Multivariate zero-and-one inflated Poisson model with applications

Chi ZHANG[a,*], **Guo-Liang TIAN**[b], **Kam Chuen YUEN**[c], **Qin WU**[d] and **Tao LI**[e]

[a] *College of Economics, Shenzhen University, Nanhai Ave 3688, Shenzhen City, Guangdong 518060, P. R. China*

[b] *Department of Statistics and Data Science, Southern University of Science and Technology, No. 1088, Xueyuan Road, Shenzhen City, Guangdong 518055, P. R. China*

[c] *Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, P. R. China*

[d] *Department of Statistics, School of Mathematical Sciences, South China Normal University, Guangzhou City, Guangdong 510631, P. R. China*

[e] *Department of Mathematics, Southern University of Science and Technology, No. 1088, Xueyuan Road, Shenzhen City, Guangdong 518055, P. R. China*

*Corresponding author's email: loezc727@163.com

Abstract: This paper extends the univariate *zero-and-one inflated Poisson* (ZOIP) distribution (Melkersson & Olsson, 1999; Zhang *et al.*, 2016) to its multivariate version, which can be used to model correlated multivariate count data with large proportions of zeros and ones marginally. More importantly, this new multivariate ZOIP distribution possesses a flexible dependency structure; i.e., the correlation coefficient between any two random components could be either positive or negative depending on the values of the parameters. The important distributional properties are explored and some useful statistical inference methods without and with covariates are developed. Simulation studies are conducted to evaluate the performance of the proposed methods. Finally, two real data sets on healthcare and insurance are used to illustrate the proposed methods.

Keywords: Expectation–maximization (EM) algorithm; Multivariate zero-and-one inflated Poisson; Univariate zero-and-one inflated Poisson; Zero-inflated Poisson.

1. Introduction

Count data on the number of sex partners among the young within a fixed period, on motor vehicle crashes among young drivers in one year and on occupational safety involving accidents or injuries in a half year exhibit the characteristics of excessive zero and excessive one observations (Lee *et al.*, 2002; Carrivick *et al.*, 2003). The traditional Poisson and *zero-inflated Poisson* (ZIP; Lambert, 1992) are no longer appropriate distributions to model such count data. Motivated by the data set on Swedish visits to a dentist with higher proportions of zeros and ones and one-visit observations being even much more frequent than zero-visits, Melkersson & Olsson (1999) proposed a so-called *zero-and-one inflated Poisson* (ZOIP) distribution as a generalization of the univariate ZIP to seize the feature of such count data. Their main objective is to fit the dentist visiting data with covariates in Sweden. Later, Saito & Rodrigues (2005) presented a Bayesian analysis of the same dentist visiting data without considering covariates by the data augmentation algorithm. Recently, Zhang *et al.* (2016) defined the univariate ZOIP distribution, denoted by $Y \sim \text{ZOIP}(\phi_0, \phi_1; \lambda)$, via the following *stochastic representation* (SR):

$$Y = Z_0 \cdot 0 + Z_1 \cdot 1 + Z_2 X = Z_1 + Z_2 X = \begin{cases} 0, & \text{with probability } \phi_0, \\ 1, & \text{with probability } \phi_1, \\ X, & \text{with probability } \phi_2, \end{cases} \tag{1.1}$$

where $\mathbf{z} = (Z_0, Z_1, Z_2)^\top \sim \text{Multinomial}\,(1; \phi_0, \phi_1, \phi_2)$, $X \sim \text{Poisson}\,(\lambda)$, and $\mathbf{z}$, X are independent (symbolized as $\mathbf{z} \perp\!\!\!\perp X$). The corresponding *probability mass function* (pmf) is

$$f(y|\phi_0, \phi_1; \lambda) = (\phi_0 + \phi_2 \mathrm{e}^{-\lambda}) I(y = 0) + (\phi_1 + \phi_2 \lambda\, \mathrm{e}^{-\lambda}) I(y = 1) + \left(\phi_2 \frac{\lambda^y \mathrm{e}^{-\lambda}}{y!}\right) I(y \geqslant 2),$$

where $I(\cdot)$ denotes the indicator function. Liu, Tang & Xu (2018) further discussed the Bayesian estimation of the ZOIP model. Tang *et al.* (2017) compared the maximum likelihood estimation with the Bayesian estimation for the ZOIP model parameters.

Extra zeros and extra ones in multivariate count data also appear frequently in practice. For instance, there exist different types of defects in manufacturing process, there are various types of injuries in accident events and so on. Sometimes, both types of defects or injuries rarely occur (in other words, there are too many $(0,0)^\top$ observations) due to excellent safety precautions; sometimes, one type of defect/injury often occurs once (i.e., there are extra $(1,0)^\top$ and/or $(0,1)^\top$ observations) because some specific defects/injuries are prone to happen and hard to be prevented; sometimes, both types of defects/injuries could simultaneously occur (namely, there are excessive $(1,1)^\top$ observations) if they are intrinsically correlated. This kind of multivariate count data share a common characteristic; i.e., each component marginally follows a univariate ZOIP distribution.

To model multivariate correlated count data, the multivariate Poisson distribution was constructed (e.g., Johnson *et al.*, 1997, p.139) by adding a common Poisson variable $X_0^* \sim \text{Poisson}(\lambda_0)$ to each $X_i^* \sim \text{Poisson}(\lambda_i)$ to form the new components $X_i = X_0^* + X_i^*$, where the correlations among $\{X_i\}$ come from the common X_0^* while the i-th component X_i is still a Poisson variable. Li *et al.* (1999) proposed a multivariate ZIP distribution to model manufacturing data with extra zeros while each marginal is a univariate ZIP distribution. Liu & Tian (2015) used the stochastic representation to construct a multivariate ZIP distribution and Tian *et al.* (2018) extended it to a multivariate *zero-adjusted Poisson* (ZAP) distribution. Later, Liu *et al.* (2018) proposed a more flexible multivariate ZAP model for multivariate count data analyses. Diallo *et al.* (2018) proposed a zero-inflated regression model for multinomial counts with joint zero-inflation.

In addition, considerable work has been concentrated on the bivariate case. For example, Walhin (2001) proposed three new bivariate ZIP models and used two real data sets to illustrate the proposed methods. Wang *et al.* (2003) applied a bivariate ZIP regression model with covariates to analyze occupational injuries data. Karlis & Ntzoufras (2005) extended the bivariate Poisson distribution by incorporating the diagonal inflation into the model to fit data with higher probabilities in diagonal elements. Deshmukh & Kasture (2002) even investigated the bivariate distribution problem with truncated Poisson marginal distributions. These researches are not available when the dimension is larger than or equal to 3. More importantly, these models can only produce zero-inflated or zero-truncated Poisson marginal distribution (except for Karlis & Ntzoufras, 2005); in other words, all above models cannot capture the characteristic of ZOIP marginal distributions. Therefore, the major objective of this article is to propose a multivariate Poisson distribution with ZOIP margins by developing its important distributional properties and the useful statistical inference methods without and with covariates. It is expected that this new multivariate model can provide a better fit especially for those correlated count data with large proportions of zeros and ones marginally.

The rest of the paper is organized as follows. In Section 2, the multivariate ZOIP distribution constructed by SR is proposed and its joint pmf is derived. In Section 3, the likelihood-based methods are developed for the general and related reduced models, including the maximum likelihood estimation, bootstrap confidence interval construction, hypothesis testing and a regression model analysis. In Sections 4, Bayesian methods are further considered. Simulations are conducted to evaluate the performance of the proposed methods in Section 5. Two real examples are used to illustrate the proposed methods in Section 6. A discussion is presented in Section 7 and some technical details are put into the Appendix.

2. Multivariate ZOIP distribution

Let $\{X_i^*\}_{i=0}^m \stackrel{\text{ind}}{\sim} \text{Poisson}\,(\lambda_i)$ and $X_i = X_0^* + X_i^*$, $i = 1, \dots, m$. Then, the discrete random vector $\mathbf{x} = (X_1, \dots, X_m)^\top$ is said to follow an m-dimensional Poisson distribution with parameters $\lambda_0 \geqslant 0$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{x} \sim \text{MP}\,(\lambda_0, \lambda_1, \dots, \lambda_m)$ or $\mathbf{x} \sim \text{MP}_m(\lambda_0, \boldsymbol{\lambda})$, accordingly. The joint pmf of $\mathbf{x}$ is

$$\Pr(\mathbf{x} = \boldsymbol{x}) = \mathrm{e}^{-\lambda_0 - \lambda_+} \sum_{k=0}^{\min(\boldsymbol{x})} \frac{\lambda_0^k}{k!} \prod_{i=1}^m \frac{\lambda_i^{x_i - k}}{(x_i - k)!}, \tag{2.1}$$

where $\boldsymbol{x} = (x_1, \dots, x_m)^\top$, $\{x_i\}_{i=1}^m$ are the corresponding realizations of $\{X_i\}_{i=1}^m$, $\lambda_+ \hat{=} \sum_{i=1}^m \lambda_i$, and $\min(\boldsymbol{x}) \hat{=} \min(x_1, \dots, x_m)$.

Let a discrete random vector $\mathbf{y} = (Y_1, \dots, Y_m)^\top$ have the following mixture distribution:

$$\begin{aligned}
\mathbf{y} \;&\sim\; (\underbrace{0, 0, \dots, 0}_{m})^\top \,\hat{=}\, \mathbf{0}_m && \text{with probability } \phi_0, \\
&\sim\; (\underbrace{0, \dots, 0}_{i-1}, 1, \underbrace{0, \dots, 0}_{m-i})^\top \,\hat{=}\, \boldsymbol{e}_m^{(i)} && \text{with probability } \phi_i, \quad 1 \leqslant i \leqslant m, \\
&\sim\; (\underbrace{1, 1, \dots, 1}_{m})^\top \,\hat{=}\, \mathbf{1}_m && \text{with probability } \phi_{m+1}, \\
&\sim\; \text{MP}_m(\lambda_0, \boldsymbol{\lambda}) && \text{with probability } \phi_{m+2},
\end{aligned}$$

where $\sum_{k=0}^{m+2} \phi_k = 1$. The joint pmf of $\mathbf{y}$ is given by (2.3), which is very complicated. To extensively explore the distributional properties and develop efficient statistical methods such as the *expectation–maximization* (EM) algorithm and the *data augmentation* (DA) algorithm, we then employ the tractable SR rather than the intractable joint pmf to define the above mixture distribution.

Definition 1 A discrete random vector $\mathbf{y} = (Y_1, \ldots, Y_m)^\top$ is said to have a multivariate ZOIP distribution with parameters $(\phi_0, \boldsymbol{\phi}, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda})$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_m)^\top$, $\phi_k \in [0, 1)$ for $k = 0, 1, \ldots, m+1$, $\phi_{m+2} \hat{=} 1 - \sum_{k=0}^{m+1} \phi_k \in (0, 1]$ and $\lambda_0 \geqslant 0$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{y} \sim \text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$, if $\mathbf{y}$ has the following SR:

$$
\begin{aligned}
\mathbf{y} &= \begin{cases} \mathbf{0}_m, & \text{with probability } \phi_0, \\ \boldsymbol{e}_m^{(i)}, & \text{with probability } \phi_i, \quad 1 \leqslant i \leqslant m, \\ \mathbf{1}_m, & \text{with probability } \phi_{m+1}, \\ \mathbf{x}, & \text{with probability } \phi_{m+2}, \end{cases} \\
&\stackrel{\text{d}}{=} Z_0 \boldsymbol{\xi}_0 + \sum_{i=1}^{m} Z_i \boldsymbol{\xi}^{(i)} + Z_{m+1} \boldsymbol{\xi}_1 + Z_{m+2} \mathbf{x} \\
&= (Z_1, \ldots, Z_m)^\top + Z_{m+1} \boldsymbol{\xi}_1 + Z_{m+2} \mathbf{x}, \qquad (2.2)
\end{aligned}
$$

where $\mathbf{z} = (Z_0, Z_1, \ldots, Z_{m+2})^\top \sim \text{Multinomial}\,(1; \phi_0, \boldsymbol{\phi}, \phi_{m+1}, \phi_{m+2})$, $\boldsymbol{\xi}_0 \sim \text{Degenerate}\,(\mathbf{0}_m)$, $\boldsymbol{\xi}^{(i)} \sim \text{Degenerate}\,(\boldsymbol{e}_m^{(i)})$ for $i = 1, \ldots, m$, $\boldsymbol{\xi}_1 \sim \text{Degenerate}\,(\mathbf{1}_m)$, $\mathbf{x} = (X_1, \ldots, X_m)^\top \sim \text{MP}_m(\lambda_0, \boldsymbol{\lambda})$ and $\mathbf{z} \perp\!\!\!\perp \mathbf{x}$. In particular, when $\phi_0 = 0$, it reduces to the one-inflated Poisson distribution. ¶

We discuss several special cases of (2.2):

(1) If $\phi_i = 0\,(i = 1, \ldots, m+1)$, then $\mathbf{y}$ has the Type II multivariate ZIP distribution, denoted by $\mathbf{y} \sim \text{ZIP}_m^{(\text{II})}(\phi_0; \lambda_0, \boldsymbol{\lambda})$, see Appendix C.

(2) If $\phi_i = 0\,(i = 1, \ldots, m+1)$ and $\lambda_0 = 0$, then $\mathbf{y}$ has the Type I multivariate ZIP distribution (Liu & Tian, 2015), denoted by $\mathbf{y} \sim \text{ZIP}_m^{(\text{I})}(\phi_0; \boldsymbol{\lambda})$, see Appendix B.

(3) If $\phi_i = 0\,(i = 0, 1, \ldots, m+1)$, then $\mathbf{y}$ follows the multivariate Poisson distribution; that is $\mathbf{y} \sim \mathrm{MP}_m(\lambda_0, \boldsymbol{\lambda})$.

