional asymptotic property of *likelihood ratio test* (LRT) is not available since the parameter values are located at the boundary of the bounded parameter space. Therefore, we only consider the score test.

4.1. Score test for testing H_0: $(\phi_0, \phi_1, \phi_2) = (0, 0, 0)$

First, we examine whether there are excessive zeros, excessive ones and excessive twos in the ZOTIP model. The hypotheses are as follows

$$H_0\colon (\phi_0, \phi_1, \phi_2) = (0, 0, 0) \quad \text{against } H_1\colon (\phi_0, \phi_1, \phi_2) \neq (0, 0, 0). \tag{16}$$

We make the following parameter transformation:

$$\begin{aligned} \theta_0 &= \frac{\phi_0}{1 - \phi_0 - \phi_1 - \phi_2}, \quad & \theta_1 &= \frac{\phi_1}{1 - \phi_0 - \phi_1 - \phi_2}, \\ \theta_2 &= \frac{\phi_2}{1 - \phi_0 - \phi_1 - \phi_2}, \quad & \beta &= \log \lambda. \end{aligned} \tag{17}$$

Then, testing H_0 specified in (16) is equivalent to testing H_0^*: $(\theta_0, \theta_1, \theta_2) = (0, 0, 0)$. The observed-data log-likelihood function of the new parameters is

$$\begin{aligned} \ell_1 \hat{=} \ell_1(\theta_0, \theta_1, \theta_2, \beta) = \sum_{i=1}^{n} \Big\{ & -\log(1 + \theta_0 + \theta_1 + \theta_2) + [\log(\theta_0 + \mathrm{e}^{-\lambda})]I(y_i = 0) \\ & + [\log(\theta_1 + \lambda\, \mathrm{e}^{-\lambda})]I(y_i = 1) + [\log(\theta_2 + \lambda^2\, \mathrm{e}^{-\lambda}/2)]I(y_i = 2) \\ & + [y_i \log \lambda - \lambda - \log(y_i!)]I(y_i \geqslant 3) \Big\}. \end{aligned}$$

The score vector is

$$U(\theta_0, \theta_1, \theta_2, \beta) = \left(\frac{\partial \ell_1}{\partial \theta_0}, \frac{\partial \ell_1}{\partial \theta_1}, \frac{\partial \ell_1}{\partial \theta_2}, \frac{\partial \ell_1}{\partial \beta} \right)^{\top},$$

where

$$\begin{aligned} \frac{\partial \ell_1}{\partial \theta_0} &= \sum_{i=1}^{n} \left[-\frac{1}{1 + \theta_0 + \theta_1 + \theta_2} + \frac{1}{\theta_0 + \mathrm{e}^{-\lambda}} I(y_i = 0) \right], \\ \frac{\partial \ell_1}{\partial \theta_1} &= \sum_{i=1}^{n} \left[-\frac{1}{1 + \theta_0 + \theta_1 + \theta_2} + \frac{1}{\theta_1 + \lambda\, \mathrm{e}^{-\lambda}} I(y_i = 1) \right], \\ \frac{\partial \ell_1}{\partial \theta_2} &= \sum_{i=1}^{n} \left[-\frac{1}{1 + \theta_0 + \theta_1 + \theta_2} + \frac{1}{\theta_2 + \lambda^2\, \mathrm{e}^{-\lambda}/2} I(y_i = 2) \right], \\ \frac{\partial \ell_1}{\partial \beta} &= \sum_{i=1}^{n} \left[-\frac{\lambda\, \mathrm{e}^{-\lambda}}{\theta_0 + \mathrm{e}^{-\lambda}} I(y_i = 0) - \frac{(\lambda^2 - \lambda)\mathrm{e}^{-\lambda}}{\theta_1 + \lambda\, \mathrm{e}^{-\lambda}} I(y_i = 1) \right. \\ & \qquad \left. - \frac{(\lambda^3/2 - \lambda^2)\mathrm{e}^{-\lambda}}{\theta_2 + \lambda^2\, \mathrm{e}^{-\lambda}/2} I(y_i = 2) + (y_i - \lambda) I(y_i \geqslant 3) \right]. \end{aligned}$$

The second derivatives are as follows:

$$\frac{\partial^2 \ell_1}{\partial \theta_0^2} = \sum_{i=1}^{n} \left[\frac{1}{(1 + \theta_0 + \theta_1 + \theta_2)^2} - \frac{1}{(\theta_0 + \mathrm{e}^{-\lambda})^2} I(y_i = 0) \right],$$

$$\frac{\partial^2 \ell_1}{\partial \theta_1^2} = \sum_{i=1}^{n} \left[\frac{1}{(1+\theta_0+\theta_1+\theta_2)^2} - \frac{1}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})^2} I(y_i=1) \right],$$

$$\frac{\partial^2 \ell_1}{\partial \theta_2^2} = \sum_{i=1}^{n} \left[\frac{1}{(1+\theta_0+\theta_1+\theta_2)^2} - \frac{1}{(\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2)^2} I(y_i=2) \right],$$

$$\begin{aligned}\frac{\partial^2 \ell_1}{\partial \beta^2} = \sum_{i=1}^{n} \Bigg\{ & -\frac{\lambda\,\mathrm{e}^{-\lambda}[\theta_0(1-\lambda)+\mathrm{e}^{-\lambda}]}{(\theta_0+\mathrm{e}^{-\lambda})^2} I(y_i=0) \\ & - \frac{[\theta_1(3\lambda^2-\lambda-\lambda^3)\mathrm{e}^{-\lambda}+\lambda^3\mathrm{e}^{-2\lambda}]}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})^2} I(y_i=1) \\ & - \frac{[\theta_2(5\lambda^3/2-\lambda^4/2-2\lambda^2)\mathrm{e}^{-\lambda}+\lambda^5\mathrm{e}^{-2\lambda}/4]}{(\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2)^2} I(y_i=2) - \lambda I(y_i \geqslant 3) \Bigg\},\end{aligned}$$

$$\frac{\partial^2 \ell_1}{\partial \theta_0 \partial \theta_1} = \sum_{i=1}^{n} \frac{1}{(1+\theta_0+\theta_1+\theta_2)^2},$$

$$\frac{\partial^2 \ell_1}{\partial \theta_0 \partial \theta_2} = \sum_{i=1}^{n} \frac{1}{(1+\theta_0+\theta_1+\theta_2)^2},$$

$$\frac{\partial^2 \ell_1}{\partial \theta_0 \partial \beta} = \sum_{i=1}^{n} \left[\frac{\lambda\,\mathrm{e}^{-\lambda}}{(\theta_0+\mathrm{e}^{-\lambda})^2} I(y_i=0) \right],$$

$$\frac{\partial^2 \ell_1}{\partial \theta_1 \partial \theta_2} = \sum_{i=1}^{n} \frac{1}{(1+\theta_0+\theta_1+\theta_2)^2},$$

$$\frac{\partial^2 \ell_1}{\partial \theta_1 \partial \beta} = \sum_{i=1}^{n} \left[\frac{(\lambda^2-\lambda)\mathrm{e}^{-\lambda}}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})^2} I(y_i=1) \right],$$

$$\frac{\partial^2 \ell_1}{\partial \theta_2 \partial \beta} = \sum_{i=1}^{n} \left[\frac{(\lambda^3/2-\lambda^2)\mathrm{e}^{-\lambda}}{(\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2)^2} I(y_i=2) \right].$$

