A new multivariate zero-adjusted Poisson model with applications to biomedicine

Yin LIU[a], **Guo-Liang TIAN**[b,*], **Man-Lai TANG**[c] and **Kam Chuen YUEN**[d]

[a]*School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, Hubei Province, P. R. China*

[b]*Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, P. R. China*

[c]*Department of Mathematics and Statistics, School of Decision Sciences, Hang Seng Management College, Siu Lek Yuen, Shatin, N.T., Hong Kong, P. R. China*

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

*Corresponding author's email: tiangl@sustc.edu.cn
Biometrical Journal first revision, 30 NOV 2017

Abstract. Recently, although advances were made on modeling multivariate count data, existing models really has several limitations: (i) The multivariate Poisson log-normal model (Aitchison and Ho, 1989) cannot be used to fit multivariate count data with excess zero-vectors; (ii) The multivariate *zero-inflated Poisson* (ZIP) distribution (Li *et al.*, 1999) cannot be used to model zero-truncated/deflated count data and it is difficult to apply to high-dimensional cases; (iii) The Type I multivariate *zero-adjusted Poisson* (ZAP) distribution (Tian *et al.*, 2017) could only model multivariate count data with a special correlation structure for random components that are all positive or negative. In this paper, we first introduce a new multivariate ZAP distribution, based on a multivariate Poisson distribution, which allows the correlations between components with a more flexible dependency structure, i.e., some of the correlation coefficients could be positive while others could be negative. We then develop its important distributional properties, and provide efficient statistical inference methods for multivariate ZAP model with or without covariates. Two real data examples in biomedicine are used to illustrate the proposed methods.

Keywords: Data augmentation algorithm; Expectation–maximization algorithm; Hypothesis testing; Multivariate zero-adjusted Poisson; Stochastic representation.

1. Introduction

The univariate zero-truncated and zero-inflated models are well studied in the past decades. In literature, a class of univariate zero-truncated discrete models such as *zero-truncated Poisson* (ZTP) distribution (David and Johnson, 1952; Moore, 1952; Rider, 1953; Cohen, 1954, 1960a; Finney and Varley, 1955; Rao and Chakravarti, 1956; Irwin, 1959; Dahiya and Gross, 1973; Gurmu, 1991; Meng, 1997; Best *et al.*, 2007), zero-truncated binomial distribution (Finney, 1949; Rider, 1955), zero-truncated negative-binomial distribution (Rider, 1955; Sampford, 1955; Hartley, 1958; Grogger and Carson, 1991), zero-truncated generalized negative-binomial distribution (Gupta, 1974), zero-truncated generalized Poisson (Medhi, 1975; Consul, 1989), and intervened Poisson distribution (Shanmugam, 1985) were developed to model count data without zero value. On the other hand, the *zero-inflated Poisson* (ZIP) model (Cohen, 1960b; Singh, 1963; Martin and Katti, 1965; Johnson and Kotz, 1969; Goraski, 1977; Kemp, 1986; Mullahy, 1986; Lambert, 1992; Böhning et al., 1999; Cheung, 2002; Deng and Paul, 2000, 2005; Winkelmann, 2004; Min and Agresti, 2005; Min and Czado, 2010; Neelon *et al.*, 2010; Li, 2012), zero-inflated generalized Poisson model (Angers and Biswas, 2003; Famoye and Singh, 2006; Cui and Yang, 2009; Xie *et al.*, 2009), zero-inflated negative binomial model (Ridout *et al.*, 2001; Yau *et al.*, 2003; Bago d'Uva, 2006; Neelon *et al.*, 2010) were proposed to fit count data with extra zeros.

However, the univariate models are not appropriate for multivariate cases while multivariate count data frequently arise in clinical health care, survival analysis, industrial production, accident analysis & prevention, and other fields. In practice, such count data often exhibit some characteristics: (i) There is no category of zero-vectors. One example is the relative effectiveness of three air samplers 1, 2, 3 to detect pathogenic bacteria in "sterile" rooms, where the data set is collected as triplets of bacterial colony counts from the three samplers in each of 50 different sterile locations (Aitchison and Ho, 1989). The second example is about the health center visit study of California (Gurmu, 1997), which reports the number of the doctor office/clinic and health center visits during a period of

4 months and the number of children in a household. The third example is about the number of accidents in 24 roads of Athens for the period 1987–1991 (Karlis, 2003), where information was recorded when accidents occurred. (ii) There exist excess zero-vectors than usual. For instance, the number of consultations with a doctor or a specialist and the total number of prescribed medications used collected during a period of time (Cameron and Trivedi, 2013), for most of the people, are both 0 because some minor ailment can be cured by themselves and it is not necessary to visit a doctor or a specialist. It is quite a challenge for us to propose a *unified* statistical distribution to model such multivariate count data with a flexible dependency structure.