(4) If $\phi_i = 0\,(i = 0, 1, \ldots, m+1)$ and $\lambda_0 = 0$, then $Y_i \overset{\text{ind}}{\sim} \text{Poisson}\,(\lambda_i)$ for $i = 1, \ldots, m$.

2.1 Joint probability mass function and mixed moments

In Appendix A.1, we show that the joint pmf of $\mathbf{y} \sim \mathrm{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$ is given by

$$
\begin{aligned}
& f(\boldsymbol{y}|\phi_0, \boldsymbol{\phi}, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda}) = \Pr(Y_1 = y_1, \ldots, Y_m = y_m) \\
= \; & a_0 I(\boldsymbol{y} = \mathbf{0}_m) + \sum_{i=1}^{m} a_i I(\boldsymbol{y} = \boldsymbol{e}_m^{(i)}) + a_{m+1} I(\boldsymbol{y} = \mathbf{1}_m) \\
& + \left[\phi_{m+2} \mathrm{e}^{-\lambda_0 - \lambda_+} \sum_{k=0}^{\min(\boldsymbol{y})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{y_i - k}}{(y_i - k)!}\right] I(\boldsymbol{y} \notin \mathcal{Y}_{01}) \qquad (2.3) \\
= \; & \phi_0 \Pr(\boldsymbol{\xi}_0 = \boldsymbol{y}) + \sum_{i=1}^{m} \phi_i \Pr(\boldsymbol{\xi}^{(i)} = \boldsymbol{y}) + \phi_{m+1} \Pr(\boldsymbol{\xi}_1 = \boldsymbol{y}) + \phi_{m+2} \Pr(\mathbf{x} = \boldsymbol{y}),
\end{aligned}
$$

where

$$
\begin{aligned}
a_0 &= \phi_0 + \phi_{m+2} \mathrm{e}^{-\lambda_0 - \lambda_+}, \quad a_i = \phi_i + \phi_{m+2} \lambda_i \mathrm{e}^{-\lambda_0 - \lambda_+}, \quad 1 \leqslant i \leqslant m, \\
a_{m+1} &= \phi_{m+1} + \phi_{m+2} \left(\lambda_0 + \textstyle\prod_{i=1}^{m} \lambda_i\right) \mathrm{e}^{-\lambda_0 - \lambda_+},
\end{aligned} \qquad (2.4)
$$

and $\mathcal{Y}_{01} \hat{=} \left\{\mathbf{0}_m, \boldsymbol{e}_m^{(1)}, \ldots, \boldsymbol{e}_m^{(m)}, \mathbf{1}_m\right\}$, $\boldsymbol{\xi}_0$, $\{\boldsymbol{\xi}^{(i)}\}_{i=1}^m$, $\boldsymbol{\xi}_1$ are defined in Definition 1.

If $\mathbf{y} \sim \mathrm{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$, according to (2.2), we have the i-th component

$$
Y_i = Z_i + Z_{m+1} + Z_{m+2} X_i \sim \mathrm{ZOIP}\left(\phi_0 + \textstyle\sum_{k=1, k\neq i}^{m} \phi_k, \phi_i + \phi_{m+1}; \lambda_0 + \lambda_i\right). \qquad (2.5)
$$

From (2.5), we can see that the marginal distributions are not necessarily identical with each other; i.e., each Y_i follows a ZOIP distribution with different zero inflation, one inflation and Poisson mean parameters.

Moreover, we have

$$\begin{cases} E(\mathbf{y}) &= \boldsymbol{\phi} + \phi_{m+1}\cdot\mathbf{1} + \phi_{m+2}(\lambda_0\cdot\mathbf{1}+\boldsymbol{\lambda}) \ \hat{=}\ \boldsymbol{\mu}, \\ E(\mathbf{y}\mathbf{y}^\top) &= \mathrm{diag}(\boldsymbol{\phi}) + \phi_{m+1}\cdot\mathbf{1}\mathbf{1}^\top \\ &\quad + \phi_{m+2}\left[(\lambda_0\cdot\mathbf{1}+\boldsymbol{\lambda})(\lambda_0\cdot\mathbf{1}+\boldsymbol{\lambda})^\top + \lambda_0\cdot\mathbf{1}\mathbf{1}^\top + \mathrm{diag}(\boldsymbol{\lambda})\right], \\ \mathrm{Var}(\mathbf{y}) &= \mathrm{diag}(\boldsymbol{\phi}) + \phi_{m+1}\cdot\mathbf{1}\mathbf{1}^\top \\ &\quad + \phi_{m+2}\left[(\lambda_0\cdot\mathbf{1}+\boldsymbol{\lambda})(\lambda_0\cdot\mathbf{1}+\boldsymbol{\lambda})^\top + \lambda_0\cdot\mathbf{1}\mathbf{1}^\top + \mathrm{diag}(\boldsymbol{\lambda})\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\top, \end{cases}$$

where $\mathbf{1} = \mathbf{1}_m$. The correlation coefficient between Y_i and Y_j for $i \neq j$ is

$$\mathrm{Corr}(Y_i, Y_j) = \frac{\phi_{m+1} + \phi_{m+2}[(\lambda_0+\lambda_i)(\lambda_0+\lambda_j)+\lambda_0] - \mu_i\mu_j}{\sqrt{\left[\mu_i - \mu_i^2 + \dfrac{(\mu_i - \phi_i - \phi_{m+1})^2}{\phi_{m+2}}\right]\left[\mu_j - \mu_j^2 + \dfrac{(\mu_j - \phi_j - \phi_{m+1})^2}{\phi_{m+2}}\right]}}, \tag{2.6}$$

where $\mu_i = \phi_i + \phi_{m+1} + \phi_{m+2}(\lambda_0 + \lambda_i)$. From (2.6), the correlation coefficient between Y_i and Y_j could be either positive or negative depending on the values of those parameters.

3. Likelihood-based methods for the general multivariate ZOIP distribution/regression model

Suppose that $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ is a random sample of size n from the $\mathrm{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$ distribution, where $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{mj})^\top$ for $j = 1, \ldots, n$. Let $\boldsymbol{y}_j = (\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{mj})^\top$ denote the realization of the random vector $\mathbf{y}_j$ and $Y_{\mathrm{obs}} = \{\boldsymbol{y}_j\}_{j=1}^n$ be the observed data. Furthermore, we define

$$\begin{aligned} \mathbb{J}_0 &= \{j | \boldsymbol{y}_j = \mathbf{0}_m,\ j = 1, \ldots, n\}, \quad n_0 \,\hat{=}\, \#\{\mathbb{J}_0\} = \sum_{j=1}^n I(\boldsymbol{y}_j = \mathbf{0}_m), \\ \mathbb{J}_i &= \{j | \boldsymbol{y}_j = \boldsymbol{e}_m^{(i)},\ j = 1, \ldots, n\}, \quad n_i \,\hat{=}\, \#\{\mathbb{J}_i\} = \sum_{j=1}^n I(\boldsymbol{y}_j = \boldsymbol{e}_m^{(i)}), \quad 1 \leqslant i \leqslant m, \\ \mathbb{J}_{m+1} &= \{j | \boldsymbol{y}_j = \mathbf{1}_m,\ j = 1, \ldots, n\}, \quad n_{m+1} \,\hat{=}\, \#\{\mathbb{J}_{m+1}\} = \sum_{j=1}^n I(\boldsymbol{y}_j = \mathbf{1}_m), \\ \mathbb{J}_{m+2} &= \{j | \boldsymbol{y}_j \notin \mathcal{Y}_{01},\ j = 1, \ldots, n\}, \quad n_{m+2} \,\hat{=}\, \#\{\mathbb{J}_{m+2}\} = n - \sum_{k=0}^{m+1} n_k. \end{aligned}$$

The observed-data likelihood function of $\boldsymbol{\theta} = \left(\phi_0, \boldsymbol{\phi}^\top, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda}^\top\right)^\top$ is given by

$$L(\boldsymbol{\theta}|Y_{\rm obs}) = \left(\prod_{k=0}^{m+1} a_k^{n_k}\right)\left(\phi_{m+2}\mathrm{e}^{-\lambda_0-\lambda_+}\right)^{n_{m+2}} \prod_{j\in\mathbb{J}_{m+2}} \sum_{k_j=0}^{\min(\boldsymbol{y}_j)} \frac{\lambda_0^{k_j}}{k_j!} \prod_{i=1}^{m} \frac{\lambda_i^{y_{ij}-k_j}}{(y_{ij}-k_j)!},$$

so that the log-likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\theta}|Y_{\rm obs}) &= \sum_{k=0}^{m+1} n_k \log a_k + n_{m+2}(\log\phi_{m+2} - \lambda_0 - \lambda_+) \\ &\quad + \sum_{j\in\mathbb{J}_{m+2}} \log\left[\sum_{k_j=0}^{\min(\boldsymbol{y}_j)} \frac{\lambda_0^{k_j}}{k_j!} \prod_{i=1}^{m} \frac{\lambda_i^{y_{ij}-k_j}}{(y_{ij}-k_j)!}\right], \end{aligned} \tag{3.1}$$

where $\{a_k\}_{k=0}^{m+1}$ are defined in (2.4).

3.1 Maximum likelihood estimation

3.1.1 MLEs via the EM algorithm

To obtain the *maximum likelihood estimates* (MLEs) of parameters, we employ the EM algorithm. We first augment $Y_{\rm obs}$ with latent variables $\{U_k\}_{k=0}^m$ that split n_k into $(U_k, n_k - U_k)$, and (W_1, W_2, W_3) that split n_{m+1} into (W_1, W_2, W_3) where $W_3 \hat{=} n_{m+1} - W_1 - W_2$, and for each $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{mj})^\top$ where $j \in \mathbb{J}_{m+2}$, we introduce latent variables $X_{0j}^* \overset{\rm ind}{\sim} \text{Poisson}\,(\lambda_0)$, $X_{ij}^* \overset{\rm ind}{\sim} \text{Poisson}\,(\lambda_i)$ for $1 \leqslant i \leqslant m$ and $X_{0j}^* \perp\!\!\!\perp X_{ij}^*$, such that

$$(x_{0j}^* + x_{1j}^*, \ldots, x_{0j}^* + x_{mj}^*)^\top = \boldsymbol{y}_j, \quad j \in \mathbb{J}_{m+2},$$

where x_{ij}^* denotes the realization of X_{ij}^*. The complete data is composed of

$$\begin{aligned} Y_{\rm com} &= \left\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n, u_0, u_1, \ldots, u_m, w_1, w_2, w_3, \{x_{0j}^*, x_{1j}^*, \ldots, x_{mj}^*\}_{j\in\mathbb{J}_{m+2}}\right\} \\ &= \left\{\{\boldsymbol{y}_j\}_{j=1}^n, \{u_k\}_{k=0}^m, w_1, w_2, w_3, \{x_{0j}^*\}_{j\in\mathbb{J}_{m+2}}\right\} \end{aligned}$$

since $x_{ij}^* = y_{ij} - x_{0j}^*$ when $j \in \mathbb{J}_{m+2}$ for $1 \leqslant i \leqslant m$. Therefore, the resultant conditional predictive distributions of $\{U_k\}_{k=0}^m$ and $\{W_k\}_{k=1}^3$ given $(Y_{\rm obs}, \boldsymbol{\theta})$ are obtained as

$$U_k|(Y_{\rm obs}, \boldsymbol{\theta}) \sim \text{Binomial}\left(n_k, \frac{\phi_k}{a_k}\right), \quad 0 \leqslant k \leqslant m, \tag{3.2}$$

$$(W_1, W_2, W_3)^\top|(Y_{\rm obs}, \boldsymbol{\theta}) \sim \text{Multinomial}\left(n_{m+1}; \frac{a_{m+1,1}}{a_{m+1}}, \frac{a_{m+1,2}}{a_{m+1}}, \frac{a_{m+1,3}}{a_{m+1}}\right), \tag{3.3}$$

where

$$a_{m+1,1} = \phi_{m+1}, \quad a_{m+1,2} = \phi_{m+2}\lambda_0\,\mathrm{e}^{-\lambda_0-\lambda_+}, \quad a_{m+1,3} = \phi_{m+2}\left(\prod_{i=1}^m \lambda_i\right)\mathrm{e}^{-\lambda_0-\lambda_+}.$$

Thus, the complete-data likelihood function is proportional to

$$\begin{aligned}
& L(\boldsymbol{\theta}|Y_{\text{com}}) \\
\propto\ & \phi_0^{u_0}\left(\phi_{m+2}\mathrm{e}^{-\lambda_0-\lambda_+}\right)^{n_0-u_0}\left[\prod_{i=1}^m \phi_i^{u_i}\left(\phi_{m+2}\lambda_i\mathrm{e}^{-\lambda_0-\lambda_+}\right)^{n_i-u_i}\right]\phi_{m+1}^{w_1}(\phi_{m+2}\lambda_0\mathrm{e}^{-\lambda_0-\lambda_+})^{w_2} \\
& \times \left[\phi_{m+2}(\textstyle\prod_{i=1}^m \lambda_i)\mathrm{e}^{-\lambda_0-\lambda_+}\right]^{w_3}(\phi_{m+2}\mathrm{e}^{-\lambda_0-\lambda_+})^{n_{m+2}}\displaystyle\prod_{j\in\mathbb{J}_{m+2}}\lambda_0^{x_{0j}^*}\prod_{i=1}^m\lambda_i^{y_{ij}-x_{0j}^*} \\
=\ & (\textstyle\prod_{k=0}^m \phi_k^{u_k})\,\phi_{m+1}^{w_1}\phi_{m+2}^{n-\sum_{k=0}^m u_k-w_1}\exp\left[-(n-\sum_{k=0}^m u_k - w_1)(\lambda_0+\lambda_+)\right] \\
& \times \lambda_0^{w_2+N_0}\displaystyle\prod_{i=1}^m \lambda_i^{w_3+n_i-u_i+N_i-N_0},
\end{aligned}$$

where $N_0 = \sum_{j\in\mathbb{J}_{m+2}} x_{0j}^*$, $N_i = \sum_{j\in\mathbb{J}_{m+2}} y_{ij} = \sum_{j=1}^n y_{ij} - n_i - n_{m+1}$ for $i = 1, \ldots, m$. Thus, the M-step is to find the complete-data MLEs

$$\begin{cases}
\hat{\phi}_k = \dfrac{u_k}{n}, \quad 0 \leqslant k \leqslant m, & \hat{\phi}_{m+1} = \dfrac{w_1}{n}, \\[2ex]
\hat{\lambda}_0 = \dfrac{w_2+N_0}{n-\sum_{k=0}^m u_k - w_1}, & \hat{\lambda}_i = \dfrac{w_3+n_i-u_i+N_i-N_0}{n-\sum_{k=0}^m u_k - w_1}, \quad 1 \leqslant i \leqslant m.
\end{cases} \tag{3.4}$$