Since,

$$E[I(y_i=0)] = \frac{\theta_0+\mathrm{e}^{-\lambda}}{1+\theta_0+\theta_1+\theta_2}, \quad E[I(y_i=1)] = \frac{\theta_1+\lambda\,\mathrm{e}^{-\lambda}}{1+\theta_0+\theta_1+\theta_2},$$
$$E[I(y_i=2)] = \frac{\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2}{1+\theta_0+\theta_1+\theta_2}, \quad E[I(y_i \geqslant 3)] = \frac{1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}-\lambda^2\,\mathrm{e}^{-\lambda}/2}{1+\theta_0+\theta_1+\theta_2},$$

we can calculate the Fisher information matrix $\mathbf{J}(\theta_0,\theta_1,\theta_2,\beta) = (J_{ij})$, where

$$\begin{aligned} J_{11} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_0^2}\right) &= \frac{n(1+\theta_1+\theta_2-\mathrm{e}^{-\lambda})}{(1+\theta_0+\theta_1+\theta_2)^2(\theta_0+\mathrm{e}^{-\lambda})}, J_{22} \\ = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_1^2}\right) &= \frac{n(1+\theta_0+\theta_2-\lambda\,\mathrm{e}^{-\lambda})}{(1+\theta_0+\theta_1+\theta_2)^2(\theta_1+\lambda\,\mathrm{e}^{-\lambda})},\end{aligned}$$

$$J_{33} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_2^2}\right) = \frac{n(1+\theta_0+\theta_1-\lambda^2\,\mathrm{e}^{-\lambda}/2)}{(1+\theta_0+\theta_1+\theta_2)^2(\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2)},$$

$$\begin{aligned}J_{44} = -E\left(\frac{\partial^2 \ell_1}{\partial \beta^2}\right) &= \frac{n\lambda\,\mathrm{e}^{-\lambda}[\theta_0(1-\lambda)+\mathrm{e}^{-\lambda}]}{(\theta_0+\mathrm{e}^{-\lambda})(1+\theta_0+\theta_1+\theta_2)} \\ &+ \frac{n[\theta_1(3\lambda^2-\lambda-\lambda^3)\mathrm{e}^{-\lambda}+\lambda^3\mathrm{e}^{-2\lambda}]}{(\theta_1+\lambda\,\mathrm{e}^{-\lambda})(1+\theta_0+\theta_1+\theta_2)} \\ &+ \frac{n[\theta_2(5\lambda^3/2-\lambda^4/2-2\lambda^2)\mathrm{e}^{-\lambda}+\lambda^5\mathrm{e}^{-2\lambda}/4]}{(\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2)(1+\theta_0+\theta_1+\theta_2)} \\ &+ \frac{n\lambda(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}-\lambda^2\,\mathrm{e}^{-\lambda}/2)}{1+\theta_0+\theta_1+\theta_2},\end{aligned}$$

$$J_{12} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_0 \partial \theta_1}\right) = -\frac{n}{(1+\theta_0+\theta_1+\theta_2)^2},$$

$$J_{13} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_0 \partial \theta_2}\right) = -\frac{n}{(1+\theta_0+\theta_1+\theta_2)^2},$$

$$J_{14} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_0 \partial \beta}\right) = -\frac{n\lambda\,\mathrm{e}^{-\lambda}}{(1+\theta_0+\theta_1+\theta_2)(\theta_0+\mathrm{e}^{-\lambda})},$$

$$J_{23} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_1 \partial \theta_2}\right) = -\frac{n}{(1+\theta_0+\theta_1+\theta_2)^2},$$

$$J_{24} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_1 \partial \beta}\right) = -\frac{n(\lambda^2-\lambda)\mathrm{e}^{-\lambda}}{(1+\theta_0+\theta_1+\theta_2)(\theta_1+\lambda\,\mathrm{e}^{-\lambda})},$$

$$J_{34} = -E\left(\frac{\partial^2 \ell_1}{\partial \theta_2 \partial \beta}\right) = -\frac{n(\lambda^3/2-\lambda^2)\mathrm{e}^{-\lambda}}{(1+\theta_0+\theta_1+\theta_2)(\theta_2+\lambda^2\,\mathrm{e}^{-\lambda}/2)}.$$

Under H_0^*, the score test statistic is

$$T_1 = U^{\top}(0,0,0,\hat{\beta})\,\mathbf{J}^{-1}(0,0,0,\hat{\beta})\,U(0,0,0,\hat{\beta}) \mathrel{\dot\sim} \chi^2(3), \tag{18}$$

where $\hat{\beta} = \log(\bar{y})$ and

$$U(0,0,0,\hat{\beta}) = (m_0\mathrm{e}^{\bar{y}}-n,\; m_1\mathrm{e}^{\bar{y}}/\bar{y}-n,\; 2m_2\mathrm{e}^{\bar{y}}/\bar{y}^2-n,\; 0)^{\top}.$$

The p-value is

$$p_{\mathrm{v}1} = \Pr(T_1 > t_1|H_0^*) = \Pr\{\chi^2(3) > t_1\}, \tag{19}$$

where t_1 is the observed value of T_1.

4.2. Likelihood ratio test for two-inflation

Now we want to test whether there exists two-inflation in the ZOTIP model. The null and alternative hypotheses are as follows:

$$H_0\colon \phi_2 = 0 \quad \text{against } H_1\colon \phi_2 > 0. \tag{20}$$

Note that under H_0, the ZOTIP$(\phi_0,\phi_1,\phi_2;\lambda)$ distribution is reduced to the zero-one inflated Poisson distribution ZOIP(ϕ_0,ϕ_1,λ). Thus, the constrained MLEs of (ϕ_0,ϕ_1,λ)

can be calculated by the following EM iterations:

$$\begin{cases} \phi_{0,H_0}^{(t+1)} &= \dfrac{m_0\phi_{0,H_0}^{(t)}}{n\left[\phi_{0,H_0}^{(t)} + \left(1-\phi_{0,H_0}^{(t)}-\phi_{1,H_0}^{(t)}\right)\mathrm{e}^{-\lambda_{H_0}^{(t+1)}}\right]}, \\ \phi_{1,H_0}^{(t+1)} &= \dfrac{m_1\phi_{1,H_0}^{(t)}}{n\left[\phi_{1,H_0}^{(t)} + \left(1-\phi_{0,H_0}^{(t)}-\phi_{1,H_0}^{(t)}\right)\lambda\,\mathrm{e}^{-\lambda_{H_0}^{(t+1)}}\right]}, \\ \lambda_{H_0}^{(t+1)} &= \dfrac{\bar{y}}{1-\phi_{0,H_0}^{(t)}-\phi_{1,H_0}^{(t)}}, \end{cases} \tag{21}$$

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

Under H_0, the LRT statistic is given by

$$T_2 = 2\{\ell(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda}|Y_{\text{obs}}) - \ell(\hat{\phi}_{0,H_0}, \hat{\phi}_{1,H_0}, 0, \hat{\lambda}_{H_0}|Y_{\text{obs}})\}, \tag{22}$$

where $(\hat{\phi}_{0,H_0}, \hat{\phi}_{1,H_0}, \hat{\lambda}_{H_0})$ are the constrained MLEs of $(\phi_0, \phi_1, \lambda)$ under H_0, and $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$ denote the unconstrained MLEs of $(\phi_0, \phi_1, \phi_2, \lambda)$ which can be obtained by either the FS algorithm (9) or the EM algorithm (11)–(12).

Since the null hypothesis in (20) means that ϕ_2 is on the boundary of the parameter space, the appropriate null distribution is a 50:50 mixture of $\chi^2(0)$ (i.e. Degenerate(0)) and $\chi^2(1)$. The corresponding p-value is

$$p_{\text{v2}} = \Pr(T_2 > t_2|H_0) = \frac{1}{2}\Pr\{\chi^2(1) > t_2\}, \tag{23}$$

where t_2 is the observed value of T_2.