In the past decades, several bivariate *zero-truncated Poisson* (ZTP) distributions (Hamdan, 1972; Dahiya, 1977; Charalambides, 1984; Deshmukh and Kasture, 2002; Piperigou and papageorgiou, 2003; Jung *et al.*, 2007) were proposed to model such correlated paired count data. However, all these papers are based on the density form of the correlated bivariate Poisson distribution introduced by Campbell (1934) and it seems quite difficult to generalize them to multivariate ($m \geqslant 3$) zero-truncated versions. On the other hand, for the zero-inflated multivariate count data, although Li *et al.* (1999) first proposed a multivariate ZIP model as a mixture of $m + 2$ components of m-dimensional discrete distributions with mixing probabilities, it still has some limitations: (i) There are too many parameters to be estimated simultaneously; (ii) The complexity of this multivariate ZIP distribution leads to less helpful distribution properties and inefficient algorithms for estimation. Tian *et al.* (2017) proposed a so-called Type I multivariate *zero-adjusted Poisson* (ZAP) distribution and Liu and Tian (2015) considered the Type I multivariate ZIP that can be used to fit the multivariate count data without or with excess zero-vectors but all the Poisson components should be independent and it could only allow a special correlation structure among random components. Hall (2000) considered to model the multivariate zero-inflated count data by using specific random effects in the regression model so that the within-subject correlation and between-subject heterogeneity in repeated measures data can be accommodated. Hall and Zhang (2004) extended the work of Hall (2000) to zero-inflated clustered data. Lee *et al.* (2006) and Wang (2010) tried to model the correlated count data with excess zeros via the multi-level ZIP model. The longitudinal zero-inflated count data had received much attention in the last 20 years and can be fitted by the two-part model or the hurdle model. Alfò and Maruotti (2010) discussed a semi-parametric estimation method for the dynamic two-part models to handle the longitudinal zero-inflated count data, based on their previous works (Alfò and Trovato, 2004; Alfò and Aitkin, 2006; Alfò, Maruotti and Trovato, 2011), and presented a comparison with other well-established alternatives. Belloc *et al.* (2013) proposed an approximate conditional dynamic finite mixture hurdle model for panel count data with excess of zeros and endogenous initial conditions. Maruotti and Raponi (2014) described a mixed-effect hurdle model for zero-inflated longitudinal count data with a baseline effect as a different form of dependence in the model. The canonical link among longitudinal zero-inflated counts for each subject is described by the generalized linear regression model but the inherent correlation among the outcomes at each time point is not really examined.

Therefore, the first objective of this paper is to propose a new multivariate discrete distribution (which is called multivariate ZAP distribution) based on the traditional multivariate Poisson distribution to model dependent count data with a zero-adjusted observation vector. The multivariate ZTP distribution, multivariate *zero-deflated Poisson* (ZDP) distribution, multivariate Poisson distribution and multivariate ZIP distribution are included as its special cases. The most important feature for the proposed multivariate ZAP distribution is that it can be employed to model zero-truncated /zero-deflated /zero-inflated count vectors with a more flexible dependency structure, i.e., some of the correlation coefficients could be positive while others could be negative. Furthermore, to analyze the aforementioned count data, we stochastically represent the multivariate ZAP random variable $\mathbf{y}$ as a mixture of a ZTP random vector $\mathbf{w}$ and a zero-vector $\mathbf{0}$ as shown in (2.4). Based on this *stochastic representation* (SR), we could derive some important distribution properties for multivariate ZAP distribution, develop an *expectation–maximization* (EM) algorithm to calculate the MLEs and posterior modes of parameters of interest and provide the *data augmentation* (DA) algorithm in the Bayesian analysis. The second objective of this paper is to investigate the multivariate ZAP regression model, which expands the application of the proposed distribution to a certain extent.

The rest of the paper is organized as follows. In Section 2, we propose a new multivariate ZAP distribution. In Section 3, we develop likelihood-based methods, Bayesian methods and hypothesis testing approaches for the proposed distribution without covariates. In Section 4, we introduce the multivariate ZAP regression model. In Section 5, two real data examples in biomedicine are used to illustrate the proposed methods. Section 6 provides a discussion. The definitions of several useful distributions, some important distributional properties for the proposed multivariate ZAP distribution are given in the Appendices.

2. The multivariate ZAP distribution

To model dependent count data with a zero-adjusted observation vector, in this section we propose a new multivariate discrete distribution (which is called the multivariate ZAP distribution) with a flexible correlation structure based on the traditional multivariate Poisson distribution (see, Appendix A.1). For this purpose, we first introduce a so-called multivariate ZTP distribution.