The E-step is to replace $\{u_k\}_{k=0}^m$, $\{w_k\}_{k=1}^3$ and $\{x_{0j}^*\}_{j\in\mathbb{J}_{m+2}}$ in (3.4) by their conditional expectations, which are given by

$$\begin{cases}
E(U_k|Y_{\text{obs}},\boldsymbol{\theta}) \overset{(3.2)}{=} \dfrac{n_k\phi_k}{a_k}, \quad 0 \leqslant k \leqslant m, \\[2ex]
E(W_k|Y_{\text{obs}},\boldsymbol{\theta}) \overset{(3.3)}{=} \dfrac{n_{m+1}a_{m+1,k}}{a_{m+1}}, \quad k = 1, 2, 3, \\[2ex]
E(X_{0j}^*|Y_{\text{obs}},\boldsymbol{\theta}) \overset{(A.5)}{=} \dfrac{\displaystyle\sum_{k_j=1}^{\min(\boldsymbol{y}_j)} \frac{\lambda_0^{k_j}}{(k_j-1)!}\prod_{i=1}^m \frac{\lambda_i^{y_{ij}-k_j}}{(y_{ij}-k_j)!}}{\displaystyle\sum_{l_j=0}^{\min(\boldsymbol{y}_j)} \frac{\lambda_0^{l_j}}{l_j!}\prod_{i=1}^m \frac{\lambda_i^{y_{ij}-l_j}}{(y_{ij}-l_j)!}}, \quad j \in \mathbb{J}_{m+2}.
\end{cases} \tag{3.5}$$

The detail for deriving (3.5) is given in Appendix A.2.

3.1.2 MLEs via the Fisher scoring algorithm

For the special case of $\lambda_0 = 0$, all $\{\lambda_i\}_{i=1}^m$ in the last term of (3.1) are multiplicatively separable which makes the calculations of the Hessian matrix (so as the Fisher information matrix) feasible, thus the Fisher scoring algorithm can also be applied to obtain the MLEs of the parameter vector $\boldsymbol{\theta} = \left(\phi_0, \boldsymbol{\phi}^\top, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda}^\top\right)^\top$ except for λ_0, denoted by $\boldsymbol{\theta}_{-\lambda_0}$. Specifically, the log-likelihood function now becomes $\ell(\boldsymbol{\theta}_{-\lambda_0}|Y_{\text{obs}})$ obtained by replacing λ_0 with zero value in (3.1). Then the score vector and the Hessian matrix are given by

$$\nabla\ell(\boldsymbol{\theta}_{-\lambda_0}|Y_{\text{obs}}) = \frac{\partial\ell(\boldsymbol{\theta}_{-\lambda_0}|Y_{\text{obs}})}{\partial\boldsymbol{\theta}_{-\lambda_0}} \quad \text{and} \quad \nabla^2\ell(\boldsymbol{\theta}_{-\lambda_0}|Y_{\text{obs}}) = \frac{\partial^2\ell(\boldsymbol{\theta}_{-\lambda_0}|Y_{\text{obs}})}{\partial\boldsymbol{\theta}_{-\lambda_0}\partial\boldsymbol{\theta}_{-\lambda_0}^\top},$$

respectively. Thus, the $(2m+2)\times(2m+2)$ Fisher information matrix is

$$\mathbf{J}(\boldsymbol{\theta}_{-\lambda_0}) = E\Big[-\nabla^2\ell(\boldsymbol{\theta}_{-\lambda_0}|Y_{\text{obs}})\Big]. \tag{3.6}$$

Let $\boldsymbol{\theta}_{-\lambda_0}^{(0)}$ be the initial value and $\boldsymbol{\theta}_{-\lambda_0}^{(t)}$ denote the t-th approximation of $\hat{\boldsymbol{\theta}}_{-\lambda_0}$, then the $(t+1)$-th approximation can be obtained by

$$\boldsymbol{\theta}_{-\lambda_0}^{(t+1)} = \boldsymbol{\theta}_{-\lambda_0}^{(t)} + \mathbf{J}^{-1}(\boldsymbol{\theta}_{-\lambda_0}^{(t)})\nabla\ell(\boldsymbol{\theta}_{-\lambda_0}^{(t)}|Y_{\text{obs}}). \tag{3.7}$$

As a by-product, the standard errors of the MLEs $\hat{\boldsymbol{\theta}}_{-\lambda_0}$ are the square roots of the diagonal elements J^{kk} of the inverse Fisher information matrix $\mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}_{-\lambda_0})$. Thus, the $100(1-\alpha)\%$ asymptotic Wald *confidence intervals* (CIs) of each component in $\boldsymbol{\theta}_{-\lambda_0}$ are given by

$$\begin{aligned} &[\hat{\phi}_{k-1} - z_{\alpha/2}\sqrt{J^{kk}}, \quad \hat{\phi}_{k-1} + z_{\alpha/2}\sqrt{J^{kk}}\,], \qquad 1 \leqslant k \leqslant m+2, \quad \text{and} \\ &[\hat{\lambda}_i - z_{\alpha/2}\sqrt{J^{m+2+i,m+2+i}}, \quad \hat{\lambda}_i + z_{\alpha/2}\sqrt{J^{m+2+i,m+2+i}}\,], \qquad 1 \leqslant i \leqslant m, \end{aligned} \tag{3.8}$$

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

3.2 Bootstrap confidence intervals

First, for the general log-likelihood function (3.1) associated with the proposed distribution (2.3), the calculation of the Hessian matrix or the Fisher information matrix seems to be too complicated, thus the standard errors of the estimators cannot be easily obtained. Second, even when they are obtainable, the resulting asymptotic CIs for parameters ϕ_k's or λ_i's are reliable only for large sample size and may become useless if either boundary for ϕ_k is beyond $[0,1]$ or the lower bound for λ_i is less than 0. Thus, under the current situation, the bootstrap method is a useful tool to find a bootstrap CI for an arbitrary function of $\boldsymbol{\theta} = (\phi_0, \boldsymbol{\phi}^\top, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda}^\top)^\top$, say, $\vartheta = h(\boldsymbol{\theta})$. Let $\hat{\vartheta} = h(\hat{\boldsymbol{\theta}})$ denote the MLE of ϑ, where $\hat{\boldsymbol{\theta}}$ represent the MLEs of $\boldsymbol{\theta}$ calculated by the EM algorithm (3.4)–(3.5). Based on the obtained MLEs $\hat{\boldsymbol{\theta}}$, by using (2.2) we can generate $\mathbf{y}_1^*, \ldots, \mathbf{y}_n^* \stackrel{\text{iid}}{\sim} \text{ZOIP}_m(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\phi}_{m+1}; \hat{\lambda}_0, \hat{\boldsymbol{\lambda}})$. Having obtained $Y_{\text{obs}}^* = \{\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_n^*\}$, we can calculate the bootstrap replications $\hat{\boldsymbol{\theta}}^*$ and get $\hat{\vartheta}^* = h(\hat{\boldsymbol{\theta}}^*)$. Independently repeat 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.,

$$\widehat{\text{se}}(\hat{\vartheta}) = \left\{ \frac{1}{G-1} \sum_{g=1}^{G} [\hat{\vartheta}_g^* - (\hat{\vartheta}_1^* + \cdots + \hat{\vartheta}_G^*)/G]^2 \right\}^{1/2}. \tag{3.9}$$

The $100(1-\alpha)\%$ bootstrap CI for ϑ is given by

$$[\hat{\vartheta}_{\text{L}},\ \hat{\vartheta}_{\text{U}}], \tag{3.10}$$

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.

3.3 Testing hypotheses for large sample sizes

Since fitting the multivariate count data with the full model specified by (2.3) strictly depends on the proportions of the data category, we first consider some reduced models. For example, we could test whether ϕ_{m+1} or λ_0 is equal to 0.

3.3.1 Likelihood ratio test for testing $\phi_{m+1} = 0$

Suppose that we want to test

$$H_0\colon\ \phi_{m+1} = 0 \quad \text{against} \quad H_1\colon\ \phi_{m+1} > 0. \tag{3.11}$$

Under H_0, the *likelihood ratio test* (LRT) statistic

$$T_1 = -2\Big\{\ell(\hat{\phi}_{0,H_0}, \hat{\boldsymbol{\phi}}_{H_0}, 0, \hat{\lambda}_{0,H_0}, \hat{\boldsymbol{\lambda}}_{H_0}|Y_{\text{obs}}) - \ell(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\phi}_{m+1}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}}|Y_{\text{obs}})\Big\}, \tag{3.12}$$

where $(\hat{\phi}_{0,H_0}, \hat{\boldsymbol{\phi}}_{H_0}, 0, \hat{\lambda}_{0,H_0}, \hat{\boldsymbol{\lambda}}_{H_0})$ denote the constrained MLEs of $(\phi_0, \boldsymbol{\phi}, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda})$ under H_0 and $(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\phi}_{m+1}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}})$ denote the unconstrained MLEs of $(\phi_0, \boldsymbol{\phi}, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda})$. Since the null hypothesis in (3.11) corresponds to ϕ_{m+1} being on the boundary of the parameter space and the appropriate null distribution is a mixture of $\chi^2(0)$ (i.e., Degenerate(0)) and $\chi^2(1)$ with equal weights (Self & Liang, 1987). Thus the corresponding p-value is

$$p_{\rm v1} = \Pr(T_1 > t_1|H_0) = \frac{1}{2}\Pr(\chi^2(1) > t_1),$$

where t_1 is the observed value of T_1.

3.3.2 Likelihood ratio test for testing $\lambda_0 = 0$

Suppose that we want to test

$$H_0\colon\ \lambda_0 = 0 \quad \text{against} \quad H_1\colon\ \lambda_0 > 0. \tag{3.13}$$

Under H_0, the LRT statistic

$$T_2 = -2\Big\{\ell(\hat{\phi}_{0,H_0}, \hat{\boldsymbol{\phi}}_{H_0}, \hat{\phi}_{m+1,H_0}, 0, \hat{\boldsymbol{\lambda}}_{H_0}|Y_{\rm obs}) - \ell(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\phi}_{m+1}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}}|Y_{\rm obs})\Big\}, \tag{3.14}$$

where $(\hat{\phi}_{0,H_0}, \hat{\boldsymbol{\phi}}_{H_0}, \hat{\phi}_{m+1,H_0}, 0, \hat{\boldsymbol{\lambda}}_{H_0})$ denote the constrained MLEs of $(\phi_0, \boldsymbol{\phi}, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda})$ under H_0 and $(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\phi}_{m+1}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}})$ denote the unconstrained MLEs of $(\phi_0, \boldsymbol{\phi}, \phi_{m+1}, \lambda_0, \boldsymbol{\lambda})$. The corresponding p-value is given by

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

where t_2 is the observed value of T_2.