4.3. Score test for two-inflation

Now we use the score statistic to test whether or not there exist extra twos. Let $(\theta_0, \theta_1, \theta_2, \beta)$ be defined by (17), so H_0 specified in (20) is equivalent to H_0^*: $\theta_2 = 0$. Under H_0^*, the score statistic is

$$T_3 = U^{\top}(\hat{\theta}_0, \hat{\theta}_1, 0, \hat{\beta})\,\mathbf{J}^{-1}(\hat{\theta}_0, \hat{\theta}_1, 0, \hat{\beta})\,U(\hat{\theta}_0, \hat{\theta}_1, 0, \hat{\beta}) \dot{\sim} \chi^2(1), \tag{24}$$

where

$$\hat{\theta}_0 = \frac{\hat{\phi}_{0,H_0}}{1-\hat{\phi}_{0,H_0}-\hat{\phi}_{1,H_0}}, \quad \hat{\theta}_1 = \frac{\hat{\phi}_{1,H_0}}{1-\hat{\phi}_{0,H_0}-\hat{\phi}_{1,H_0}} \quad \text{and} \quad \hat{\beta} = \log(\hat{\lambda}_{H_0})$$

denote the MLEs of θ_0, θ_1 and β under H_0^*. $(\hat{\phi}_{0,H_0}, \hat{\phi}_{1,H_0}, \hat{\lambda}_{H_0})$ are determined by (21). The score vector $U(\theta_0, \theta_1, \theta_2, \beta)$ evaluated at $(\theta_0, \theta_1, \theta_2, \beta) = (\hat{\theta}_0, \hat{\theta}_1, 0, \hat{\beta})$ is given by

$$U(\hat{\theta}_0, \hat{\theta}_1, 0, \hat{\beta}) = \left(0,\ 0,\ -\frac{n}{1+\hat{\theta}_0+\hat{\theta}_1} + \frac{2m_2}{\hat{\lambda}_{H_0}^2\mathrm{e}^{-\hat{\lambda}_{H_0}}},\ 0\right)^{\top},$$

where $m_2 = \sum_{i=1}^{n} I(y_i = 2)$. The corresponding p-value is

$$p_{\text{v3}} = \Pr(T_3 > t_3|H_0^*) = \Pr(\chi^2(1) > t_3), \tag{25}$$

where t_3 is the observed value of T_3.

Table 1. Comparisons between MLEs and MSEs of parameters in ZOTIP distribution.

		Parameter				Iteration no.	
Sample size	Method	$\phi_0 = 0.25$	$\phi_1 = 0.25$	$\phi_2 = 0.25$	$\lambda = 9$	FS	EM
$n = 30$	MLE	0.2470	0.2521	0.2459	9.0277	4.89	7.00
	MSE	(0.0060)	(0.0063)	(0.0057)	(1.2463)		
$n = 50$	MLE	0.2503	0.2478	0.2511	9.0171	4.99	6.83
	MSE	(0.0038)	(0.0036)	(0.0036)	(0.7417)		
$n = 100$	MLE	0.2525	0.2494	0.2485	8.9989	5.00	6.75
	MSE	(0.0019)	(0.0020)	(0.0020)	(0.3840)		
$n = 200$	MLE	0.2496	0.2494	0.2509	9.0093	5.00	6.69
	MSE	(0.0009)	(0.0009)	(0.0010)	(0.2015)		
$n = 500$	MLE	0.2491	0.2495	0.2512	9.0104	5.00	6.70
	MSE	(0.0004)	(0.0004)	(0.0004)	(0.0783)		

Note: MLE is the average of 1000 MLEs of the parameters via FS and EM algorithms; MSE is the average of 1000 MSEs of the parameters; Iteration no. is the average iterative times.

5. Numerical experiments

In this section, we first numerically assess the performance of the proposed methods including the algorithms of FS and EM for calculating MLEs. Second, we compare the type I error rate and the power of the two tests LRT and the score test.

5.1. Simulations for likelihood-based methods

Given combinations of $(n, \phi_0, \phi_1, \phi_2, \lambda)$, we draw $\mathbf{z}_1^{(l)}, \ldots, \mathbf{z}_n^{(l)} \overset{\text{iid}}{\sim} \text{Multinomial}\,(1; \phi_0, \phi_1, \phi_2, \phi_3)$ for $l = 1, \ldots, L$, where $\mathbf{z}_i^{(l)} = (Z_{0i}^{(l)}, Z_{1i}^{(l)}, Z_{2i}^{(l)}, Z_{3i}^{(l)})^\top$, $i = 1, \ldots, n$, and generate $X_1^{(l)}, \ldots, X_n^{(l)} \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$. Finally, we set

$$Y_i^{(l)} = Z_{1i}^{(l)} + 2 \cdot Z_{2i} + Z_{3i}^{(l)} \cdot X_i^{(l)}, \quad i = 1, \ldots, n; \quad l = 1, \ldots, L, \ (L = 1000).$$

Based on the generated samples, we can obtain *mean square errors* (MSEs) and MLEs of the parameters by (9) and (11)–(12), respectively. The average iterative times of the MLEs with the same precision (10^{-6}) obtained by the two methods can also be calculated, which are defined as dividing the sum numbers of all simulation iterations by L, respectively. We set the sample size $n = 30,\ 50,\ 100,\ 200,\ 500$. The simulation results are summarized in Tables 1 and 2.

We note that the MSEs become smaller as the sample size increases. In addition, as the FS algorithm converges quadratically [26], it has a better performance in the number of iterations.

5.2. Tests for two-inflation

To compare the ZOTIP with the ZOIP distributions, we conduct some simulation studies for the null hypothesis H_0: $\phi_2 = 0$. To investigate the performance of the LRT and the score test, we compare their type I error rates and their powers of the two tests.

The sample sizes are set to be $n = 50, 100, 150, \ldots, 500$ and the values of ϕ_2 in H_1 are chosen to be $0.01, 0.03, 0.05, 0.07, 0.10, 0.15$. Other parameters are fixed as $\phi_0 = \phi_1 = 0.3, \lambda = 3$. All hypothesis testings are conducted at the significant level $\alpha = 0.05$. Let r_k

Table 2. Comparisons between MLEs and MSEs of parameters in ZOTIP distribution.

		Parameter				Iteration no.	
Sample size	Method	$\phi_0 = 0.2$	$\phi_1 = 0.1$	$\phi_2 = 0.2$	$\lambda = 5$	FS	EM
$n = 30$	MLE	0.1935	0.0889	0.1823	4.8972	5.00	15.92
	MSE	(0.0052)	(0.0031)	(0.0081)	(0.5642)		
$n = 50$	MLE	0.1999	0.0985	0.2011	5.0128	5.10	17.72
	MSE	(0.0034)	(0.0023)	(0.0047)	(0.3200)		
$n = 100$	MLE	0.1995	0.1003	0.1967	4.9870	5.00	18.24
	MSE	(0.0017)	(0.0011)	(0.0021)	(0.1538)		
$n = 200$	MLE	0.2005	0.1003	0.1994	5.0011	5.00	17.94
	MSE	(0.0008)	(0.0006)	(0.0011)	(0.0731)		
$n = 500$	MLE	0.2000	0.0998	0.1999	4.9949	5.00	17.76
	MSE	(0.0003)	(0.0002)	(0.0004)	(0.0291)		

Note: MLE is the average of 1000 MLEs of the parameters via FS and EM algorithms; MSE is the average of 1000 MSEs of the parameters; Iteration no. is the average iterative times.

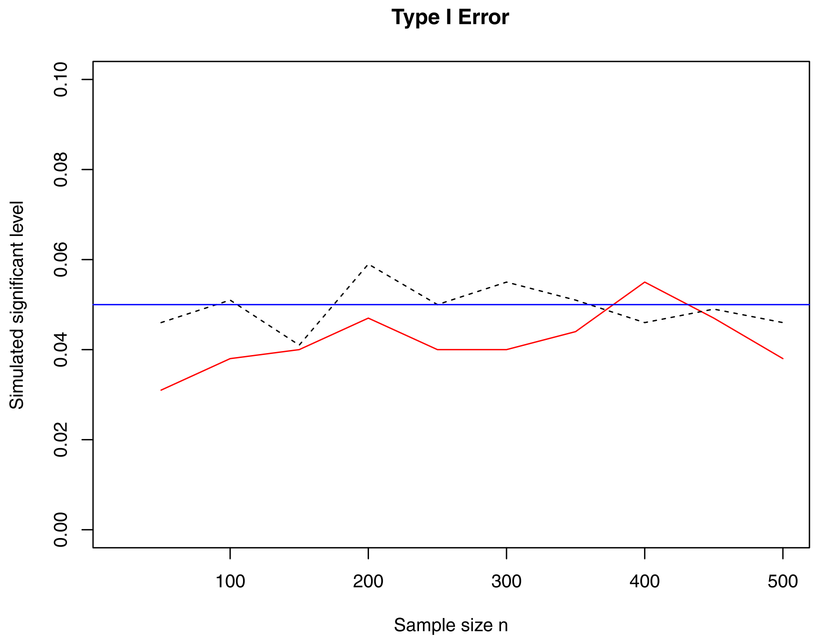

Figure 1. Comparison of type I error rates between the LRT (solid line) and the score test (dotted line). The blue line is set as the predetermined significance level of $\alpha = 0.05$.

denote the number of rejecting the null hypothesis H_0: $\phi_2 = 0$ by the test statistics T_k $(k = 2, 3)$ given by (22) and (24). In this way, the actual significance level can be estimated by r_k/L with $\phi_2 = 0$ and the power of the test statistic T_k can be estimated by r_k/L with $\phi_2 > 0$. We repeat the process of estimating each significance level and each power for L = 1000 times respectively, to obtain $\{\hat{\alpha}_k\}_{k=1}^{1000}$ and $\{1 - \hat{\beta}_k\}_{k=1}^{1000}$.