2.1 The multivariate ZTP distribution

Definition 1 Let $\mathbf{x} \sim \text{MP}(\lambda_0, \lambda_1, \dots, \lambda_m)$. A discrete random vector $\mathbf{w} = (W_1, \dots, W_m)^\top$ is said to follow the multivariate ZTP distribution with the parameter $\lambda_0 \geqslant 0$ and the parameter vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{w} \sim \text{ZTP}(\lambda_0, \lambda_1, \dots, \lambda_m)$ or $\mathbf{w} \sim \text{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$, if

$$\mathbf{x} \stackrel{\text{d}}{=} U\,\mathbf{w} = \begin{cases} \mathbf{0}, & \text{with probability } \psi, \\ \mathbf{w}, & \text{with probability } 1-\psi, \end{cases} \tag{2.1}$$

where $U \sim \text{Bernoulli}\,(1-\psi)$ with $\psi = \text{e}^{-\lambda_0-\lambda_+}$, $\lambda_+ = \sum_{i=1}^m \lambda_i \,\hat{=}\, \|\boldsymbol{\lambda}\|_1$, and $U \perp\!\!\!\perp \mathbf{w}$. ¶

From (2.1) and (A.1), the joint *probability mass function* (pmf) of $\mathbf{w} \sim \mathrm{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$ is

$$
\begin{aligned}
f(\boldsymbol{w}; \lambda_0, \boldsymbol{\lambda}) &= \Pr(\mathbf{w} = \boldsymbol{w}) \stackrel{(2.1)}{=} \frac{\Pr(\mathbf{x} = \boldsymbol{w})}{\Pr(U = 1)} \\
&\stackrel{(A.1)}{=} \frac{1}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+}} \sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k} \mathrm{e}^{-\lambda_i}}{(w_i - k)!},
\end{aligned} \tag{2.2}
$$

where $\|\boldsymbol{w}\|_1 \neq 0$ and $\min(\boldsymbol{w}) \,\hat{=}\, \min(w_1, \ldots, w_m)$.

It is clear that the multivariate ZTP reduces to the Type I multivariate ZTP distribution (see, Appendix A.3) of Tian *et al.* (2017) if and only if $\lambda_0 = 0$. Let $\mathbf{w} \sim \mathrm{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$, then, we have $\Pr(\mathbf{w} = \mathbf{0}) = 0$ and

$$
\mathbf{w} \stackrel{\mathrm{d}}{=} \mathbf{x} | (\mathbf{x} \neq \mathbf{0}), \tag{2.3}
$$

where $\mathbf{x} \sim \mathrm{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$. The SR (2.3) can be used to generate the ZTP random vector $\mathbf{w}$ via the generation of the multivariate Poisson random vector $\mathbf{x}$, while the SR (2.1) is useful in deriving important distributional properties below and in developing a novel EM algorithm in Section 3.1. Moreover, besides coming from the missing zero vector, the correlation between any two components of the $\mathbf{w}$ may come from the common random variable $X_0^* \sim \mathrm{Poisson}(\lambda_0)$. Other important distributional properties of the multivariate ZTP distribution are provided in Appendix C.

2.2 The multivariate ZAP distribution

Next, in Appendix B, we define the univariate ZAP distribution. Motivated by (B.1), we naturally introduce its multivariate generalization on the basis of Definition 1 as follows.

Definition 2 A discrete random vector $\mathbf{y} = (Y_1, \ldots, Y_m)^\top$ is said to have the multivariate ZAP distribution with parameters $\varphi \in [0, 1)$, $\lambda_0 \geqslant 0$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{y} \sim \mathrm{ZAP}(\varphi\,; \lambda_0, \lambda_1, \ldots, \lambda_m)$ or $\mathbf{y} \sim \mathrm{ZAP}_m(\varphi\,; \lambda_0, \boldsymbol{\lambda})$, if

$$
\mathbf{y} \stackrel{\mathrm{d}}{=} Z'\mathbf{w} = \begin{cases} \mathbf{0}, & \text{with probability } \varphi, \\ \mathbf{w}, & \text{with probability } 1 - \varphi, \end{cases} \tag{2.4}
$$

where $Z' \sim \text{Bernoulli}(1-\varphi)$, $\mathbf{w} \sim \text{ZTP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$, and $Z' \perp\!\!\!\perp \mathbf{w}$. The random vector $\mathbf{w}$ is called the base vector of the $\mathbf{y}$. ¶

When $\lambda_0 = 0$, the multivariate ZAP is reduced to the Type I multivariate ZAP (Tian *et al.*, 2017). It is easy to show that the joint pmf of $\mathbf{y} \sim \text{ZAP}(\varphi\,; \lambda_0, \lambda_1, \ldots, \lambda_m)$ is

$$\Pr(\mathbf{y} = \boldsymbol{y}) = \varphi I(\boldsymbol{y} = \mathbf{0}) + \left[\frac{1-\varphi}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \sum_{k=0}^{\min(\boldsymbol{y})} \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{y_i-k}\mathrm{e}^{-\lambda_i}}{(y_i-k)!}\right] I(\boldsymbol{y} \neq \mathbf{0}). \qquad (2.5)$$

We consider the following special cases of (2.4) or (2.5):

(i) If $\varphi = 0$, then $\mathbf{y} \stackrel{\mathrm{d}}{=} \mathbf{w} \sim \text{ZTP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$, i.e., the multivariate ZTP distribution is a special member of the family of the Type multivariate ZAP distributions;