3.4 Multivariate ZOIP regression model

In this subsection, we extend the proposed distribution by incorporating covariates into the data analysis. Considering that the full model is rarely satisfied and for simplicity of model formulation, we focus on the case with $\lambda_0 = 0$. We use the multinomial logistic regression to link $(\phi_0, \phi_1, \ldots, \phi_{m+2})$ with the covariates via the logit transformation. Moreover, the Poisson parameters $\boldsymbol{\lambda}$ can be modeled by the ordinary log-linear regression. Thus, we consider the following multivariate ZOIP regression model:

$$\begin{cases} \mathbf{y}_j & \stackrel{\text{ind}}{\sim} & \text{ZOIP}_m(\phi_{0j}, \boldsymbol{\phi}_j, \phi_{m+1,j}; 0, \boldsymbol{\lambda}_j), \quad 1 \leqslant j \leqslant n, \\ \phi_{kj} & = & \dfrac{\exp(\boldsymbol{w}_j^\top \boldsymbol{\gamma}_k)}{1 + \sum_{i=0}^{m+1} \exp(\boldsymbol{w}_j^\top \boldsymbol{\gamma}_i)}, \quad 0 \leqslant k \leqslant m+1, \\ \phi_{m+2,j} & = & \dfrac{1}{1 + \sum_{i=0}^{m+1} \exp(\boldsymbol{w}_j^\top \boldsymbol{\gamma}_i)}, \\ \lambda_{ij} & = & \exp(\boldsymbol{x}_j^\top \boldsymbol{\beta}_i), \quad 1 \leqslant i \leqslant m, \end{cases}$$

where $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{mj})^\top$ is the response vector of the subject j, $\boldsymbol{\phi}_j = (\phi_{1j}, \ldots, \phi_{mj})^\top$, $\boldsymbol{\lambda}_j = (\lambda_{1j}, \ldots, \lambda_{mj})^\top$, $\boldsymbol{w}_j = (1, w_{1j}, \ldots, w_{pj})^\top$ and $\boldsymbol{x}_j = (1, x_{1j}, \ldots, x_{qj})^\top$ are not necessarily identical covariate vectors associated with the subject j, $\boldsymbol{\gamma}_k = (\gamma_{k0}, \gamma_{k1}, \ldots, \gamma_{kp})^\top$ and $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \ldots, \beta_{iq})^\top$ are vectors of regression coefficients, respectively. Note that the component $\phi_{m+2,j}$ is taken as the baseline for the multinomial logit model. Thus, the logarithm for other components relative to $\phi_{m+2,j}$ is

$$\log\left(\frac{\phi_{kj}}{\phi_{m+2,j}}\right) = \boldsymbol{w}_j^\top \boldsymbol{\gamma}_k, \quad 0 \leqslant k \leqslant m+1.$$

First, we define $I_{0j} \hat{=} I(\boldsymbol{y}_j = \mathbf{0}_m)$, $I_{ij} \hat{=} I(\boldsymbol{y}_j = \boldsymbol{e}_m^{(i)})$ for $1 \leqslant i \leqslant m$, $I_{m+1,j} \hat{=} I(\boldsymbol{y}_j = \mathbf{1}_m)$ and $I_{m+2,j} \hat{=} I(\boldsymbol{y}_j \notin \mathcal{Y}_{01})$. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_{m+1}^\top)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_m^\top)^\top$, $Y_{\text{obs}} = \{\boldsymbol{y}_j, \boldsymbol{w}_j, \boldsymbol{x}_j\}_{j=1}^n$. Then, the observed-data likelihood function is

$$L_1'(\boldsymbol{\gamma}, \boldsymbol{\beta} | Y_{\text{obs}}) = \prod_{j=1}^n \left[\left(\prod_{i=0}^{m+1} b_{ij}^{I_{ij}} \right) \left(\phi_{m+2,j} \mathrm{e}^{-\lambda_{+j}} \prod_{i=1}^m \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right)^{I_{m+2,j}} \right],$$

where

$$\begin{aligned} b_{0j} &= \phi_{0j} + \phi_{m+2,j} \mathrm{e}^{-\lambda_{+j}}, \quad b_{ij} = \phi_{ij} + \phi_{m+2,j} \lambda_{ij} \mathrm{e}^{-\lambda_{+j}}, \quad 1 \leqslant i \leqslant m, \\ b_{m+1,j} &= \phi_{m+1,j} + \phi_{m+2,j} \mathrm{e}^{-\lambda_{+j}} \prod_{i=1}^m \lambda_{ij}, \end{aligned}$$

and $\lambda_{+j} = \sum_{i=1}^m \lambda_{ij}$. Similarly, we augment Y_{obs} with $(m+2) \times n$ latent variables U_{kj}'s for $k = 0, 1, \ldots, m+1$ and $j = 1, \ldots, n$. The conditional predictive distributions of $\{U_{kj}\}$ given

$(Y_{\rm obs}, \boldsymbol{\gamma}, \boldsymbol{\beta})$ are

$$U_{kj}|(Y_{\rm obs}, \boldsymbol{\gamma}, \boldsymbol{\beta}) \sim \text{Bernoulli}\left(\frac{\phi_{kj}}{b_{kj}}\right), \quad 0 \leqslant k \leqslant m+1,$$

Denote the missing data by $Y_{\rm mis} = \{\{u_{kj}\}_{k=0}^{m+1}\}_{j=1}^{n}$, $Y_{\rm com} = \{Y_{\rm obs}, Y_{\rm mis}\}$, then the complete-data likelihood function is

$$\begin{aligned}
L_1'(\boldsymbol{\gamma}, \boldsymbol{\beta}|Y_{\rm com}) \;\propto\; & \prod_{j=1}^{n} \Bigg[\phi_{0j}^{u_{0j}I_{0j}} \left(\phi_{m+2,j}\mathrm{e}^{-\lambda_{+j}}\right)^{(1-u_{0j})I_{0j}} \prod_{i=1}^{m} \phi_{ij}^{u_{ij}I_{ij}} \left(\phi_{m+2,j}\lambda_{ij}\mathrm{e}^{-\lambda_{+j}}\right)^{(1-u_{ij})I_{ij}} \\
& \times \phi_{m+1,j}^{u_{m+1,j}I_{m+1,j}} \left(\phi_{m+2,j}\mathrm{e}^{-\lambda_{+j}} \textstyle\prod_{i=1}^{m} \lambda_{ij}\right)^{(1-u_{m+1,j})I_{m+1,j}} \\
& \times \left(\phi_{m+2,j}\mathrm{e}^{-\lambda_{+j}} \textstyle\prod_{i=1}^{m} \lambda_{ij}^{y_{ij}}\right)^{I_{m+2,j}} \Bigg],
\end{aligned}$$

and the complete-data log-likelihood is

$$\begin{aligned}
\ell_1'(\boldsymbol{\gamma}, \boldsymbol{\beta}|Y_{\rm com}) \;=\; & \sum_{j=1}^{n} \Bigg[\sum_{k=0}^{m+1} u_{kj}I_{kj}\log\phi_{kj} + \sum_{k=0}^{m+1}(1-u_{kj})I_{kj}\log\phi_{m+2,j} + I_{m+2,j}\log\phi_{m+2,j} \\
& - \sum_{k=0}^{m+1}(1-u_{kj})I_{kj}\lambda_{+j} - I_{m+2,j}\lambda_{+j} + \sum_{i=1}^{m}(1-u_{ij})I_{ij}\log\lambda_{ij} \\
& + \sum_{i=1}^{m}(1-u_{m+1,j})I_{m+1,j}\log\lambda_{ij} + \sum_{i=1}^{m} I_{m+2,j}y_{ij}\log\lambda_{ij} \Bigg] \\
\hat{=}\; & \ell_{11}(\boldsymbol{\gamma}|Y_{\rm com}) + \ell_{12}(\boldsymbol{\beta}|Y_{\rm com}),
\end{aligned}$$

where

$$\begin{aligned}
\ell_{11} &= \ell_{11}(\boldsymbol{\gamma}|Y_{\rm com}) \\
&= \sum_{j=1}^{n} \left[\sum_{k=0}^{m+1} u_{kj}I_{kj}\log\phi_{kj} + \sum_{k=0}^{m+1}(1-u_{kj})I_{kj}\log\phi_{m+2,j} + I_{m+2,j}\log\phi_{m+2,j} \right], \\
\ell_{12} &= \ell_{12}(\boldsymbol{\beta}|Y_{\rm com}) = -\sum_{j=1}^{n} \Bigg[\sum_{k=0}^{m+1}(1-u_{kj})I_{kj}\lambda_{+j} + I_{m+2,j}\lambda_{+j} - \sum_{i=1}^{m}(1-u_{ij})I_{ij}\log\lambda_{ij} \\
&\quad - \sum_{i=1}^{m}(1-u_{m+1,j})I_{m+1,j}\log\lambda_{ij} - \sum_{i=1}^{m} y_{ij}I_{m+2,j}\log\lambda_{ij} \Bigg],
\end{aligned}$$

which only involves ϕ_{ij}'s and λ_{ij}'s, respectively. For convenience, we define a new operator "$\circ$" by $\boldsymbol{u} \circ \boldsymbol{y}_i = (u_1 y_{i1}, \ldots, u_n y_{in})^{\top}$. Then we have

$$
\begin{aligned}
\frac{\partial \ell_{11}}{\partial \boldsymbol{\gamma}_k} &= \sum_{j=1}^{n} (u_{kj} I_{kj} \boldsymbol{w}_j - \phi_{kj} \boldsymbol{w}_j) = \mathbf{W}^{\top} \left(\boldsymbol{u}_{(k)} \circ \boldsymbol{i}_{(k)} - \boldsymbol{\phi}_{(k)} \right), \quad 0 \leqslant k \leqslant m+1, \\
\frac{\partial \ell_{12}}{\partial \boldsymbol{\beta}_i} &= -\sum_{j=1}^{n} \Bigg[\lambda_{ij} \boldsymbol{x}_j - \sum_{k=0}^{m+1} u_{kj} I_{kj} \lambda_{ij} \boldsymbol{x}_j - (1-u_{ij}) I_{ij} \boldsymbol{x}_j \\
&\quad - (1-u_{m+1,j}) I_{m+1,j} \boldsymbol{x}_j - I_{m+2,j} y_{ij} \boldsymbol{x}_j \Bigg] \\
&= -\mathbf{X}^{\top} \Bigg[\boldsymbol{\lambda}_{(i)} - \sum_{k=0}^{m+1} \left(\boldsymbol{u}_{(k)} \circ \boldsymbol{i}_{(k)} \right) \circ \boldsymbol{\lambda}_{(i)} - \left(\mathbf{1} - \boldsymbol{u}_{(i)} \right) \circ \boldsymbol{i}_{(i)} \\
&\quad - \left(\mathbf{1} - \boldsymbol{u}_{(m+1)} \right) \circ \boldsymbol{i}_{(m+1)} - \boldsymbol{y}_{(i)} \circ \boldsymbol{i}_{(m+2)} \Bigg], \quad 1 \leqslant i \leqslant m, \\
\frac{\partial^2 \ell_{11}}{\partial \boldsymbol{\gamma}_k \partial \boldsymbol{\gamma}_k^{\top}} &= -\sum_{j=1}^{n} \phi_{kj} (1-\phi_{kj}) \boldsymbol{w}_j \boldsymbol{w}_j^{\top} = -\mathbf{W}^{\top} \mathrm{diag} \left[\boldsymbol{\phi}_{(k)} \circ \left(\mathbf{1} - \boldsymbol{\phi}_{(k)} \right) \right] \mathbf{W}, \\
\frac{\partial^2 \ell_{11}}{\partial \boldsymbol{\gamma}_k \partial \boldsymbol{\gamma}_{k'}^{\top}} &= \sum_{j=1}^{n} \phi_{kj} \phi_{k'j} \boldsymbol{w}_j \boldsymbol{w}_j^{\top} = \mathbf{W}^{\top} \mathrm{diag} \left(\boldsymbol{\phi}_{(k)} \circ \boldsymbol{\phi}_{(k')} \right) \mathbf{W}, \quad k \neq k', \\
\frac{\partial^2 \ell_{12}}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i^{\top}} &= -\sum_{j=1}^{n} \left(\lambda_{ij} \boldsymbol{x}_j \boldsymbol{x}_j^{\top} - \sum_{k=0}^{m+1} u_{kj} I_{kj} \lambda_{ij} \boldsymbol{x}_j \boldsymbol{x}_j^{\top} \right) \\
&= -\mathbf{X}^{\top} \mathrm{diag} \Bigg[\boldsymbol{\lambda}_{(i)} - \sum_{k=0}^{m+1} \left(\boldsymbol{u}_{(k)} \circ \boldsymbol{i}_{(k)} \right) \circ \boldsymbol{\lambda}_{(i)} \Bigg] \mathbf{X}, \\
\frac{\partial^2 \ell_{12}}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_{i'}^{\top}} &= \mathbf{0}, \quad i \neq i',
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{W} &= (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)^{\top}, \quad \boldsymbol{u}_{(k)} = (u_{k1}, \ldots, u_{kn})^{\top}, \\
\boldsymbol{i}_{(l)} &= (I_{l1}, \ldots, I_{ln})^{\top}, \quad \boldsymbol{\phi}_{(k)} = (\phi_{k1}, \ldots, \phi_{kn})^{\top}, \\
\mathbf{X} &= (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^{\top}, \quad \boldsymbol{\lambda}_{(i)} = (\lambda_{i1}, \ldots, \lambda_{in})^{\top}, \\
\boldsymbol{y}_{(i)} &= (y_{i1}, \ldots, y_{in})^{\top},
\end{aligned}
$$

and $l = 0, 1, \ldots, m+2$. The M-step is to embed the Newton–Raphson algorithm to update each iteration and E-step is to replace all u_{kj}'s by their conditional expectations.

After we obtained the MLEs of $(\boldsymbol{\gamma}_k, \boldsymbol{\beta}_i)$, denoted by $(\hat{\boldsymbol{\gamma}}_k, \hat{\boldsymbol{\beta}}_i)$, we are interested in finding the standard errors of $(\hat{\gamma}_k, \hat{\beta}_i)$. Therefore, we need to derive the observed information matrix. Let $\boldsymbol{\theta}$ be the parameters of interest. According to Louis (1982), the observed information matrix can be calculated as

$$\mathbf{I}(\hat{\boldsymbol{\theta}}|Y_{\text{obs}}) = \Big\{E[-\nabla\ell^2(\boldsymbol{\theta}|Y_{\text{com}})|Y_{\text{obs}}, \boldsymbol{\theta}] - E\{[\nabla\ell(\boldsymbol{\theta}|Y_{\text{com}})]^{\bigotimes 2}|Y_{\text{obs}}, \boldsymbol{\theta}\}\Big\}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{3.15}$$

where $\boldsymbol{a}^{\bigotimes 2} = \boldsymbol{a}\boldsymbol{a}^{\top}$, $\nabla\ell^2(\boldsymbol{\theta}|Y_{\text{com}})$ and $\nabla\ell(\boldsymbol{\theta}|Y_{\text{com}})$ are the Hessian matrix and the gradient vector of the complete-data log-likelihood function. Note that the key point in (3.15) is the computation of the expectations of the terms involving the latent variables U_{kj}'s. Since U_{kj} for $j = 1, \ldots, n$ independently follows the Bernoulli distribution, thus we have

$$E(U_{kj}^2) = E(U_{kj}) \quad \text{and} \quad E(U_{kj}U_{k'j'}) = E(U_{kj})E(U_{k'j'})$$

for $j \neq j'$ and $k, k' = 0, 1, \ldots, m+1$. The estimated standard errors are the square roots of the diagonal elements of the inverse observed information matrix $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}|Y_{\text{obs}})$.