Figure 1 shows the comparison of type I error rates between the LRT and the score test for testing the two-inflation in the ZOTIP model with H_0: $\phi_2 = 0$ against H_1: $\phi_2 > 0$ for various sample sizes. We can see that both the LRT and the score test have the correct size around $\alpha = 0.05$. However, the score test has a better performance in controlling its type I error rates around the pre-chosen nominal level than the LRT.

Table 3. The LRT statistic T_2 based on 1000 replications for $\phi_0 = \phi_1 = 0.3$ and $\lambda = 3$.

		Empirical power					
Sample size (n)	Empirical level	$\phi_2 = 0.01$	$\phi_2 = 0.03$	$\phi_2 = 0.05$	$\phi_2 = 0.07$	$\phi_2 = 0.10$	$\phi_2 = 0.15$
50	0.031	0.055	0.095	0.140	0.172	0.286	0.464
100	0.038	0.065	0.134	0.212	0.321	0.452	0.744
150	0.040	0.063	0.162	0.270	0.418	0.665	0.868
200	0.047	0.063	0.188	0.338	0.521	0.751	0.955
250	0.040	0.071	0.216	0.393	0.638	0.847	0.981
300	0.040	0.083	0.236	0.501	0.686	0.911	0.990
350	0.044	0.102	0.269	0.552	0.739	0.952	0.998
400	0.055	0.103	0.320	0.582	0.808	0.965	1.000
450	0.047	0.104	0.329	0.617	0.838	0.975	1.000
500	0.038	0.109	0.373	0.659	0.886	0.979	1.000

Table 4. The score statistic T_3 based on 1000 replications for $\phi_0 = \phi_1 = 0.3$ and $\lambda = 3$.

		Empirical power					
Sample size (n)	Empirical level	$\phi_2 = 0.01$	$\phi_2 = 0.03$	$\phi_2 = 0.05$	$\phi_2 = 0.07$	$\phi_2 = 0.10$	$\phi_2 = 0.15$
50	0.046	0.051	0.062	0.083	0.109	0.210	0.406
100	0.051	0.060	0.086	0.130	0.220	0.344	0.636
150	0.041	0.047	0.098	0.175	0.307	0.558	0.784
200	0.059	0.039	0.117	0.239	0.390	0.638	0.916
250	0.050	0.053	0.141	0.298	0.521	0.773	0.960
300	0.055	0.054	0.162	0.379	0.569	0.842	0.980
350	0.051	0.068	0.154	0.432	0.619	0.906	0.993
400	0.046	0.077	0.211	0.464	0.730	0.927	0.997
450	0.049	0.057	0.229	0.502	0.766	0.946	0.999
500	0.046	0.065	0.257	0.539	0.809	0.961	1.000

Figure 2 gives the comparison of powers between the LRT and the score test for testing the two-inflation in the ZOTIP model with H_0: $\phi_2 = 0$ against H_1: $\phi_2 > 0$ for different values of $\phi_2 > 0$. It is obvious to find that the LRT is more powerful than the score test. The reason maybe that the LRT considers estimates under both the null and alternative hypotheses while the score test only considers estimates under the null hypothesis. So the score test is easier to make type II error rates than the LRT.

Tables 3 and 4 are the empirical levels/powers of the LRT statistic T_2 and the score test statistic T_3 for six scenarios: $\phi_2 = 0.01, 0.03, 0.05, 0.07, 0.10, 0.15$.

6. Applications

In this section, two real datasets are used to illustrate the proposed statistical inference methods for the ZOTIP distribution. We use the FS and EM algorithms to find the MLEs of parameters instead of the Newton–Raphson algorithm, because the corresponding observed information matrix are nearly singular in the first examples.

6.1. Stillbirths in litters of New Zealand white rabbits

The data in Table 5 give the distribution of stillbirths in 402 litters of New Zealand white rabbits, originally reported by Morgan et al. [24]. Now we will show that the proposed

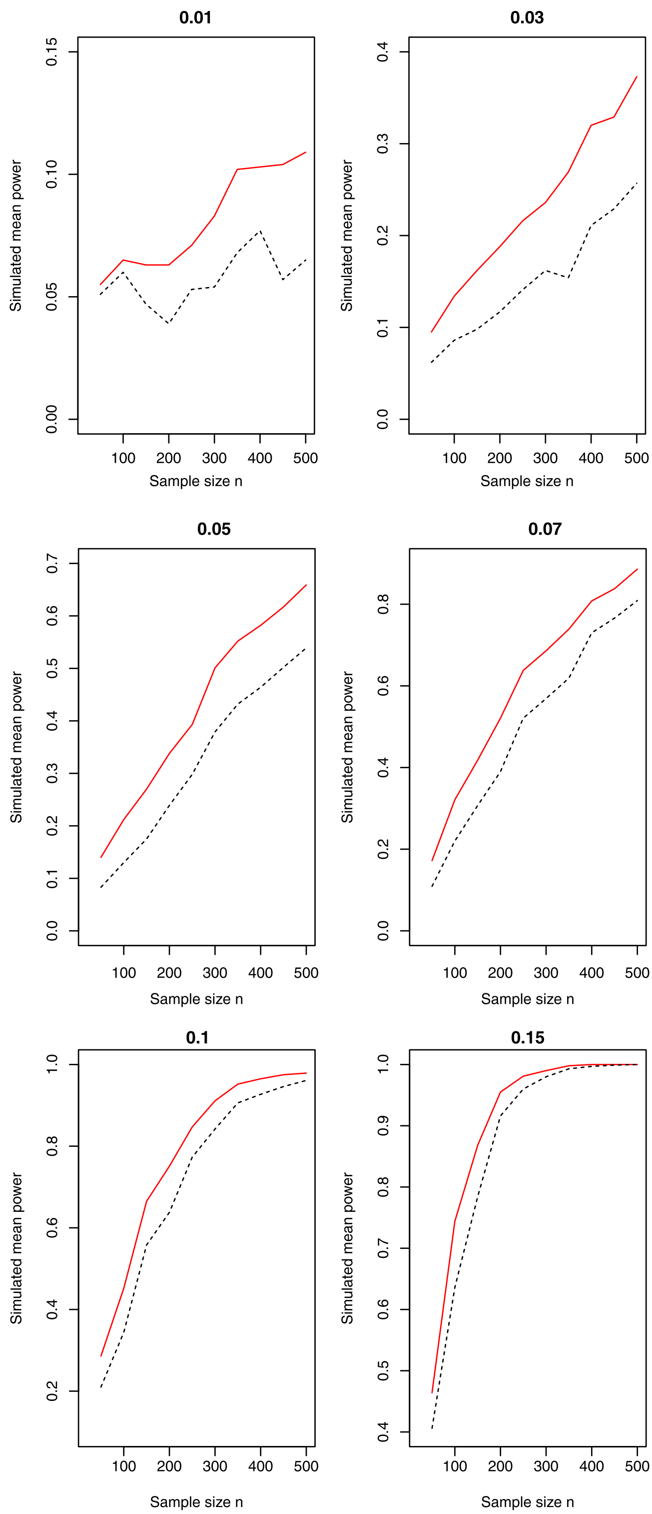

Figure 2. Comparison of powers between the LRT (solid line) and the score test (dotted line) for different values of $\phi_2 = 0.01, 0.03, 0.05, 0.07, 0.10, 0.15$.