(ii) If $\varphi \in (0, \mathrm{e}^{-\lambda_0-\lambda_+})$, then $\mathbf{y}$ follows the multivariate ZDP distribution with parameters $(\varphi, \lambda_0, \boldsymbol{\lambda})$, denoted by $\mathbf{y} \sim \text{ZDP}(\varphi\,; \lambda_0, \lambda_1, \ldots, \lambda_m)$ or $\mathbf{y} \sim \text{ZDP}_m(\varphi\,; \lambda_0, \boldsymbol{\lambda})$;

(iii) If $\varphi = \mathrm{e}^{-\lambda_0-\lambda_+}$, then $\mathbf{y}$ follows the multivariate Poisson distribution with parameters $\lambda_0 \geqslant 0$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, i.e., $\mathbf{y} \sim \text{MP}_m(\lambda_0, \boldsymbol{\lambda})$;

(iv) If $\varphi \in (\mathrm{e}^{-\lambda_0-\lambda_+}, 1)$, then $\mathbf{y}$ follows the multivariate ZIP distribution with parameters $\phi \hat{=} (\varphi - \mathrm{e}^{-\lambda_0-\lambda_+})/(1-\mathrm{e}^{-\lambda_0-\lambda_+})$, $\lambda_0 \geqslant 0$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^m$, denoted by $\mathbf{y} \sim \text{ZIP}(\phi; \lambda_0, \lambda_1, \ldots, \lambda_m)$.

From (2.4) and (C.1), we immediately obtain

$$\begin{cases} E(\mathbf{y}) &= \dfrac{1-\varphi}{1-\psi}(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda}), \\ E(\mathbf{y}\mathbf{y}^\top) &= \dfrac{1-\varphi}{1-\psi}[\lambda_0 \cdot \mathbf{1}\mathbf{1}^\top + \text{diag}(\boldsymbol{\lambda}) + (\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})^\top], \\ \text{Var}(\mathbf{y}) &= \dfrac{1-\varphi}{1-\psi}\left[\lambda_0 \cdot \mathbf{1}\mathbf{1}^\top + \text{diag}(\boldsymbol{\lambda}) - \dfrac{\psi-\varphi}{1-\psi}(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})^\top\right]. \end{cases} \qquad (2.6)$$

Thus, $i, j = 1, \ldots, m$ and $i \neq j$, we have

$$\text{Corr}(Y_i, Y_j) = \frac{\lambda_0 - (\lambda_0+\lambda_i)(\lambda_0+\lambda_j)(\psi-\varphi)/(1-\psi)}{\sqrt{[\lambda_0+\lambda_i-(\lambda_0+\lambda_i)^2(\psi-\varphi)/(1-\psi)]\,[\lambda_0+\lambda_j-(\lambda_0+\lambda_j)^2(\psi-\varphi)/(1-\psi)]}}. \qquad (2.7)$$

From (2.7), we can see that the multivariate ZAP distribution can be used to model zero-adjusted count vectors with a more flexible dependency structure, i.e., the correlation coefficient between Y_i and Y_j could be positive or negative depending on the values of the parameters $(\varphi, \lambda_0, \boldsymbol{\lambda})$. In particular, if $\lambda_0 = 0$, we obtain

$$\mathrm{Corr}(Y_i, Y_j) = \frac{\lambda_i \lambda_j (\varphi - \psi)/(1-\psi)}{\sqrt{\left[\lambda_i - \lambda_i^2(\psi - \varphi)/(1-\psi)\right]\left[\lambda_j - \lambda_j^2(\psi - \varphi)/(1-\psi)\right]}}.$$

Furthermore, if $\lambda_i = \lambda_j = \lambda$, then

$$\mathrm{Corr}(Y_i, Y_j) = \frac{\lambda(\varphi - \psi)/(1-\psi)}{[1 - \lambda(\psi - \varphi)/(1-\psi)]}.$$

For any $r_1, \ldots, r_m \geqslant 0$, the mixed moments of $\mathbf{y}$ are given by

$$E\left(\prod_{i=1}^m Y_i^{r_i}\right) = (1-\varphi) E\left(\prod_{i=1}^m W_i^{r_i}\right) = \frac{1-\varphi}{1-\psi} E\left(\prod_{i=1}^m X_i^{r_i}\right). \tag{2.8}$$