Alternatively, we use the square roots of the diagonal elements of the inversed complete information matrix, i.e., $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}|Y_{\text{com}})$, to approximate the estimated errors which is

$$\mathbf{I}(\hat{\boldsymbol{\theta}}|Y_{\text{com}}) = E[-\nabla\ell^2(\boldsymbol{\theta}|Y_{\text{com}})|Y_{\text{obs}}, \boldsymbol{\theta}]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{3.16}$$

4. Bayesian methods

For the reduced model with $\lambda_0 = 0$, we could consider Bayesian methods to compute the posterior modes and generate posterior samples from which we see that all results have explicit expressions.

4.1 Posterior modes via the EM algorithm

To derive the posterior modes of $\boldsymbol{\theta}_{-\lambda_0} = \big(\phi_0, \boldsymbol{\phi}^{\top}, \phi_{m+1}, \boldsymbol{\lambda}^{\top}\big)^{\top}$, we employ the EM algorithm again. Similar to the way of introducing latent variables in Section 3.1.1, in the current case we introduce $\{U_k\}_{k=0}^{m+1}$ to split n_k into U_k and $n_k - U_k$, respectively. The conditional predictive distributions of $\{U_k\}_{k=0}^{m+1}$ are given by

$$U_k|(Y_{\text{obs}}, \boldsymbol{\theta}_{-\lambda_0}) \sim \text{Binomial}\left(n_k, \frac{\phi_k}{b_k}\right), \quad 0 \leqslant k \leqslant m+1. \tag{4.1}$$

To assign priors for the parameters, a Dirichlet$(\delta_0, \delta_1, \ldots, \delta_{m+2})$ is adopted as the prior distribution of $(\phi_0, \phi_1, \ldots, \phi_{m+2})^\top$, a Gamma$(\alpha_i, \beta_i)$ is adopted as the prior distribution of λ_i for $1 \leqslant i \leqslant m$, and they are mutually independent. Then, the complete-data posterior distributions are given by

$$\begin{aligned}(\phi_0, \ldots, \phi_{m+2})^\top | Y_{\text{com}} &\sim \text{Dirichlet}\,(u_0 + \delta_0, u_1 + \delta_1, \ldots, u_{m+1} + \delta_{m+1}, n' + \delta_{m+2}), \\ \lambda_i | Y_{\text{com}} &\sim \text{Gamma}\,(n_i - u_i + n_{m+1} - u_{m+1} + N_i + \alpha_i, n' + \beta_i),\end{aligned} \tag{4.2}$$

for $i = 1, \ldots, m$, where $n' = n - \sum_{k=0}^{m+1} u_k$. The M-step is to calculate the complete-data posterior modes of $(\phi_0, \phi_1, \ldots, \phi_{m+2})$ and $\boldsymbol{\lambda}$, which are given by

$$\begin{cases} \phi_k &= \dfrac{u_k + \delta_k - 1}{n + \delta_+ - m - 3}, \quad 0 \leqslant k \leqslant m+1, \\ \phi_{m+2} &= \dfrac{n' + \delta_{m+2} - 1}{n + \delta_+ - m - 3}, \\ \lambda_i &= \dfrac{N_i + n_{m+1} - u_{m+1} + n_i - u_i + \alpha_i - 1}{n' + \beta_i}, \quad 1 \leqslant i \leqslant m, \end{cases} \tag{4.3}$$

where $\delta_+ = \sum_{k=0}^{m+2} \delta_k$, and the E-step is to replace $\{u_k\}_{k=0}^{m+1}$ by their conditional expectations, i.e., $n_k \phi_k / b_k$, directly derived from (4.1).

4.2 Generation of posterior samples via the DA algorithm

To make a full Bayesian inference on the parameters $\boldsymbol{\theta}_{-\lambda_0}$, we need to generate posterior samples from the observed posterior distribution $f(\boldsymbol{\theta}_{-\lambda_0} | Y_{\text{obs}})$ by using the *data augmentation* (DA) algorithm (Tanner & Wong, 1987). The I-step of the DA algorithm is to draw the missing values of $\{U_k = u_k\}_{k=0}^{m+1}$ for given $(Y_{\text{obs}}, \boldsymbol{\theta}_{-\lambda_0})$ from (4.1), and the P-step is to draw $\boldsymbol{\theta}_{-\lambda_0}$ from (4.2) for given $(Y_{\text{obs}}, u_0, u_1 \ldots, u_{m+1})$.

5. Simulations studies

To assess the performance of the proposed methods in Section 3 for the multivariate ZOIP distribution, we first concern the accuracy of the point estimators and the interval estimators, and then investigate the performance of the proposed LRTs in Section 3.3 by calculating their levels and powers via simulations.

5.1 Accuracy of the point and interval estimators

To evaluate the accuracy of the point and interval estimators of parameters, we consider two cases for the dimension: $m = 2$ and $m = 3$. Four combinations of parameter configurations are set as follows:

(1) Case I: When $m = 2$, $\phi_0 = 0.3$, $(\phi_1, \phi_2) = (0.2, 0.2)$, $\phi_3 = 0.1$, $\lambda_0 = 2$, $(\lambda_1, \lambda_2) = (2, 3)$;

(2) Case Ⅱ: When $m = 2$, $\phi_0 = 0.65$, $(\phi_1, \phi_2) = (0.05, 0.08)$, $\phi_3 = 0.07$, $\lambda_0 = 6$, $(\lambda_1, \lambda_2) = (2, 3)$;

(3) Case Ⅲ: When $m = 3$, $\phi_0 = 0.3$, $(\phi_1, \phi_2, \phi_3) = (0.2, 0.1, 0.1)$, $\phi_4 = 0.1$, $\lambda_0 = 1$, $(\lambda_1, \lambda_2, \lambda_3) = (2, 1, 3)$;

(4) Case Ⅳ: When $m = 3$, $\phi_0 = 0.5$, $(\phi_1, \phi_2, \phi_3) = (0.08, 0.06, 0.05)$, $\phi_4 = 0.06$, $\lambda_0 = 6$, $(\lambda_1, \lambda_2, \lambda_3) = (4, 2, 3)$.

The sample size n is set to be 200 and 800. For each scenario and a given sample size n, we first generate $\{\mathbf{y}_j\}_{j=1}^n \overset{\text{iid}}{\sim} \text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$ and then use the EM algorithm specified by (3.4)–(3.5) to calculate the MLEs of the parameters. With the MLEs, by generating $G = 500$ bootstrap samples we obtain the 95% bootstrap CI specified by (3.10) for each parameter based on these bootstrap replications. Independently repeat this process 500 times. Finally, based on 500 repetitions, the resultant mean of the MLEs (denoted by MLE), the mean squared error (denoted by MSE, equals to the sum of the variance and the squared bias of the estimator) of the estimators, the coverage probability (denoted by CP) and the average width (denoted by Width) of bootstrap CIs under each parameter configuration are reported in Tables 1-4, respectively.

The results reveal that under different parameter configurations, all MLEs of parameters are close to their true values and the corresponding coverage probabilities of the interval estimators are quite satisfactory for both small and large sample size situations. More specifically, as the sample size increases, the MLEs are more accurate since the differences between estimated values and their true values become smaller and the corresponding MSEs also drop significantly. The interval estimators are more precise as the average widths become more narrow when the sample size is increased.

Table 1: Simulation results on accuracy of MLEs and interval estimators for Case I

		$n = 200$				$n = 800$			
Parameter	True	MLE	MSE	CP	Width	MLE	MSE	CP	Width
ϕ_0	0.3	0.299473	0.001182	0.932	0.126431	0.300089	0.000249	0.946	0.063291
ϕ_1	0.2	0.200015	0.000867	0.934	0.110085	0.199680	0.000209	0.938	0.055277
ϕ_2	0.2	0.200885	0.000864	0.936	0.110769	0.199475	0.000200	0.946	0.055263
ϕ_3	0.1	0.100587	0.000459	0.936	0.083341	0.098490	0.000108	0.942	0.041514
λ_0	2	2.077578	0.291295	0.946	2.005189	2.013730	0.082810	0.938	1.085131
λ_1	2	1.906087	0.286596	0.928	1.999224	1.972283	0.081504	0.934	1.103672
λ_2	3	2.908938	0.326888	0.944	2.116052	2.988352	0.090192	0.944	1.153983

Table 2: Simulation results on accuracy of MLEs and interval estimators for Case II

		$n = 200$				$n = 800$			
Parameter	True	MLE	MSE	CP	Width	MLE	MSE	CP	Width
ϕ_0	0.65	0.648977	0.001066	0.942	0.131677	0.649747	0.000298	0.930	0.066018
ϕ_1	0.05	0.051314	0.000235	0.930	0.059572	0.050105	0.000063	0.946	0.030122
ϕ_2	0.08	0.080211	0.000339	0.940	0.074213	0.079644	0.000084	0.940	0.037331
ϕ_3	0.07	0.069743	0.000345	0.926	0.069246	0.069702	0.000078	0.948	0.035186
λ_0	6	6.033782	0.709480	0.940	3.016558	6.008797	0.135546	0.952	1.496788
λ_1	2	1.982282	0.505431	0.912	2.554871	1.984519	0.096169	0.954	1.301682
λ_2	3	3.000144	0.554380	0.934	2.669882	2.970556	0.103760	0.954	1.350646

5.2 Performance of the LRT

In Section 3.3, the LRT is developed for testing H_0: $\phi_{m+1} = 0$ in (3.11) and H_0: $\lambda_0 = 0$ in (3.13). To evaluate the performance of the proposed LRT, we calculate the levels and powers for different sample sizes via simulations. We only consider $m = 2$ with sample sizes set to be $n = 100(50)500$.

Table 3: Simulation results on accuracy of MLEs and interval estimators or Case III

		$n = 200$				$n = 800$			
Parameter	True	MLE	MSE	CP	Width	MLE	MSE	CP	Width
ϕ_0	0.3	0.298799	0.001074	0.942	0.126305	0.300315	0.000247	0.962	0.063417
ϕ_1	0.2	0.201679	0.000868	0.938	0.110736	0.200352	0.000190	0.960	0.055456
ϕ_2	0.1	0.099344	0.000444	0.942	0.081668	0.099730	0.000108	0.950	0.041447
ϕ_3	0.1	0.100098	0.000502	0.932	0.082638	0.099765	0.000113	0.946	0.041415
ϕ_4	0.1	0.100359	0.000465	0.944	0.082945	0.100536	0.000118	0.932	0.041820
λ_0	1	1.028665	0.087442	0.918	1.066990	1.008615	0.018221	0.946	0.536210
λ_1	2	1.976059	0.109587	0.940	1.255372	2.009634	0.023395	0.956	0.631360
λ_2	1	0.978098	0.084634	0.910	1.060033	0.993418	0.019133	0.938	0.539105
λ_3	3	2.932917	0.145493	0.934	1.413747	2.992506	0.034368	0.952	0.711005

Table 4: Simulation results on accuracy of MLEs and interval estimators for Case IV

		$n = 200$				$n = 800$			
Parameter	True	MLE	MSE	CP	Width	MLE	MSE	CP	Width
ϕ_0	0.5	0.502220	0.001212	0.960	0.137620	0.501070	0.000307	0.940	0.069106
ϕ_1	0.08	0.079400	0.000387	0.912	0.073645	0.079600	0.000097	0.932	0.037270
ϕ_2	0.06	0.060160	0.000295	0.940	0.064190	0.060487	0.000071	0.938	0.032821
ϕ_3	0.05	0.048940	0.000225	0.932	0.058270	0.049827	0.000054	0.944	0.029955
ϕ_4	0.06	0.060277	0.000278	0.930	0.064636	0.059833	0.000072	0.948	0.032663
λ_0	6	6.033335	0.268890	0.954	1.992526	6.012531	0.069173	0.948	0.985235
λ_1	4	3.964745	0.211920	0.934	1.804585	3.974121	0.052280	0.946	0.901372
λ_2	2	1.965569	0.175774	0.928	1.615118	1.985290	0.041220	0.958	0.810607
λ_3	3	2.958785	0.199463	0.940	1.712874	2.987533	0.047092	0.944	0.858310

First, we investigate the type I error rates (with H_0: $\phi_{m+1} = 0$) and powers (with H_1: $\phi_{m+1} > 0$), where the values of ϕ_{m+1} in H_1 are chosen to be 0.01, 0.05, 0.1. For a given combination of $\big(n, \phi_0 = 0.3, \boldsymbol{\phi}^\top = (0.2, 0.2), \phi_{m+1}, \lambda_0 = 2, \boldsymbol{\lambda}^\top = (2, 3)\big)$, we first generate $\mathbf{y}_1^{(l)}, \ldots, \mathbf{y}_n^{(l)} \overset{\text{iid}}{\sim} \text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$ for $l = 1, \ldots, L$ ($L = 500$). For each group of samples $\{\mathbf{y}_j^{(l)}\}_{j=1}^n$, we conduct the testing hypothesis. Let r_1 denote the number of rejecting the null hypothesis H_0: $\phi_{m+1} = 0$ by the test statistic T_1 given by (3.12). Then the empirical level can be estimated by r_1/L with $\phi_{m+1} = 0$ and the power of the test statistic T_1 can be estimated by r_1/L with $\phi_{m+1} > 0$.