Table 5. The stillbirths in litters of New Zealand white rabbits [24].

No. of stillbirths	0	1	2	3	4	5	6	7	8	9	10	11
No. of litters	314	48	20	7	5	2	2	1	2	0	0	1

Table 6. MLEs and CIs of parameters for the data of stillbirths in litters.

	parameter			
	ϕ_0	ϕ_1	ϕ_2	λ
MLE	0.78005843	0.11513307	0.04095272	4.1211694
std^F	0.02041905	0.01649636	0.01165951	0.5518462
95% asymptotic Wald CI	[0.740,0.820]	[0.083,0.147]	[0.018,0.064]	[3.040,5.203]
std^B	0.02654	0.02577	0.01838	1.65442
95% bootstrap CI†	[0.724,0.824]	[0.057,0.158]	[0,0.072]	[1.082,7.568]
95% bootstrap CI‡	[0.713,0.821]	[0.041,0.148]	[0,0.067]	[1.798,7.996]

Notes: std^F: Square roots of the diagonal elements of the inverse Fisher information matrix $\mathbf{J}^{-1}(\phi_0, \phi_1, \phi_2, \lambda)$. std^B: The sample standard deviation of the bootstrap samples, see (13). CI†: Normal-based bootstrap CI, see (14). CI‡: Non-normal-based bootstrap CI, see (15).

Table 7. Comparison of different Poisson distributions for the data of stillbirths in litters.

Count	Observed frequency	Poisson	ZIP	ZOIP	ZOTIP
0	314	253.73	314	314	314
1	48	116.76	32.82	48	48
2	20	26.87	28.38	12.35	**20**
3	7	4.12	16.36	11.56	4.86
4	5	0.47	7.07	8.11	5.01
5-11	8	0.04	3.37	7.97	10.13
ϕ_0			0.733884	0.77329474	0.78005843
ϕ_1				0.09750703	0.11513307
ϕ_2					0.04095272
λ		0.460199	1.7293184	2.80725198	4.1211694
Pearson's χ^2		1686.280261	21.817805	7.730145	**1.390192**
d.f.		4	3	2	1
p-value		0	0.00007118	0.0209614	**0.23837**
AIC		883.687	718.3784	695.1769	**684.1728**
BIC		887.6834	726.3713	707.1662	**700.1586**

ZOTIP distribution is more appropriate to fit this data set than other models such ZIP or ZOIP.

6.1.1. Likelihood-based inferences

Table 5 shows that there are a lot of zero stillbirths, one stillbirths and two stillbirths in the data set. Therefore, the ZOTIP distribution could be considered to model the data.

Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$. To find the MLEs of parameters, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \lambda^{(0)}) = (0.1, 0.2, 0.3, 3)$ as their initial values. To calculate the MLEs of $(\phi_0, \phi_1, \phi_2, \lambda)$, we use the FS algorithm (9) and the EM algorithm (11)–(12), which converged to the same $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$ as shown in the third row of Table 6. The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$ are given in the fourth row and 95% asymptotic Wald CIs (specified by (10)) of the four parameters are listed in the fifth row of Table 6. With $G = 6000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \phi_2, \lambda)$ are showed in the last two rows of Table 6.

Table 8. The dentist visiting data from Swedish Level of Living Surveys in 1981.

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

To examine whether there are excessive zeros, excessive ones and excessive twos in the observations, we test H_0: $(\phi_0, \phi_1, \phi_2) = (0, 0, 0)$ against H_1: $(\phi_0, \phi_1, \phi_2) \neq (0, 0, 0)$. We calculate the value of the score test statistic which is given by $t_1 = 104.1386$ according to (18), and from (19) we have $p_{v1} = 0 \ll 0.05$. Thus, the H_0 should be rejected at the significance level $\alpha = 0.05$, which means at least one of the ϕ_0, ϕ_1, ϕ_2 are positive.

To explore if the ZOIP distribution can be used to fit the data set, i.e.to examine if or not there are extra twos, we test the null hypothesis H_0: $\phi_2 = 0$ against H_1: $\phi_2 > 0$ at $\alpha = 0.05$. According to (22) and (24), we calculate the values of the LRT statistic and the score test statistic, which are given by $t_2 = 13.00406$ and $t_3 = 58.82448$, respectively. Then, from (23) and (25), we have $p_{v2} = 0.000155409 \ll \alpha = 0.05$ and $p_{v3} = 1.72085E - 14 \ll \alpha = 0.05$, resulting in a rejection of H_0 as there are excessive twos.

6.1.2. Model comparisons

We assess the goodness-of-fit by Pearson's chi-square test [27] via the predicted counts and we also use the Akaike information criterion (AIC; [28]) and Bayesian information criterion (BIC; [29]) to compare models. The comparisons of the fitted Poisson, ZIP, ZOIP and ZOTIP distributions are shown in Table 7. We found that the value of Pearson's chi-square statistic for the ZOTIP distribution is the smallest and the corresponding p-value 0.23837 is the unique p-value higher than the significance $\alpha = 0.05$. Besides, both AIC and BIC still favour the ZOTIP distribution.

6.2. Dentist visiting data in Sweden

Eriksson and Åberg [25] reported a data set that consists of the frequencies of a two-year panel observations from Swedish Level of Living Surveys in 1974 and 1981 to investigate the long-term effect of the regular dentist visits. Zhang et al. [7] only used the sample with different visit frequencies to a dentist in 1981 in Table 8 to model the distribution of visits to a dentist. To demonstrate that the ZOTIP model is more appropriate than other models such as ZIP or ZOIP, we still use the same data set as Zhang et al. [7] do.

6.2.1. Likelihood-based inferences

To find the MLEs of parameters, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \lambda^{(0)}) = (0.25, 0.25, 0.25, 3)$ as their initial values. To calculate the MLEs of $(\phi_0, \phi_1, \phi_2, \lambda)$, we employ the FS algorithm (9) and the EM algorithm (11)–(12), which converged to the same $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$ as shown in the third row of Table 9. The standard errors of the MLEs $(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2, \hat{\lambda})$ are given in the fourth row and 95% asymptotic Wald CIs (specified by (10)) of the four parameters are listed in the fifth row of Table 9. With $G = 6000$ bootstrap replications, the two 95% bootstrap CIs of $(\phi_0, \phi_1, \phi_2, \lambda)$ are showed in the last two rows of Table 9.

To examine whether there are extra zeros, extra ones and extra twos in the observations simultaneously, we consider to test H_0: $(\phi_0, \phi_1, \phi_2) = (0, 0, 0)$ against H_1: $(\phi_0, \phi_1, \phi_2) \neq (0, 0, 0)$. We calculate the value of the score test statistic which is given by $t_1 = 62.70993$

Table 9. MLEs and CIs of parameters for the dentist visiting data in Sweden.

	parameter			
	ϕ_0	ϕ_1	ϕ_2	λ
MLE	0.1721312	0.39716666	0.16550188	4.5496009
std^F	0.0137530	0.01856817	0.01565415	0.1934232
95% asymptotic Wald CI	[0.145,0.199]	[0.361,0.434]	[0.135,0.196]	[4.171,4.929]
std^B	0.01460445	0.02240059	0.02140815	0.49413620
95% bootstrap CI†	[0.143,0.200]	[0.351,0.439]	[0.122,0.205]	[3.595,5.532]
95% bootstrap CI‡	[0.143,0.201]	[0.348,0.437]	[0.118,0.202]	[3.628,5.556]

Notes: std^F: Square roots of the diagonal elements of the inverse Fisher information matrix $\mathbf{J}^{-1}(\phi_0,\phi_1,\phi_2,\lambda)$. std^B: The sample standard deviation of the bootstrap samples, see (13). CI†: Normal-based bootstrap CI, see (14). CI‡: Non-normal-based bootstrap CI, see (15).

Table 10. Comparison of different Poisson distributions for the dentist visiting data.