By using the formula of $E(\xi) = E[E(\xi|Z')]$, the *moment generating function* (mgf) of $\mathbf{y}$ is

$$\begin{aligned} M_{\mathbf{y}}(\boldsymbol{t}) &= E[\exp(\boldsymbol{t}^\top \mathbf{y})] = E[\exp(Z' \cdot \boldsymbol{t}^\top \mathbf{w})] = E\Big\{E[\exp(Z' \boldsymbol{t}^\top \mathbf{w})|Z']\Big\} \\ &= E[M_{\mathbf{w}}(Z'\boldsymbol{t})] = \varphi M_{\mathbf{w}}(\mathbf{0}) + (1-\varphi) M_{\mathbf{w}}(\boldsymbol{t}) = \varphi + (1-\varphi) M_{\mathbf{w}}(\boldsymbol{t}) \\ &= \varphi + \frac{1-\varphi}{1-\psi}\left[\exp\left(\sum_{i=1}^m \lambda_i \mathrm{e}^{t_i} + \lambda_0 \mathrm{e}^{t_+} - \lambda_+ - \lambda_0\right) - \mathrm{e}^{-\lambda_0 - \lambda_+}\right]. \end{aligned} \tag{2.9}$$

Other important distributional properties of the multivariate ZAP distribution are summarized in Appendix D.

3. Statistical methods for the multivariate ZAP distribution without covariates

Suppose that $\mathbf{y}_j \stackrel{\text{ind}}{\sim} \mathrm{ZAP}(\varphi; \lambda_0 t_j, \lambda_1 t_j, \ldots, \lambda_m t_j)$, where $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{mj})^\top$ for $j = 1, \ldots, n$ and $\{t_j\}_{j=1}^n$ are positive and known constants. Let $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{mj})^\top$ denote the realization of the random vector $\mathbf{y}_j$, and $Y_{\text{obs}} = \{t_j, \boldsymbol{y}_j\}_{j=1}^n$ be the observed data.

Furthermore, let $\mathbb{J} = \{j|\boldsymbol{y}_j = \mathbf{0}, j = 1, \ldots, n\}$ and $m_0 = \sum_{j=1}^n I(\boldsymbol{y}_j = \mathbf{0})$ denote the number of elements in $\mathbb{J}$. Then, the observed-data likelihood function is proportion to

$$\begin{aligned} L(\varphi, \lambda_0, \boldsymbol{\lambda}|Y_{\text{obs}}) \;\propto\; & \varphi^{m_0}(1-\varphi)^{n-m_0} \\ & \times \left\{ \prod_{j \notin \mathbb{J}} \frac{\mathrm{e}^{-(\lambda_0+\lambda_+)t_j}}{1-\mathrm{e}^{-(\lambda_0+\lambda_+)t_j}} \left[\sum_{k=0}^{\min(\boldsymbol{y}_j)} \frac{(\lambda_0 t_j)^k}{k!} \prod_{i=1}^m \frac{(\lambda_i t_j)^{y_{ij}-k}}{(y_{ij}-k)!} \right] \right\}. \end{aligned} \tag{3.1}$$

Thus, we can write the log-likelihood function into two parts:

$$\ell(\varphi, \lambda_0, \boldsymbol{\lambda}|Y_{\text{obs}}) = \ell_1(\varphi|Y_{\text{obs}}) + \ell_2(\lambda_0, \boldsymbol{\lambda}|Y_{\text{obs}}),$$

where

$$\begin{aligned} \ell_1(\varphi|Y_{\text{obs}}) &= m_0 \log \varphi + (n-m_0)\log(1-\varphi) \quad \text{and} \\ \ell_2(\lambda_0, \boldsymbol{\lambda}|Y_{\text{obs}}) &= -(\lambda_0+\lambda_+)\sum_{j \notin \mathbb{J}} t_j - \sum_{j \notin \mathbb{J}} \log[1-\mathrm{e}^{-(\lambda_0+\lambda_+)t_j}] \\ &\quad + \sum_{j \notin \mathbb{J}} \log \left[\sum_{k=0}^{\min(\boldsymbol{y}_j)} \frac{(\lambda_0 t_j)^k}{k!} \prod_{i=1}^m \frac{(\lambda_i t_j)^{y_{ij}-k}}{(y_{ij}-k)!} \right]. \end{aligned} \tag{3.2}$$

In other words, the parameter φ and the parameter vector $(\lambda_0, \boldsymbol{\lambda})$ can be estimated separately. Obviously, the MLE of φ has an explicit solution

$$\hat{\varphi} = \frac{m_0}{n}, \tag{3.3}$$

but the closed-form MLEs of $(\lambda_0, \boldsymbol{\lambda})$ are not yet available.