Next, we investigate the type I error rates (with H_0: $\lambda_0 = 0$) and powers (with H_1: $\lambda_0 > 0$), where the values of λ_0 in H_1 are chosen to be 1, 3, 5. For a given combination of $\big(n, \phi_0 = 0.3, \boldsymbol{\phi}^\top = (0.2, 0.2), \phi_{m+1} = 0.1, \lambda_0, \boldsymbol{\lambda}^\top = (2, 3)\big)$, we first generate $\mathbf{y}_1^{(l)}, \ldots, \mathbf{y}_n^{(l)} \overset{\text{iid}}{\sim} \text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$ for $l = 1, \ldots, L$ ($L = 500$). For each group of samples $\{\mathbf{y}_j^{(l)}\}_{j=1}^n$, we conduct the testing hypothesis. Let r_2 denote the number of rejecting the null hypothesis H_0: $\lambda_0 = 0$ by the test statistic T_2 given by (3.14). Then the empirical level can be estimated by r_2/L with $\lambda_0 = 0$ and the power of the test statistic T_2 can be estimated by r_2/L with $\lambda_0 > 0$.

The empirical levels/powers of the LRT statistics T_1 and T_2 are summarized in Table 5. Figure 1 displays the type I error rates and powers of the LRT in testing H_0: $\phi_{m+1} = 0$ against H_1: $\phi_{m+1} > 0$ with three different values of $\phi_{m+1} > 0$ for various sample sizes. Figure 2 displays the type I error rates and the powers of the LRT in testing H_0: $\lambda_0 = 0$ against H_1: $\lambda_0 > 0$ with three different values of $\lambda_0 > 0$ for various sample sizes. From both of the two figures, we can see that the lines for levels of LRT in two tests fluctuate near the line of $\alpha = 0.05$, indicating that they perform well in controlling the type I error rates around the pre-chosen nominal level. Besides, the LRT in all of six scenarios tend to be more powerful as the sample size n turns larger.

[Insert Figures 1 and 2 here]

6. Applications

6.1 Health care utilization data

Cameron & Trivedi (2013) reported data concerning the demand for Health Care in Australia which refers to the Australian Health survey for 1977-1978. Let Y_1 denote the number of consultations with a doctor or a specialist and Y_2 denote the total number of prescribed medications used in past two days. The data are given in Table 6.

Table 5: Empirical levels/powers of the LRT statistics T_1 and T_2 based on 500 replications

Sample size (n)	Empirical level	Empirical power ϕ_3			Empirical level	Empirical power λ_0		
		0.01	0.05	0.10		1	3	5
100	0.038	0.326	0.970	1.000	0.061	0.396	0.840	0.962
150	0.040	0.512	0.988	1.000	0.050	0.504	0.946	0.994
200	0.048	0.532	0.996	1.000	0.048	0.570	0.982	1.000
250	0.056	0.660	1.000	1.000	0.054	0.600	0.994	1.000
300	0.036	0.694	1.000	1.000	0.050	0.674	1.000	1.000
350	0.030	0.742	1.000	1.000	0.052	0.712	1.000	1.000
400	0.048	0.812	1.000	1.000	0.046	0.828	1.000	1.000
450	0.060	0.842	1.000	1.000	0.050	0.844	1.000	1.000
500	0.036	0.892	1.000	1.000	0.048	0.892	1.000	1.000

Table 6: Cross tabulation of the health care utilization data in the Australian Health Survey for 1977-1978 (Cameron & Trivedi, 2013)

$Y_1 \backslash Y_2$	0	1	2	3	4	5	6	7	8	Total
0	2789	726	307	171	76	32	16	15	9	4141
1	224	212	149	85	50	35	13	5	9	782
2	49	34	38	11	23	7	5	3	4	174
3	8	10	6	2	1	1	2	0	0	30
4	8	8	2	2	3	1	0	0	0	24
5	3	3	2	0	1	0	0	0	0	9
6	2	0	1	3	1	2	2	0	1	12
7	1	0	3	2	1	2	1	0	2	12
8	1	1	1	0	1	0	1	0	0	5
9	0	0	0	0	0	0	0	0	1	1
Total	3085	994	509	276	157	80	40	23	26	5190

6.1.1 Likelihood-based inferences without covariates

Through some endeavor of trying the models proposed in Section 2, the model $\text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; 0, \boldsymbol{\lambda})$ works well in fitting this data. The procedure of model selection is listed in Table

9. To find the MLEs of $(\phi_0, \phi_1, \phi_2, \phi_3, \lambda_1, \lambda_2)$, we choose $(\phi_0^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \phi_3^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}) = (0.2, 0.1, 0.1, 0.1, 2, 2)$ as their initial values. The MLEs converged to the values shown in the second column of Table 7 in 31 iterations for the Fisher scoring algorithm (3.7) and in 72 iterations for the EM algorithm (3.4)–(3.5), while the Newton–Raphson method is not available because the observed information matrix is nearly singular. The standard errors of the estimators are given in the third column and 95% asymptotic Wald CIs (specified by (3.8)) of the six parameters are listed in the fourth column of Table 7. With $G = 6,000$ bootstrap replications, the corresponding standard deviations and 95% bootstrap CIs are presented in last two columns of Table 7.

Table 7: MLEs and CIs of parameters for the Australian health survey data without covariates

Parameter	MLE	std$^{\mathrm{F}}$	95% Wald CI	std$^{\mathrm{B}}$	95% bootstrap CI
ϕ_0	0.5214	0.0081	[0.5056, 0.5372]	0.0072	[0.5073, 0.5354]
ϕ_1	0.0307	0.0030	[0.0249, 0.0366]	0.0030	[0.0250, 0.0365]
ϕ_2	0.1039	0.0061	[0.0921, 0.1158]	0.0055	[0.0931, 0.1145]
ϕ_3	0.0128	0.0031	[0.0067, 0.0189]	0.0031	[0.0067, 0.0188]
λ_1	0.7798	0.0236	[0.7336, 0.8260]	0.0237	[0.7341, 0.8258]
λ_2	2.2526	0.0494	[2.1558, 2.3493]	0.0499	[2.1539, 2.3498]

std$^{\mathrm{F}}$: Square roots of the diagonal elements of the inverse Fisher information matrix, c.f. (3.6); std$^{\mathrm{B}}$: The sample standard deviation of the bootstrap samples, c.f. (3.9); bootstrap CI: c.f. (3.10).

6.1.2 Bayesian methods

In the setting of Bayesian analysis, we adopt Dirichlet$\,(1, 1, 1, 1, 1)$ as the prior distribution of $(\phi_0, \phi_1, \phi_2, \phi_3, \phi_4)^\top$ and independent Gamma$\,(1, 1)$ as the prior distributions of both λ_1 and λ_2. Using $(\phi_0^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \phi_3^{(0)}) = (0.2, 0.1, 0.1, 0.1)$ and $(\lambda_1^{(0)}, \lambda_2^{(0)}) = (2, 2)$ as the initial values, the EM algorithm specified by (4.3) converged to the posterior modes in 68 iterations which are presented in the second column of Table 8.

To calculate the Bayesian credible intervals of $(\phi_0, \phi_1, \phi_2, \phi_3, \phi_4, \lambda_1, \lambda_2)$, we use the DA algorithm to generate $L = 60,000$ posterior samples for each of these parameters based on

(4.1) and (4.2). By discarding the first half of the samples, we can calculate the posterior means, the posterior standard deviations and the 95% Bayesian credible intervals of them, which are given in Table 8.

Table 8: Posterior estimates of parameters for the Australian health survey data

Parameter	Posterior mode	Posterior mean	Posterior std	95% Bayesian credible interval
ϕ_0	0.5213	0.5193	0.0073	[0.5049, 0.5336]
ϕ_1	0.0307	0.0296	0.0030	[0.0239, 0.0356]
ϕ_2	0.1038	0.1004	0.0055	[0.0898, 0.1111]
ϕ_3	0.0128	0.0106	0.0029	[0.0051, 0.0166]
ϕ_4	0.3314	0.3401	0.0094	[0.3220, 0.3590]
λ_1	0.7790	0.7698	0.0234	[0.7252, 0.8164]
λ_2	2.2498	2.2092	0.0478	[2.1164, 2.3037]

6.1.3 Model selection and comparison

In model selection, we begin with the full model $\text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$, but it does not converge. To remove insignificant parameters step by step in the model, we start with models of $\lambda_0 = 0$ and $\phi_3 = 0$, respectively. Based on the LRT results in Table 9, the null hypothesis H_0: $\lambda_0 = 0$ cannot be rejected, the null hypothesis H_0: $\phi_3 = 0$ should be rejected at 5% significance level, and no parameter can be removed any more, so we select the ZOIP $(\phi_0, \phi_1, \phi_2, \phi_3; 0, \lambda_1, \lambda_2)$ model.

Table 9: Likelihood ratio test in model selection

Null hypothesis	Alternative model	LRT statistic	p-value
H_0: $\lambda_0 = 0$	ZOIP $(\phi_0, \phi_1, \phi_2, 0; \lambda_0, \lambda_1, \lambda_2)$	1.1298	0.1439
H_0: $\phi_3 = 0$	ZOIP $(\phi_0, \phi_1, \phi_2, \phi_3; 0, \lambda_1, \lambda_2)$	18.7925	< 0.001
H_0: $\phi_1 = 0$	ZOIP $(\phi_0, \phi_1, \phi_2, \phi_3; 0, \lambda_1, \lambda_2)$	152.2099	< 0.001
H_0: $\phi_1 = \phi_2 = 0$	ZOIP $(\phi_0, \phi_1, \phi_2, \phi_3; 0, \lambda_1, \lambda_2)$	506.6042	< 0.001
H_0: $\phi_1 = \phi_2 = \phi_3 = 0$	ZOIP $(\phi_0, \phi_1, \phi_2, \phi_3; 0, \lambda_1, \lambda_2)$	398.2568	< 0.001

We choose the Akaike information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978) to compare models. Karlis & Ntzoufras (2005) used the *bivariate Poisson* (BP) and *diagonal inflated bivariate Poisson* (DIBP) models with covariates to fit the data set. To illustrate the fit of models, we just concentrate on the original models without covariates. Liu & Tian (2015) proposed a new Type I multivariate ZIP distribution to fit the data. The MLEs of parameters in BP model are estimated by $\hat{\lambda}_0 = 0.1256, \hat{\lambda}_1 = 0.1761, \hat{\lambda}_2 = 0.7370$. The fitted DIBP model led to zero-inflated model with only (0,0) inflated and the MLEs of parameters are estimated by $\hat{\phi}_0 = 0.4763, \hat{\lambda}_1 = 0.5017, \hat{\lambda}_2 = 1.5727, \hat{\lambda}_0 = 0.0745$. The MLEs of parameters in Type I ZIP model are $\hat{\phi} = 0.4830, \hat{\lambda}_1 = 0.5836, \hat{\lambda}_2 = 1.6685$. The values of AIC and BIC for the BP model, DIBP model, Type I ZIP model and bivariate ZOIP model are summarized in Table 10. The bivariate ZOIP model is selected by both AIC and BIC.

Table 10: Comparison by AIC and BIC of the four models

	Criterion	
Model	AIC	BIC
BP model	22542.71	22562.38
DIBP model	20529.92	20556.14
Type I bivariate ZIP model	20565.82	20585.48
Bivariate ZOIP model	**20173.56**	**20212.89**

BP: see Karlis & Ntzoufras (2005); DIBP: see Karlis & Ntzoufras (2005); Type I bivariate ZIP: see Liu & Tian (2015).

6.1.4 Marginal analysis

The sample correlation coefficient in the health survey data is $r = 0.307779$. By performing the correlation test on the correlation coefficient between Y_1 and Y_2, the corresponding p-value is far less than 0.05, indicating a positive correlation between Y_1 and Y_2. Therefore, it is not appropriate to fit the data by two independent ZOIP distributions. The bivariate ZOIP distribution gives an estimated value $\hat{\rho} = 0.388162$.

To evaluate the performance of the proposed model from the view of marginal fitting, we compare the theoretical marginal distribution with univariate ZOIP distribution. From

Table 7, we know that $(Y_1, Y_2)^\top$ follows the bivariate ZOIP distribution with $\hat{\phi}_0 = 0.5214$, $(\hat{\phi}_1, \hat{\phi}_2) = (0.0307, 0.1039)$, $\hat{\phi}_3 = 0.0128$ and $(\hat{\lambda}_1, \hat{\lambda}_2) = (0.7798, 2.2526)$. According to (2.5), we have the marginal distribution for each component of the multivariate ZOIP is $Y_i \sim$ ZOIP $(\phi_0 + \sum_{k=1,k\neq i}^{m} \phi_k, \phi_i + \phi_{m+1}; \lambda_0 + \lambda_i)$, then the corresponding marginal distribution of Y_1 and Y_2 are estimated to be ZOIP $(\hat{\phi}_0 = 0.6253, \hat{\phi}_1 = 0.0435; \hat{\lambda}_1 = 0.7798)$ and ZOIP $(\hat{\phi}_0 = 0.5521, \hat{\phi}_1 = 0.1167; \hat{\lambda}_2 = 2.2526)$, respectively. If we fit Y_1 and Y_2 with univariate ZOIP distributions, the estimates are given as ZOIP $(\hat{\phi}_0^{\text{M}} = 0.7869, \hat{\phi}_1^{\text{M}} = 0.1282; \hat{\lambda}_1^{\text{M}} = 2.0425)$ and ZOIP $(\hat{\phi}_0^{\text{M}} = 0.5651, \hat{\phi}_1^{\text{M}} = 0.1221, \hat{\lambda}_2^{\text{M}} = 2.3676)$. Both results show that Y_1 and Y_2 follow the ZOIP distributions with different zero inflation, one inflation and Poisson mean parameters.