Count	Observed frequency	Poisson	ZIP	ZOIP	ZOTIP
0	134	110.66	134	134	134
1	314	214.1	192.69	314	314
2	149	207.11	196.55	81.88	**149**
3	69	133.57	133.66	86.2	33.71
4	32	64.6	68.17	68.05	38.34
5	26	25	27.81	42.98	34.88
6	14	8.06	9.46	22.62	26.45
7	6	2.23	2.76	10.21	17.19
8–20	22	0.68	0.9	6.06	18.42
ϕ_0			0.05162344	0.1534964	0.1721312
ϕ_1				0.3422204	0.39716666
ϕ_2					0.16550188
λ		1.934726	2.04004	3.157959	4.5496009
Pearson's χ^2		794.74226	639.1248	131.20756	54.09342
d.f.		7	6	5	4
p-value		< 0.001	< 0.001	< 0.001	< 0.001
AIC		3182.059	3175.778	2963.108	**2839.008**
BIC		3186.7	3185.061	2977.031	**2857.573**

according to (18), and from (19) we have $p_{v1} = 1.548761\mathrm{e}^{-13} \ll 0.05$. Thus, the H_0 should be rejected at the significance level of $\alpha = 0.05$, which means at least one of the ϕ_0, ϕ_1, ϕ_2 are positive.

To explore if the ZOIP model can be used to fit the data set, i.e. to examine if or not there are excessive twos, we test the null hypothesis H_0: $\phi_2 = 0$ against H_1: $\phi_2 > 0$ at $\alpha = 0.05$. According to (22) and (24), we calculate the values of the LRT statistic and the score test statistic, which are given by $t_2 = 126.0995$ and $t_3 = 540.5837$, respectively. Then, from (23) and (25), we have $p_{v2} = p_{v3} = 0 \ll \alpha = 0.05$, resulting in a rejection of H_0 as there are excessive twos.

6.2.2. *Model comparisons*

We assess the goodness-of-fit by Pearson's chi-square test via the predicted counts and we also use the AIC and BIC to compare models. The comparisons of the fitted Poisson, ZIP, ZOIP and ZOTIP distributions are shown in Table 10. We found that the value of Pearson's chi-square statistic for the ZOTIP distribution is the smallest, though all four p-values are less than the significance $\alpha = 0.05$. Besides, both AIC and BIC favour the ZOTIP distribution.

7. Discussion

In this paper, we introduce the zero-one-two inflated Poisson distribution for the first time to model count data with excess zeros, ones and twos. First, we develop some important distributional properties for the ZOTIP distribution by means of establishing different but equivalent stochastic representations. It is obvious that the first and third SRs are important to calculate explicit expressions of the moments and the moment generating function.

The MLEs of the four parameters in the ZOTIP distribution can be easily obtained by the FS algorithm or the EM algorithm. The bootstrap confidence intervals are also provided in this paper. The SR (A1) reveals the relationship between ZOIP and ZOTIP distributions, explaining that the ZOTIP is an extension of the ZOIP model.

In this paper, we only discussed the likelihood-based inference methods for one sample problem. It would be our further interest to discuss two-sample problems and multiple-sample problems by both likelihood-based and Bayesian methods. Besides, we did not consider covariates in the ZOTIP model. Therefore, the generalization of the ZOTIP distribution to the multivariate version and the consideration of covariates will be the research interest of our next stage.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors would like to thank editors and two referees for their helpful comments and valuable suggestions. Shishun Zhao's research was partially supported by a grant from the National Natural Science Foundation of China [grant number 11671168] and a grant from the Science and Technology Developing Plan of Jilin Province [grant number 20200201258JC]. Guo-Liang Tian's research was fully supported by a grant from the National Natural Science Foundation of China [grant number 11771199] and a grant from University Stability Support Plan of Shenzhen City [grant number 20200925153807002]. The work of Man-Lai Tang was partially supported through grants from the Research Grant Council of the Hong Kong Special Administrative Region [grant numbers UGC/FDS14/P01/16, UGC/FDS14/P02/18 and The Research Matching Grant Scheme (RMGS)] and a grant from the National Natural Science Foundation of China [grant number 11871124]. The computing facilities/software were supported by SAS Viya and the Big Data Intelligence Centre at the Hang Seng, University of Hong Kong.

ORCID

Yuan Sun http://orcid.org/0000-0003-2485-8290

References

[1] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34(1):1–14.
[2] Minami M, Lennert-Cody CE, Gao W, et al. Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. Fish Res. 2007;84(2):210–221.
[3] Famoye F, Singh KP. Zero-inflated generalized Poisson regression model with an application to domestic violence data. J Data Sci. 2006;4:117–130.
[4] Melkersson M, Olsson C. Is visiting the dentist a good habit? Analyzing count data with excess zeros and excess ones. Umeå: UmeåUniversity; 1999. (UmeåEconomic Studies, No. 492).

[5] Saito MY, Rodrigues J. A Bayesian analysis of zero and one inflated distributions. Rev Mat Estat. 2005;23(1):47–57.
[6] Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. J Am Stat Assoc. 1987;82(398):528–540.
[7] Zhang C, Tian GL, Ng KW. Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods. Stat Interface. 2016;9(1):11–32.
[8] Alshkaki RSA. On the zero-one inflated Poisson distribution. Int J Stat Distribut Appl. 2016;2(4):42–48.
[9] Tang YC, Liu WC, Xu AC. Statistical inference for zero-and-one-inflated Poisson models. Stat Theory Related Fields. 2017;1(2):216–226.
[10] Rahayu LP, Sadik K, Indahwati. Overdispersion study of Poisson and zero-inflated Poisson regression for some characteristics of the data on lamda, n, p. Int J Adv Intel Inform. 2016;2(3):140–148.
[11] Pittman B, Buta E, Krishnan-Sarin S, et al. Models for analyzing zero-inflated and overdispersed count data: an application to cigarette and marijuana use. Nicotine Tobacco Res Official J Soc Res Nicotine Tobacco. 2018;22(8):1390–1398.
[12] Hayati M, Kurnia A. Conwey-maxwell Poisson distribution: approach for over- and-under-dispersed count data modelling. IOP Conf Ser Earth Environ Sci. 2018;187(1): 012039.
[13] Guo ZJ, Small DS, Cheng J. Mediation analysis for count and zero-inflated count data without sequential ignorability and its application in dental studies. Stat Papers. 2018;67(2):371–394.
[14] Cheng J, Cheng NF, Guo Z, et al. Mediation analysis for count and zero-inflated count data. Stat Methods Med Res. 2018;27(9):2756–2774.
[15] Sakthivel K, Rajitha C. Model selection for count data with excess number of zero counts. Am J Appl Math Stat. 2019;7(1):43–51.
[16] Lee KH, Coull BA, Moscicki AB, et al. Bayesian variable selection for multivariate zero-inflated models: application to microbiome count data. Biostatistics. 2020;21(3):499–517.
[17] Carrivick PJW, Lee AH, Yau KKW. Zero-inflated Poisson modeling to evaluate occupational safety interventions. Saf Sci. 2003;41(1):53–63.
[18] Dieckmann A. Ein einfaches stochastisches modell zur analyse von häufigkeitsverteilungen abweichenden verhaltens. Zeitschrift für Soziologie. 1981;10(3):319–325.
[19] Greenwood M, Yule GU. 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. J R Stat Soc. 1920;83(2):255–279.
[20] Gupta PL, Gupta RC, Tripathi RC. Analysis of zero-adjusted count data. Comput Stat Data Anal. 1996;23(2):207–218.
[21] Hsu WW, Todem D, Mawella NR, et al. A robust score test of homogeneity for zero-inflated count data. Stat Methods Med Res. 2020;29(12):3653–3665.
[22] Leroux BG, Puterman ML. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. Biometrics. 1992;48:545–558.
[23] Schilling W. A frequency distribution represented as the sum of two Poisson distributions. J Am Stat Assoc. 1947;42:407–424.
[24] Morgan BJT, Palmer KJ, Ridout MS. Negative score test statistics. Am Stat. 2007;61:285–288.
[25] Eriksson R, Åberg R, editors. Welfare in transition: a survey of living conditions in Sweden 1968–1981. Oxford: Clarendon Press; 1987.
[26] Tanner MA. Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions. 3rd ed. New York: Springer; 1996.
[27] Pearson K. 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. Philos Mag Ser 5. 1900;50(302):157–175.
[28] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19(6):716–723.
[29] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–464.