3.1 MLEs via the EM algorithm

The objective of this subsection is to find the MLEs of $(\lambda_0, \boldsymbol{\lambda})$ based on (3.2). The SR (2.1) can motivate a novel EM algorithm, where some latent variables are independent of the observed variables. For each $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{mj})^\top \neq \mathbf{0}_m$, $j \notin \mathbb{J}$, we introduce latent variables $U_j \overset{\text{ind}}{\sim} \text{Bernoulli}(1-\psi_j)$ with $\psi_j = \mathrm{e}^{-(\lambda_0+\lambda_+)t_j}$, $X_{0j}^* \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_0 t_j)$, $X_{ij}^* \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_i t_j)$ for $i = 1, \ldots, m$, and $X_{0j}^* \perp\!\!\!\perp X_{ij}^*$, such that

$$(x_{0j}^* + x_{1j}^*, \ldots, x_{0j}^* + x_{mj}^*)^\top = u_j \boldsymbol{y}_j,$$

where u_j and x^*_{ij} denote the realizations of U_j and X^*_{ij}, respectively. We denote the latent/missing data by $Y_{\text{mis}} = \{u_j, x^*_{0j}, x^*_{1j}, \dots, x^*_{mj}\}_{j \notin \mathbb{J}}$, so that the complete data are

$$\begin{aligned} Y_{\text{com}} &= Y_{\text{obs}} \cup Y_{\text{mis}} = \{t_j, \boldsymbol{y}_j, u_j, x^*_{0j}, x^*_{1j}, \dots, x^*_{mj}\}_{j \notin \mathbb{J}} \\ &= \{t_j, x^*_{0j}, x^*_{1j}, \dots, x^*_{mj}\}_{j \notin \mathbb{J}} = \{t_j, x^*_{0j}, u_j, \boldsymbol{y}_j\}_{j \notin \mathbb{J}}, \end{aligned}$$

where $x^*_{ij} = u_j y_{ij} - x^*_{0j}$ for $j \notin \mathbb{J}$ and $i = 1, \dots, m$. Thus, the complete-data likelihood function for the non-zero vectors is given by

$$\begin{aligned} L_2(\lambda_0, \boldsymbol{\lambda} | Y_{\text{com}}) &= \prod_{j \notin \mathbb{J}} \left[\frac{(\lambda_0 t_j)^{x^*_{0j}} \mathrm{e}^{-\lambda_0 t_j}}{x^*_{0j}!} \prod_{i=1}^m \frac{(\lambda_i t_j)^{x^*_{ij}} \mathrm{e}^{-\lambda_i t_j}}{x^*_{ij}!} \right] \\ &= \prod_{j \notin \mathbb{J}} \left[\frac{(\lambda_0 t_j)^{x^*_{0j}} \mathrm{e}^{-\lambda_0 t_j}}{x^*_{0j}!} \prod_{i=1}^m \frac{(\lambda_i t_j)^{u_j y_{ij} - x^*_{0j}} \mathrm{e}^{-\lambda_i t_j}}{(u_j y_{ij} - x^*_{0j})!} \right] \\ &\propto \lambda_0^{(n-m_0)\bar{x}^*_0} \mathrm{e}^{-\lambda_0 \sum_{j \notin \mathbb{J}} t_j} \prod_{i=1}^m \lambda_i^{\sum_{j \notin \mathbb{J}} u_j y_{ij} - (n-m_0)\bar{x}^*_0} \mathrm{e}^{-\lambda_i \sum_{j \notin \mathbb{J}} t_j}, \end{aligned} \tag{3.4}$$

where $\bar{x}^*_0 = \sum_{j \notin \mathbb{J}} x^*_{0j}/(n - m_0)$ and the complete-data log-likelihood function for the non-zero vectors is

$$\begin{aligned} \ell_2(\lambda_0, \boldsymbol{\lambda} | Y_{\text{com}}) &= (n - m_0)\bar{x}^*_0 \log(\lambda_0) - \lambda_0 \sum_{j \notin \mathbb{J}} t_j \\ &\quad + \sum_{i=1}^m \left\{ \left[\sum_{j \notin \mathbb{J}} u_j y_{ij} - (n - m_0)\bar{x}^*_0 \right] \log(\lambda_i) - \lambda_i \sum_{j \notin \mathbb{J}} t_j \right\}. \end{aligned}$$

The M-step is to calculate the complete-data MLEs:

$$\hat{\lambda}_0 = \frac{\sum_{j \notin \mathbb{J}} x^*_{0j}}{\sum_{j \notin \mathbb{J}} t_j} \quad \text{and} \quad \hat{\lambda}_i = \frac{\sum_{j \notin \mathbb{J}} u_j y_{ij}}{\sum_{j \notin \mathbb{J}} t_j} - \hat{\lambda}_0, \qquad i = 1, \dots, m, \tag{3.5}$$

and the E-step is to replace $\{u_j\}_{j\notin\mathbb{J}}$ and $\{x_{0j}^*\}_{j\notin\mathbb{J}}$ in (3.5) by their conditional expectations:

$$E(U_j|Y_{\rm obs},\lambda_0,\boldsymbol{\lambda}) \quad = \quad E(U_j) = 1-{\rm e}^{-(\lambda_0+\lambda_+)t_j}, \quad \text{and} \tag{3.6}$$