6.1.5 Likelihood-based inferences with covariates

We choose the following covariates. Let V_1 denote the gender, where $V_1 = 1$ if female and $V_1 = 0$ if male. Let V_2 denote the age in years divided by 100. let V_3 denote the annual income in Australian dollars divided by 1000, which measured as midpoint of coded ranges: 200—1000, 1001—2000, 2001—3000, 3001—4000, 4001—5000, 5001—6000, 6001—7000, 7001—8000, 8001—10000, 10001—12000, 12001—14000, with 14001+ treated as 15000. Let $\mathbf{w} = (1, V_1)^\top$ and $\mathbf{x} = (1, V_1, V_2, V_3)^\top$. By adopting the model of special case 1 that incorporated with covariates to fit the data, the MLEs and corresponding confidence intervals of the regression coefficients for parameters are listed in Table 11.

6.2 Automobile third party liability insurance data

The data are claims of a large automobile portfolio in France which including 181038 liability policies in 1989 provided by Vernic (1997). The corresponding claim frequencies were divided into material damage only (type I) denoted by Y_1 and bodily injury (type II) claims denoted by Y_2, as shown in Table 12. Note that the three categories (0,0), (1,0) and (0,1) have comparative frequencies than the other cells, our model should be considered in the first place.

Table 11: MLEs and estimated errors of parameters for the Australian health survey data with covariates

Parameter	Coefficients	MLE	std_1^{F}	std_2^{F}
$\log(\phi_0/\phi_4)$	Constant	0.819103	0.090515	0.045131
	Sex (Female)	−1.482019	0.108103	0.062272
$\log(\phi_1/\phi_4)$	Constant	−2.702688	0.303373	0.150015
	Sex (Female)	−1.661920	0.545856	0.268987
$\log(\phi_2/\phi_4)$	Constant	−1.786845	0.209480	0.099266
	Sex (Female)	−0.033092	0.238233	0.119766
$\log(\phi_3/\phi_4)$	Constant	−4.261099	1.092479	0.318782
	Sex (Female)	−2.154864	4.824744	0.696513
λ_1	Constant	−0.249535	0.123691	0.096908
	Sex (Female)	−0.308786	0.071620	0.056172
	Age	0.365522	0.150968	0.130522
	Income	−0.294971	0.090352	0.082081
λ_2	Constant	−0.680362	0.096899	0.072440
	Sex (Female)	0.088192	0.055689	0.037419
	Age	2.427721	0.107850	0.090117
	Income	−0.099109	0.060352	0.053422

std_1^{F}: Square roots of the diagonal elements of $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}|Y_{\text{obs}})$, c.f. (3.15); std_2^{F}: Square roots of the diagonal elements of $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}|Y_{\text{com}})$, c.f. (3.16).

Table 12: Cross tabulation of the automobile third party liability insurance data (Vernic, 1997)

$Y_1 \backslash Y_2$	0	1	2 and above	Total
0	171345	918	2	172265
1	8273	73	0	8346
2	389	5	0	394
3	31	1	0	32
4 and above	1	0	0	1

6.2.1 Likelihood-based inferences

As the data are characterized by the first three highest frequencies locating at categories (0,0), (1,0) and (0,1), the model $\text{ZOIP}_m(\phi_0, \boldsymbol{\phi}, 0; 0, \boldsymbol{\lambda})$ is the most appropriate after some calculations and comparisons which are shown in Section 6.2.2. We choose $(\phi_0^{(0)}, \phi_1^{(0)}, \phi_2^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}) = (0.2, 0.1, 0.1, 1, 1)$ as their initial values and the MLEs converged to the results shown in the second column of Table 13 in 9 iterations for the Fisher scoring algorithm (3.7), while the EM algorithm (3.4)–(3.5) does not work well for the really slow speed of convergence and the Newton–Raphson method is not available either due to the singularity of the observed information matrix. The standard errors of the estimators and 95% asymptotic Wald CIs (specified by (3.8)) are given in the third and fourth columns of Table 13. With $G = 6,000$ bootstrap replications, the corresponding standard deviations and 95% bootstrap CIs are reported in last two columns of Table 13. Since the EM algorithm does not work well in parameters convergence, we do not consider the Bayesian methods in parameters estimation for this data set.

The sample correlation coefficient in the insurance data is $r = 0.011191$ and the correlation test suggests to reject the independency between Y_1 and Y_2. Thus, Y_1 and Y_2 has a positive and low correlation. While the estimated value given by the above model is $\hat{\rho} = 0.011022$, which is very close to that from the samples.

Table 13: MLEs and CIs of parameters $(\phi_0, \phi_1, \phi_2, \lambda_1, \lambda_2)$ for the automobile third party liability insurance data

Parameter	MLE	std$^{\mathrm{F}}$	95% Wald CI	std$^{\mathrm{B}}$	95% bootstrap CI
ϕ_0	0.8496	0.0316	[0.7877, 0.9115]	0.0386	[0.7476, 0.8927]
ϕ_1	0.0251	0.0036	[0.0181, 0.0322]	0.0039	[0.0158, 0.0308]
ϕ_2	0.0033	0.0004	[0.0025, 0.0041]	0.0004	[0.0023, 0.0040]
λ_1	0.2118	0.0329	[0.1473, 0.2763]	0.0331	[0.1498, 0.2792]
λ_2	0.0183	0.0034	[0.0116, 0.0250]	0.0034	[0.0122, 0.0255]

std$^{\mathrm{F}}$: Square roots of the diagonal elements of the inverse Fisher information matrix; std$^{\mathrm{B}}$: The sample standard deviation of the bootstrap samples, c.f. (3.9); bootstrap CI: c.f. (3.10).

6.2.2 Model selection and comparison

In model selection, we begin with the full model $\mathrm{ZOIP}_m(\phi_0, \boldsymbol{\phi}, \phi_{m+1}; \lambda_0, \boldsymbol{\lambda})$, but it does not converge. We first restrict λ_0 to be zero. Based on the LRT results in Table 14, the null hypothesis H_0: $\phi_3 = 0$ cannot be rejected at 5% significance level and no parameter can be removed any more, so we select the ZOIP $(\phi_0, \phi_1, \phi_2, 0; 0, \lambda_1, \lambda_2)$ model.

Table 14: Likelihood ratio test in model selection

Null hypothesis	Null model	Alternative model	LRT statistic	p-value
H_0: $\phi_3 = 0$	ZOIP $(\phi_0, \phi_1, \phi_2; \lambda_1, \lambda_2)$	ZOIP $(\phi_0, \phi_1, \phi_2, \phi_3; \lambda_1, \lambda_2)$	1.8555	0.0866
H_0: $\phi_1 = 0$	ZOIP $(\phi_0, \phi_2; \lambda_1, \lambda_2)$	ZOIP $(\phi_0, \phi_1, \phi_2; \lambda_1, \lambda_2)$	17.6799	< 0.001
H_0: $\phi_1 = \phi_2 = 0$	ZOIP $(\phi_0; \lambda_1, \lambda_2)$	ZOIP $(\phi_0, \phi_1, \phi_2; \lambda_1, \lambda_2)$	23.2810	< 0.001

We evaluate models by AIC and BIC. For purpose of comparison, from Vernic (1997), *bivariate generalized Poisson distribution* (BGPD) was adopted to fit the data. The MLEs of parameters in BGPD model are estimated by $\hat{\lambda}_1 = 0.0495, \hat{\lambda}_2 = 0.0054, \hat{\lambda}_3 = 0.0002$ and $\hat{\theta}_1 = 0.0270, \hat{\theta}_2 = -0.0027, \hat{\theta}_3 = 0.0498$. The values of AIC and BIC are summarized in Table 15. As suggested from AIC and BIC, the bivariate ZOIP model gives a better fit.

Table 15: Comparison by AIC and BIC of the two models

	Criterion	
Model	AIC	BIC
BGPD model	86309.19	86369.83
Bivariate ZOIP	**86286.63**	**86337.16**

BGPD: see Vernic (1997).

7. Discussion

This paper extends the univariate zero-and-one inflated Poisson distribution to a multivariate version by considering inflation at several categories simultaneously. This new multivariate ZOIP distribution has a flexible dependency structure; i.e., the correlation coefficient between any two random components could be either positive or negative depending on the values of the parameters, as shown in (2.6). The marginal distributions are not necessarily identical with each other; i.e., each random component follows a ZOIP distribution with different zero inflation, one inflation and Poisson mean parameters as shown in (2.5). The distributional theories are explored profoundly and statistical inference methods are provided explicitly.

The multivariate regression model with covariates is also investigated in Section 3.4 and the estimates of those regression coefficients are obtained through an EM algorithm embedded with the Newton–Raphson algorithm. However, in our real examples, the bootstrap method is not available for calculating the standard errors of the coefficient estimates due to the singularity of the observed information matrix. Instead, we calculate the observed information matrix using the method of Louis (1982) by subtracting the missing information from the complete information. Because of the complexity of Louis's method, sometimes we may calculate the complete information and use it to approximate the observed information.

Appendix A: Some technical derivations

A.1 Derivation of the joint probability mass function (2.3)

If $\boldsymbol{y} = \mathbf{0}_m$, we have

$$\begin{aligned}\Pr(\mathbf{y} = \mathbf{0}_m) &= \Pr(Z_0 = 1) + \Pr(Z_{m+2} = 1,\, X_1 = 0, \dots, X_m = 0) \\ &\overset{(2.1)}{=} \phi_0 + \phi_{m+2}\mathrm{e}^{-\lambda_0-\lambda_+}. \end{aligned} \tag{A.1}$$

If $\boldsymbol{y} = \boldsymbol{e}_m^{(i)}$, we obtain

$$\begin{aligned}\Pr(\mathbf{y} = \boldsymbol{e}_m^{(i)}) &= \Pr(Z_i = 1) + \Pr(Z_{m+2} = 1,\, X_1 = 0, \dots, X_i = 1, \dots, X_m = 0) \\ &\overset{(2.1)}{=} \phi_i + \phi_{m+2}\lambda_i\mathrm{e}^{-\lambda_0-\lambda_+}, \qquad i = 1, \dots, m. \end{aligned} \tag{A.2}$$

If $\boldsymbol{y} = \mathbf{1}_m$, we have

$$\begin{aligned}\Pr(\mathbf{y} = \mathbf{1}_m) &= \Pr(Z_{m+1} = 1) + \Pr(Z_{m+2} = 1,\, X_1 = 1, \dots, X_m = 1) \\ &\overset{(2.1)}{=} \phi_{m+1} + \phi_{m+2}\left(\lambda_0 + \textstyle\prod_{i=1}^m \lambda_i\right)\mathrm{e}^{-\lambda_0-\lambda_+}. \end{aligned} \tag{A.3}$$

If $\boldsymbol{y} \notin \left\{\mathbf{0}_m, \boldsymbol{e}_m^{(1)}, \dots, \boldsymbol{e}_m^{(m)}, \mathbf{1}_m\right\}$, then we have

$$\begin{aligned}\Pr(\mathbf{y} = \boldsymbol{y}) &= \Pr(Z_{m+2} = 1,\, X_1 = y_1, \dots, X_m = y_m) \\ &= \phi_{m+2}\mathrm{e}^{-\lambda_0-\lambda_+}\sum_{k=0}^{\min(\boldsymbol{y})}\frac{\lambda_0^k}{k!}\prod_{i=1}^m \frac{\lambda_i^{y_i-k}}{(y_i-k)!}. \end{aligned} \tag{A.4}$$

By combining (A.1)–(A.4), we obtain (2.3).

A.2 Derivation of the conditional expectation (3.5)

To derive the fourth formula in (3.5), we have if $\boldsymbol{y} \notin \mathcal{Y}_{01}$, then

$$\Pr(X_0^* = l | \mathbf{y} = \boldsymbol{y} \notin \mathcal{Y}_{01}) = \frac{\Pr(X_0^* = l, \mathbf{y} = \boldsymbol{y})}{\Pr(\mathbf{y} = \boldsymbol{y})}$$

$$= \frac{\Pr(Z_{m+2} = 1, X_0^* = l, X_1^* = y_1 - l, \dots, X_m^* = y_m - l)}{\Pr(Z_{m+2} = 1, \mathbf{x} = \boldsymbol{y})}$$

$$= \frac{\dfrac{\lambda_0^l}{l!} \displaystyle\prod_{i=1}^{m} \frac{\lambda_i^{y_i - l}}{(y_i - l)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{y})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{y_i - k}}{(y_i - k)!}} \mathrel{\hat{=}} q_l(\boldsymbol{y}, \lambda_0, \boldsymbol{\lambda}), \tag{A.5}$$

for $l = 0, 1, \dots, \min(\boldsymbol{y})$, which implying[1]

$$X_0^* | (\mathbf{y} = \boldsymbol{y} \notin \mathcal{Y}_{01}) \sim \text{Finite}\,(l, q_l(\boldsymbol{y}, \lambda_0, \boldsymbol{\lambda}); \quad l = 0, 1, \dots, \min(\boldsymbol{y})).$$

Appendix B: Definition of Type I multivariate ZIP distribution

Definition 2 An m-dimensional discrete random vector $\mathbf{y} = (Y_1, \dots, Y_m)^\top$ is said to have a Type I multivariate zero-inflated Poisson distribution (Liu & Tian, 2015) with parameters $\phi \in [0, 1)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}_+^m$ if

$$\mathbf{y} \stackrel{\text{d}}{=} Z\,\mathbf{x} = \begin{cases} \mathbf{0}, & \text{with probability } \phi, \\ \mathbf{x}, & \text{with probability } 1 - \phi, \end{cases}$$

where $Z \sim \text{Bernoulli}\,(1 - \phi)$, $\mathbf{x} = (X_1, \dots, X_m)^\top$, $X_i \sim \text{Poisson}\,(\lambda_i)$ for $i = 1, \dots, m$, and $(Z, X_1, \dots, X_m)$ are mutually independent. We will write $\mathbf{y} \sim \text{ZIP}_m^{(\text{I})}(\phi; \boldsymbol{\lambda})$. ¶