Appendices

Appendix 1. The other two different but equivalent SRs

A.1 The second stochastic representation

To derive the second SR of $Y \sim \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$, we first introduce the *zero-and-one-inflated Poisson* (ZOIP) distribution. A discrete random variable $W \sim \text{ZOIP}(\phi_0^*, \phi_1^*; \lambda)$ if its pmf is [4]

$$\Pr(W = w) = (\phi_0^* + \phi_2^* \mathrm{e}^{-\lambda}) I(w = 0) + (\phi_1^* + \phi_2^* \lambda\, \mathrm{e}^{-\lambda}) I(w = 1) \\ + (\phi_2^* \lambda^w \mathrm{e}^{-\lambda}/w!) I(w \geqslant 2),$$

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 $\phi_2^* \hat{=} 1 - \phi_0^* - \phi_1^* \in (0, 1]$.

Let $Z \sim \text{Bernoulli}(1 - \phi)$, $W \sim \text{ZOIP}(\phi_0^*, \phi_1^*; \lambda)$, and $Z \perp\!\!\!\perp W$. Then

$$Y \stackrel{\mathrm{d}}{=} (1 - Z) \cdot 2 + ZW = \begin{cases} 2, & \text{with probability } \phi, \\ W, & \text{with probability } 1 - \phi \end{cases} \tag{A1}$$

follows the distribution $\text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ with $\phi_0 = (1 - \phi)\phi_0^*$, $\phi_1 = (1 - \phi)\phi_1^*$ and $\phi_2 = \phi$. This fact can be shown as follows:

$$\begin{cases} \Pr(Y = 0) & = \Pr(Z = 1, W = 0) = (1 - \phi)(\phi_0^* + \phi_2^* \mathrm{e}^{-\lambda}), \\ \Pr(Y = 1) & = \Pr(Z = 1, W = 1) = (1 - \phi)(\phi_1^* + \phi_2^* \lambda\, \mathrm{e}^{-\lambda}), \\ \Pr(Y = 2) & = \Pr(Z = 0) + \Pr(Z = 1, W = 2) = \phi + (1 - \phi)\phi_2^* \lambda^2 \mathrm{e}^{-\lambda}/2, \\ \Pr(Y = y) & = \Pr(Z = 1, W = y) = (1 - \phi)\phi_2^* \lambda^y \mathrm{e}^{-\lambda}/y!, \quad y \geqslant 3. \end{cases} \tag{A2}$$

By comparing (A2) with (3), we obtain a one-to-one mapping between (ϕ_0, ϕ_1, ϕ_2) and $(\phi_0^*, \phi_1^*, \phi)$:

$$\begin{cases} (1 - \phi)\phi_0^* = \phi_0, \\ (1 - \phi)\phi_1^* = \phi_1, \\ \phi = \phi_2, \\ (1 - \phi)\phi_2^* = \phi_3, \end{cases} \iff \begin{cases} \phi = \phi_2, \\ \phi_0^* = \dfrac{\phi_0}{1 - \phi_2}, \\ \phi_1^* = \dfrac{\phi_1}{1 - \phi_2}. \end{cases}$$

The SR (A1) indicates that $Y \sim \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ is a mixture of the Degenerate (2) distribution and the $\text{ZOIP}(\phi_0^*, \phi_1^*; \lambda)$ distribution.

A.2 The third stochastic representation

A discrete random variable V is said to have the *three points* (TP) distribution, denoted by $V \sim \text{TP}(p_0, p_1)$, if its pmf is defined as

$$\Pr(V = k) = \begin{cases} p_0, & \text{if } k = 0, \\ p_1, & \text{if } k = 1, \\ 1 - p_0 - p_1, & \text{if } k = 2. \end{cases}$$

Let $Z \sim \text{Bernoulli}(1 - \phi)$, $V \sim \text{TP}(p_0, p_1)$, $X \sim \text{Poisson}(\lambda)$ and (Z, V, X) be mutually independent. Then

$$Y \stackrel{\mathrm{d}}{=} (1 - Z)V + ZX = \begin{cases} V, & \text{with probability } \phi, \\ X, & \text{with probability } 1 - \phi \end{cases} \tag{A3}$$

follows the distribution ZOTIP$(\phi_0, \phi_1, \phi_2; \lambda)$ with $\phi_0 = \phi p_0$, $\phi_1 = \phi p_1$ and $\phi_2 = \phi(1 - p_0 - p_1)$. This fact can be verified as follows:

$$\begin{cases} \Pr(Y=0) & = \Pr(Z=0, V=0) + \Pr(Z=1, X=0) \\ & = \phi p_0 + (1-\phi)\mathrm{e}^{-\lambda}, \\ \Pr(Y=1) & = \Pr(Z=0, V=1) + \Pr(Z=1, X=1) \\ & = \phi p_1 + (1-\phi)\lambda\,\mathrm{e}^{-\lambda}, \\ \Pr(Y=2) & = \Pr(Z=0, V=2) + \Pr(Z=1, X=2) \\ & = \phi(1-p_0-p_1) + (1-\phi)\lambda^2\mathrm{e}^{-\lambda}/2, \\ \Pr(Y=y) & = \Pr(Z=1, X=y) = (1-\phi)\lambda^y\mathrm{e}^{-\lambda}/y!, \quad y \geqslant 3. \end{cases} \tag{A4}$$

By comparing (A4) with (3), we obtain a one-to-one mapping between (ϕ, p_0, p_1) and (ϕ_0, ϕ_1, ϕ_2):

$$\begin{cases} \phi p_0 = \phi_0, \\ \phi p_1 = \phi_1, \\ \phi(1-p_0-p_1) = \phi_2, \\ 1-\phi = \phi_3, \end{cases} \iff \begin{cases} \phi = \phi_0 + \phi_1 + \phi_2, \\ p_0 = \dfrac{\phi_0}{\phi_0+\phi_1+\phi_2}, \\ p_1 = \dfrac{\phi_1}{\phi_0+\phi_1+\phi_2}. \end{cases}$$

The SR (A3) means that $Y \sim \text{ZOTIP}(\phi_0, \phi_1, \phi_2; \lambda)$ is also a mixture of the TP$(\phi_0/(\phi_0 + \phi_1 + \phi_2), \phi_1/(\phi_0 + \phi_1 + \phi_2))$ distribution and the Poisson(λ) distribution.

Appendix 2. Proofs of Propositions 2.1–3.1

Proof of Proposition 2.1: The joint conditional pmf of $\mathbf{z}|(Y = y)$ is given by

$$\Pr(\mathbf{z} = z|Y = y) = \frac{\Pr(Z_0 = z_0, Z_1 = z_1, Z_2 = z_2, Z_3 = z_3, Y = y)}{\Pr(Y = y)}.$$

If $y = 0$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0,0)^\top \mid Y=0\} \overset{(1)}{=} \dfrac{\phi_0}{\phi_0 + \phi_3\mathrm{e}^{-\lambda}} \overset{(6)}{=} \psi_1, \\ \Pr\{\mathbf{z} = (0,1,0,0)^\top \mid Y=0\} = 0, \\ \Pr\{\mathbf{z} = (0,0,1,0)^\top \mid Y=0\} = 0, \\ \Pr\{\mathbf{z} = (0,0,0,1)^\top \mid Y=0\} = \dfrac{\phi_3\mathrm{e}^{-\lambda}}{\phi_0 + \phi_3\mathrm{e}^{-\lambda}} = 1 - \psi_1, \end{cases}$$

which imply the first assertion of (5).