$$E(X_{0j}^*|Y_{\rm obs},\lambda_0,\boldsymbol{\lambda}) \overset{\rm (C.23)}{=} \frac{[1-{\rm e}^{-(\lambda_0+\lambda_+)t_j}]\displaystyle\sum_{k_j=1}^{\min(\boldsymbol{y}_j)}\frac{(\lambda_0 t_j)^{k_j}}{(k_j-1)!}\prod_{i=1}^m\frac{(\lambda_i t_j)^{y_{ij}-k_j}}{(y_{ij}-k_j)!}}{\displaystyle\sum_{k_j=0}^{\min(\boldsymbol{y}_j)}\frac{(\lambda_0 t_j)^{k_j}}{k_j!}\prod_{i=1}^m\frac{(\lambda_i t_j)^{y_{ij}-k_j}}{(y_{ij}-k_j)!}} \times I(\min(\boldsymbol{y}_j)\geqslant 1), \tag{3.7}$$

respectively. An important feature of this EM algorithm is that the latent variables $\{U_j\}_{j\notin\mathbb{J}}$ are independent of the observed variables $\{\mathbf{y}_j\}_{j\notin\mathbb{J}}$.

3.2 Bootstrap confidence intervals

When other approaches are not available, the bootstrap method is a useful tool to find *confidence intervals* (CIs) for an arbitrary function of $(\varphi,\lambda_0,\boldsymbol{\lambda})$, say, $\vartheta=h(\varphi,\lambda_0,\boldsymbol{\lambda})$. Let $(\hat\varphi,\hat\lambda_0,\hat{\boldsymbol{\lambda}})$ be the MLEs of $(\varphi,\lambda_0,\boldsymbol{\lambda})$ calculated by (3.3) and by the EM algorithm (3.5)–(3.7), then $\hat\vartheta=h(\hat\varphi,\hat\lambda_0,\hat{\boldsymbol{\lambda}})$ is the MLE of ϑ. Based on $(\hat\varphi,\hat\lambda_0,\hat{\boldsymbol{\lambda}})$, we can independently generate $\mathbf{y}_j^*\overset{\rm ind}{\sim}{\rm ZAP}(\hat\varphi,\hat\lambda_0 t_j,\hat\lambda_1 t_j,\dots,\hat\lambda_m t_j)$ via the SR (2.4) and the SR (2.3) for $j=1,\dots,n$. Having obtained $Y_{\rm obs}^*=\{t_j,\boldsymbol{y}_j^*\}_{j=1}^n$, we can calculate the bootstrap replication $(\hat\varphi^*,\hat\lambda_0^*,\hat{\boldsymbol{\lambda}}^*)$ and get $\hat\vartheta^*=h(\hat\varphi^*,\hat\lambda_0^*,\hat{\boldsymbol{\lambda}}^*)$. Independently repeating this process G times, we obtain G bootstrap replications $\{\hat\vartheta_g^*\}_{g=1}^G$. Consequently, the standard error, ${\rm se}(\hat\vartheta)$, of $\hat\vartheta$ can be estimated by the sample standard deviation of the G replications, i.e.,

$$\widehat{\rm se}(\hat\vartheta)=\left\{\frac{1}{G-1}\sum_{g=1}^G[\hat\vartheta_g^*-(\hat\vartheta_1^*+\cdots+\hat\vartheta_G^*)/G]^2\right\}^{1/2}. \tag{3.8}$$

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

$$\left[\hat\vartheta-z_{\alpha/2}\cdot\widehat{\rm se}(\hat\vartheta),\ \hat\vartheta+z_{\alpha/2}\cdot\widehat{\rm se}(\hat\vartheta)\right]. \tag{3.9}$$

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

$$[\hat\vartheta_{\rm L},\ \hat\vartheta_{\rm U}], \tag{3.10}$$

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

3.3 Bayesian methods

3.3.1 Posterior modes via the EM algorithm

Without loss of generality, let $\mathbb{J} = \{1, 2, \ldots, m_0\}$. Then, the EM algorithm introduced in Section 3.1 can be re-expressed as follows. The observed data $Y_{\rm obs}$ were augmented by the latent variables $\{U_j\}_{j=m_0+1}^n \stackrel{\rm ind}{\sim} \text{Bernoulli}(1-{\rm e}^{-(\lambda_0+\lambda_+)t_j})$ and $\{X_{0j}^*\}_{j=m_0+1}^n \stackrel{\rm ind}{\sim} \text{Poisson}(\lambda_0 t_j)$ to form the complete data

$$Y_{\rm com} = \{\underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{m_0}, \{\boldsymbol{y}_j, t_j\}_{j=m_0+1}^n;\ \{u_j, x_{0j}^*\}_{j=m_0+1}^n\}.$$

From (3.4) and the observed-data likelihood function (3.1), we know that the complete-data likelihood function is proportional to

$$\begin{aligned} L(\varphi, \lambda_0, \boldsymbol{\lambda}|Y_{\rm com}) \ &\propto\ \varphi^{m_0}(1-\varphi)^{n-m_0} \cdot \lambda_0^{\sum_{j=m_0+1}^n x_{0j}^*} {\rm e}^{-\lambda_0 \sum_{j=m_0+1}^n t_j} \\ &\quad \times \prod_{i=1}^m \lambda_i^{\sum_{j=m_0+1}^n (u_j y_{ij} - x_{0j}^*)} {\rm e}^{-\lambda_i \sum_{j=m_0+1}^n t_j}. \end{aligned} \tag{3.11}$$