Appendix C: Definition of Type II multivariate ZIP distribution

Definition 3 An m-dimensional discrete random vector $\mathbf{y} = (Y_1, \dots, Y_m)^\top$ is said to have a Type II multivariate zero-inflated Poisson distribution with parameters $\phi \in [0, 1)$, $\lambda_0 \geqslant 0$

[1]A discrete random variable X is said to have the *general finite* distribution, denoted by $X \sim \text{Finite}\,(x_k, p_k;\ k = 0, 1, \dots, K)$, if $\Pr(X = x_k) = p_k \in [0, 1]$ and $\sum_{k=1}^{K} p_k = 1$.

and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$ if

$$\mathbf{y} \stackrel{\mathrm{d}}{=} Z\,\mathbf{x} = \begin{cases} \mathbf{0}, & \text{with probability } \phi, \\ \mathbf{x}, & \text{with probability } 1-\phi, \end{cases}$$

where $Z \sim \text{Bernoulli}\,(1-\phi)$, $\mathbf{x} = (X_1, \ldots, X_m)^\top \sim \text{MP}\,(\lambda_0, \boldsymbol{\lambda})$, $X_i = X_0^* + X_i^*$ for $i = 1, \ldots, m$, $\{X_i^*\}_{i=0}^m \sim \text{Poisson}\,(\lambda_i)$ and $Z \perp\!\!\!\perp \mathbf{x}$. We will write $\mathbf{y} \sim \text{ZIP}_m^{(\mathrm{II})}(\phi; \lambda_0, \boldsymbol{\lambda})$. ¶

Acknowledgments

Chi ZHANG's research was supported by National Natural Science Foundation of China (Grant no. 11801380) and Natural Science Foundation of SZU (Grant no. 2019069). Guo-Liang TIAN's research was fully supported by National Natural Science Foundation of China (Grant no. 11771199). Kam Chuen YUEN's research was supported by a Seed Fund for Basic Research of the University of Hong Kong. Qin WU's research was fully supported by National Natural Science Foundation of China (Grant no. 11401226).

References

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

Cameron, A. and Trivedi, P. (2013). *Regression Analysis of Count Data* (2nd Edition). Cambridge University Press, New York.

Carrivick, P.J.W., Lee, A.H. and Yau, K.K.W. (2003). Zero-inflated Poisson modeling to evaluate occupational safety interventions. *Safety Science* **41**(1), 53–63.

Deshmukh, S.R. and Kasture, M.S. (2002). Bivariate distribution with truncated Poisson marginal distributions. *Communications in Statistics—Theory and Methods* **31**(4), 527–534.

Diallo, A.O., Diop, A. and Dupuy, J.F. (2018). Analysis of multinomial counts with joint zero-inflation, with an application to health economics. *Journal of Statistical Planning and Inference* **194**, 85–105.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions.* Wiley, New York.

Karlis, D. and Ntzoufras, I (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software* **14**(10), 1–36.

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

Lee, A.H., Stevenson, M.R., Wang, K. and Yau, K.K.W. (2002). Modeling young driver motor vehicle crashes: data with extra zeros. *Accident Analysis and Prevention* **34**, 515–521.

Li, C.S., Lu, J.C., Park, J., Kim, K., Brinkley, P.A. and Peterson, J.P. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics* **41**(1), 29–38.

Liu, W., Tang, Y. and Xu, A. (2018). A zero-and-one inflated Poisson model and its application. *Statistics and Its Interface* **11**(2), 339–351.

Liu, Y. and Tian, G.L. (2015). Type I multivariate zero-inflated Poisson distribution with applications. *Computational Statistics & Data Analysis* **83**, 200–222.

Liu, Y., Tian, G.L., Tang, M.L. and Yuen, K.C. (2018). A new multivariate zero-adjusted Poisson model with applications to biomedicine. *Biometrical Journal.* Epub ahead of print 25 May 2018. DOI: 10.1002/bimj.201700144.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 226–233.

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.

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

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

Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American*

Statistical Association **82**(398), 605–610.

Tang, Y., Liu, W. and Xu, A. (2017). Statistical inference for zero-and-one-inflated Poisson models. *Statistical Theory and Related Fields* **1**(2), 216–226.

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.

Tian, G.L., Liu, Y., Tang, M.L. and Jiang, X.J. (2018). Type I multivariate zero-truncated/adjusted Poisson distributions with applications. *Journal of Computational and Applied Mathematics* **344**, 132–153.

Vernic, R. (1997). On the bivariate generalized Poisson distribution. *Astin Bulletin* **27**(1), 23–32.

Walhin, J.F. (2001). Bivariate ZIP models. *Biometrical Journal* **43**(2), 147–160.

Wang, K., Lee, A.H., Yau, K.K.W. and Carrivick, P.J.W. (2003). A bivariate zero-inflated Poisson regression model to analyze occupational injuries. *Accident Analysis & Prevention* **35**(4), 625–629.

Zhang, C., Tian, G.L. and Ng, K.W. (2016). Properties of the zero-and-one inflated Poisson distribution and likelihood-based inference methods. *Statistics and Its Interface* **9**(1), 11–32.

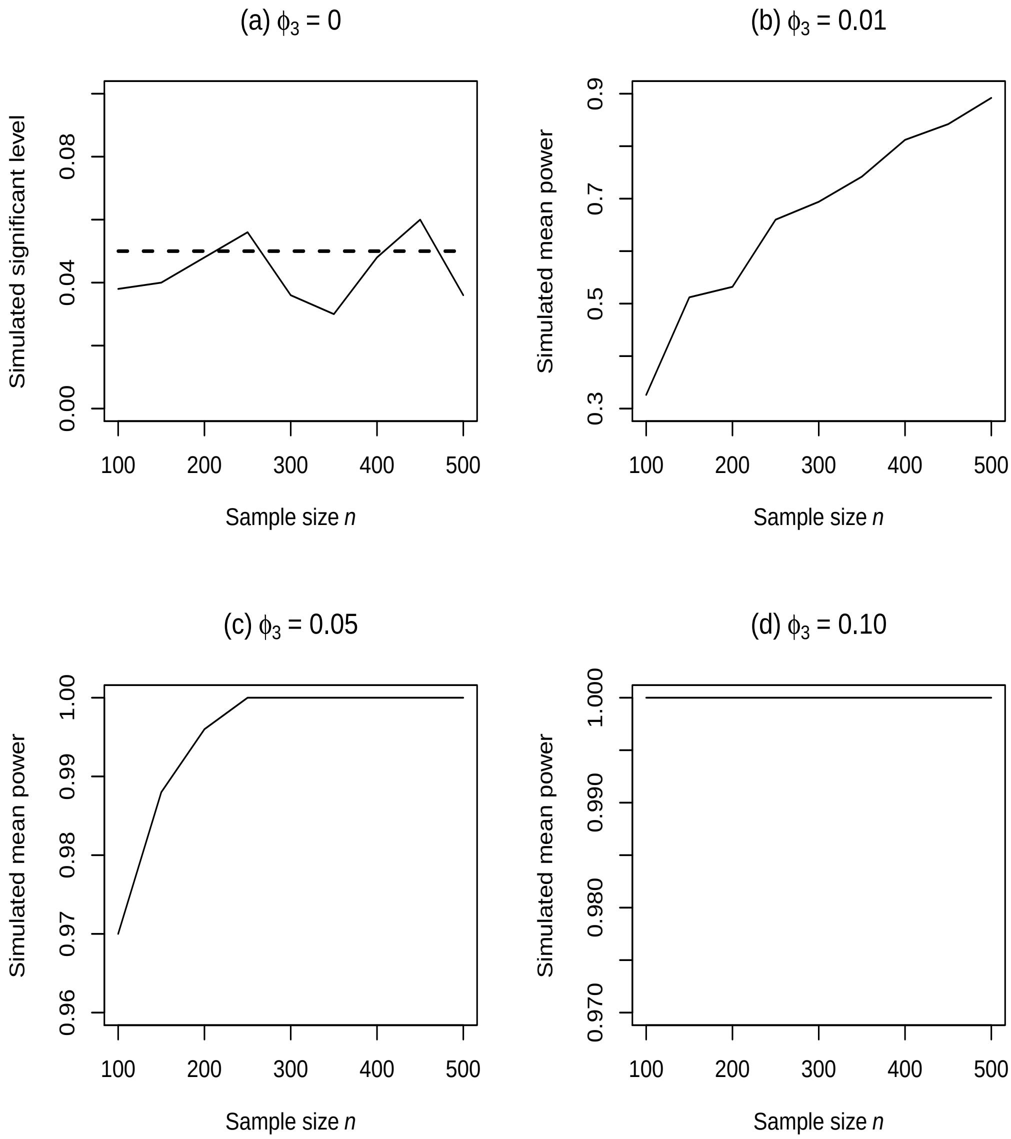

Figure 1 (a) The type I error rates for testing H_0: $\phi_{m+1} = 0$ against H_1: $\phi_{m+1} > 0$ in the multivariate ZOIP model and the dashed line is set as the predetermined significance level of $\alpha = 0.05$; (b) the powers when $\phi_{m+1} = 0.01$ in H_1; (c) the powers when $\phi_{m+1} = 0.05$ in H_1; (d) the powers when $\phi_{m+1} = 0.10$ in H_1.

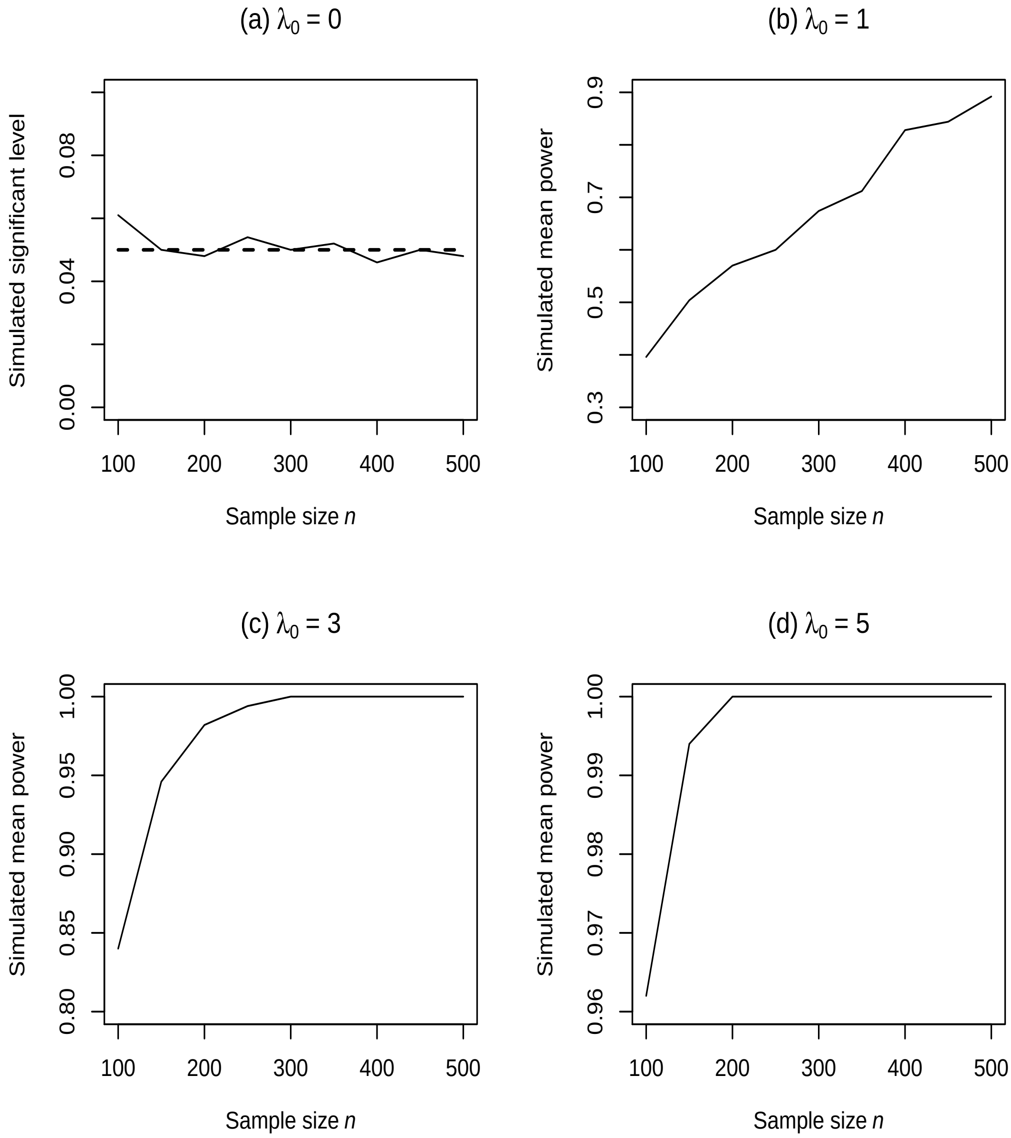

Figure 2 (a) The type I error rates for testing H_0: $\lambda_0 = 0$ against H_1: $\lambda_0 > 0$ in the multivariate ZOIP model and the dashed line is set as the predetermined significance level of $\alpha = 0.05$; (b) the powers when $\lambda_0 = 1$ in H_1; (c) the powers when $\lambda_0 = 3$ in H_1; (d) the powers when $\lambda_0 = 5$ in H_1.