If $y = 1$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0,0)^\top \mid Y=1\} = 0, \\ \Pr\{\mathbf{z} = (0,1,0,0)^\top \mid Y=1\} \overset{(1)}{=} \dfrac{\phi_1}{\phi_1 + \phi_3\lambda\,\mathrm{e}^{-\lambda}} \overset{(6)}{=} \psi_2, \\ \Pr\{\mathbf{z} = (0,0,1,0)^\top \mid Y=1\} = 0, \\ \Pr\{\mathbf{z} = (0,0,0,1)^\top \mid Y=1\} = \dfrac{\phi_3\lambda\,\mathrm{e}^{-\lambda}}{\phi_1 + \phi_3\lambda\,\mathrm{e}^{-\lambda}} = 1 - \psi_2, \end{cases}$$

which imply the second assertion of (5).

If $y = 2$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0,0)^\top \mid Y = 2\} = 0, \\ \Pr\{\mathbf{z} = (0,1,0,0)^\top \mid Y = 2\} = 0, \\ \Pr\{\mathbf{z} = (0,0,1,0)^\top \mid Y = 2\} \stackrel{(1)}{=} \dfrac{\phi_2}{\phi_2 + \phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} \stackrel{(6)}{=} \psi_3, \\ \Pr\{\mathbf{z} = (0,0,0,1)^\top \mid Y = 2\} = \dfrac{\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2}{\phi_2 + \phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} = 1 - \psi_3, \end{cases}$$

which imply the third assertion of (5).

If $y \geqslant 3$, then

$$\begin{cases} \Pr\{\mathbf{z} = (1,0,0,0)^\top \mid Y = y\} = 0, \\ \Pr\{\mathbf{z} = (0,1,0,0)^\top \mid Y = y\} = 0, \\ \Pr\{\mathbf{z} = (0,0,1,0)^\top \mid Y = y\} = 0, \\ \Pr\{\mathbf{z} = (0,0,0,1)^\top \mid Y = y\} \stackrel{(1)}{=} \dfrac{\phi_3\lambda^y\mathrm{e}^{-\lambda}/y!}{\phi_3\lambda^y\mathrm{e}^{-\lambda}/y!} = 1, \end{cases}$$

implying the last assertion of (5). ■

Proof of Proposition 2.2: If $y = 0$, we have

$$\begin{aligned} \Pr(X = x|Y = 0) &= \frac{\Pr(X = x, Y = 0)}{\Pr(Y = 0)} \\ &= \frac{\Pr(X = 0, Z_1 = Z_2 = 0)}{f(0|\phi_0,\phi_1,\phi_2;\lambda)} I(x = 0) + \frac{\Pr(X = x, Z_0 = 1)}{f(0|\phi_0,\phi_1,\phi_2;\lambda)} I(x \neq 0) \\ &\stackrel{(1)}{=} \frac{(1 - \phi_1 - \phi_2)\mathrm{e}^{-\lambda}}{\phi_0 + \phi_3\mathrm{e}^{-\lambda}} I(x = 0) + \frac{\phi_0\lambda^x\mathrm{e}^{-\lambda}/x!}{\phi_0 + \phi_3\mathrm{e}^{-\lambda}} I(x \neq 0) \\ &\stackrel{(6)}{=} (1 - \psi_1 + \psi_1\mathrm{e}^{-\lambda}) I(x = 0) + (\psi_1\lambda^x\mathrm{e}^{-\lambda}/x!) I(x \neq 0), \end{aligned}$$

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

If $y = 1$, we have

$$\begin{aligned} \Pr(X = x|Y = 1) &= \frac{\Pr(X = x, Y = 1)}{\Pr(Y = 1)} \\ &= \frac{\Pr(X = 1, Z_0 = Z_2 = 0)}{f(1|\phi_0,\phi_1,\phi_2;\lambda)} I(x = 1) + \frac{\Pr(X = x, Z_1 = 1)}{f(1|\phi_0,\phi_1,\phi_2;\lambda)} I(x \neq 1) \\ &\stackrel{(1)}{=} \frac{(1 - \phi_0 - \phi_2)\lambda\,\mathrm{e}^{-\lambda}}{\phi_1 + \phi_3\lambda\,\mathrm{e}^{-\lambda}} I(x = 1) + \frac{\phi_1\lambda^x\mathrm{e}^{-\lambda}/x!}{\phi_1 + \phi_3\lambda\,\mathrm{e}^{-\lambda}} I(x \neq 1) \\ &\stackrel{(6)}{=} (1 - \psi_2 + \psi_2\lambda\,\mathrm{e}^{-\lambda}) I(x = 1) + (\psi_2\lambda^x\mathrm{e}^{-\lambda}/x!) I(x \neq 1), \end{aligned}$$

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

If $y = 2$, we have

$$\begin{aligned} \Pr(X = x|Y = 2) &= \frac{\Pr(X = x, Y = 2)}{\Pr(Y = 2)} \\ &= \frac{\Pr(X = 2, Z_0 = Z_1 = 0)}{f(2|\phi_0,\phi_1,\phi_2;\lambda)} I(x = 2) + \frac{\Pr(X = x, Z_2 = 1)}{f(2|\phi_0,\phi_1,\phi_2;\lambda)} I(x \neq 2) \\ &\stackrel{(1)}{=} \frac{(1 - \phi_0 - \phi_1)\lambda^2\,\mathrm{e}^{-\lambda}/2}{\phi_2 + \phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} I(x = 2) + \frac{\phi_2\lambda^x\mathrm{e}^{-\lambda}/x!}{\phi_2 + \phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2} I(x \neq 2) \end{aligned}$$

$$\overset{(6)}{=} (1-\psi_3+\psi_3\lambda\,\mathrm{e}^{-\lambda})I(x=2)+(\psi_3\lambda^x\mathrm{e}^{-\lambda}/x!)I(x\neq 2),$$

implying $X|(Y=2) \sim \mathrm{TIP}(1-\psi_3,\lambda)$.

If $y \geqslant 3$, we have

$$\Pr(X=x|Y=y)=\frac{\Pr(X=x,Y=y)}{f(y|\phi_0,\phi_1,\phi_2;\lambda)} \overset{(1)}{=} \frac{\Pr(X=y,Z_3=1)}{\phi_3\lambda^y\mathrm{e}^{-\lambda}/y!}=1,$$

implying $X|(Y=y\geqslant 3) \sim \mathrm{Degenerate}(y)$. ■

Proof of Proposition 3.1: It is easy to know the expressions of m_0, m_1 and m_2, so we focus on the last one and give two ways to prove it. The first way is as follows. Since

$$E(Y_i|Y_i\geqslant 3)=\frac{\lambda(1-\mathrm{e}^{-\lambda}-\lambda\mathrm{e}^{-\lambda})}{1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}-\lambda^2\,\mathrm{e}^{-\lambda}/2},\quad i\notin \mathbb{I}_0\cup\mathbb{I}_1\cup\mathbb{I}_2,$$

we have

$$\begin{aligned}
E(N) &= E[E(N|m_0,m_1,m_2)] \\
&= E\left[\frac{(n-m_0-m_1-m_2)\lambda(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda})}{1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}-\lambda^2\,\mathrm{e}^{-\lambda}/2}\right] \\
&= \frac{[n-E(m_0)-E(m_1)-E(m_2)]\,\lambda\,(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda})}{1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}-\lambda^2\,\mathrm{e}^{-\lambda}/2} \\
&\overset{(8)}{=} n\phi_3\lambda(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}).
\end{aligned}$$

In the second way, it is obvious that $N=\sum_{i\notin\mathbb{I}_0\cup\mathbb{I}_1\cup\mathbb{I}_2} y_i=\sum_{i=1}^n y_i-m_1-2m_2$. Thus, we obtain:

$$\begin{aligned}
E(N) &= nE(Y_1)-E(m_1)-2E(m_2) \\
&\overset{(4)}{=} n(\phi_1+2\phi_2+\phi_3\lambda)-n(\phi_1+\phi_3\lambda\,\mathrm{e}^{-\lambda})-2n(\phi_2+\phi_3\lambda^2\,\mathrm{e}^{-\lambda}/2) \\
&= n\phi_3\lambda(1-\mathrm{e}^{-\lambda}-\lambda\,\mathrm{e}^{-\lambda}),
\end{aligned}$$

which implies (8). ■