If we adopt $\text{Beta}(a, b)$ as the prior distribution of φ and $\text{Gamma}(a_i, b_i)$ as the priors of λ_i, $i = 0, 1, \ldots, m$, from (3.11), the complete-data posterior distributions are given by

$$\begin{aligned} \varphi|(Y_{\rm obs}, \boldsymbol{u}, \boldsymbol{x}_0^*) \ &\sim\ \text{Beta}(a+m_0,\ b+n-m_0), \\ \lambda_0|(Y_{\rm obs}, \boldsymbol{u}, \boldsymbol{x}_0^*) \ &\sim\ \text{Gamma}\left(a_0 + \sum_{j=m_0+1}^n x_{0j}^*,\ b_0 + \sum_{j=m_0+1}^n t_j\right), \\ \lambda_i|(Y_{\rm obs}, \boldsymbol{u}, \boldsymbol{x}_0^*) \ &\stackrel{\rm ind}{\sim}\ \text{Gamma}\left(a_i + \sum_{j=m_0+1}^n (u_j y_{ij} - x_{0j}^*),\ b_i + \sum_{j=m_0+1}^n t_j\right), \end{aligned} \tag{3.12}$$

for $i = 1, \ldots, m$, where $\boldsymbol{u} = (u_{m_0+1}, \ldots, u_n)^\top$ and $\boldsymbol{x}_0^* = (x_{0,m_0+1}^*, \ldots, x_{0n}^*)^\top$. The M-step of the EM algorithm is to calculate the complete-data posterior modes of φ, λ_0 and $\boldsymbol{\lambda}$, which are given by

$$\begin{aligned} \tilde{\varphi} \ &=\ \frac{a+m_0-1}{a+b+n-2}, \qquad \tilde{\lambda}_0 = \frac{a_0 + \sum_{j=m_0+1}^n x_{0j}^* - 1}{b_0 + \sum_{j=m_0+1}^n t_j}, \\ \tilde{\lambda}_i \ &=\ \frac{a_i + \sum_{j=m_0+1}^n (u_j y_{ij} - x_{0j}^*) - 1}{b_i + \sum_{j=m_0+1}^n t_j}, \qquad i = 1, \ldots, m. \end{aligned} \tag{3.13}$$

And the E-step is to replace $\{u_j\}_{j=m_0+1}^n$ and $\{x_{0j}^*\}_{j=m_0+1}^n$ by their conditional expectation (3.6) and (3.7).

3.3.2 Generation of posterior samples via the DA algorithm

The conditional predictive distributions of $\{U_j\}_{j=m_0+1}^n$ and the conditional predictive distributions of $\{X_{0j}^*\}_{j=m_0+1}^n$ based on (C.21) are given by

$$U_j|(Y_{\text{obs}}, \varphi, \lambda_0, \boldsymbol{\lambda}) \overset{\text{ind}}{\sim} \text{Bernoulli}(1 - \mathrm{e}^{-(\lambda_0+\lambda_+)t_j}), \quad \text{and} \tag{3.14}$$

$$X_{0j}^*|(Y_{\text{obs}}, \varphi, \lambda_0, \boldsymbol{\lambda}) \overset{\text{ind}}{\sim} \text{Finite}(l, p_l(\boldsymbol{y}_j, \lambda_0 t_j, \boldsymbol{\lambda} t_j);\ l = 0, 1, \ldots, \min(\boldsymbol{y}_j)), \tag{3.15}$$

for $j = m_0+1, \ldots, n$, where the functions $\{p_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda})\}$ are defined by (C.24). To make a full Bayesian inference on the parameters $(\varphi, \lambda_0, \boldsymbol{\lambda})$, we need to generate posterior samples from the observed posterior distribution $f(\varphi, \lambda_0, \boldsymbol{\lambda}|Y_{\text{obs}})$ by using the DA algorithm. The I-step is to independently draw the latent variables $\{u_j\}_{j=m_0+1}^n$ from (3.14) for given $(Y_{\text{obs}}, \varphi, \lambda_0, \boldsymbol{\lambda})$ and $\{x_{0j}^*\}_{j=m_0+1}^n$ from (3.15), and the P-step is to independently draw φ, λ_0 and $\boldsymbol{\lambda}$ from (3.12) for given $(Y_{\text{obs}}, \boldsymbol{u}, \boldsymbol{x}_0^*)$.

3.4 Hypothesis testing

Let $\mathbf{y}_j \overset{\text{ind}}{\sim} \text{ZAP}(\varphi; \lambda_0 t_j, \lambda_1 t_j, \ldots, \lambda_m t_j)$, where $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{mj})^\top$ for $j = 1, \ldots, n$ and $\{t_j\}_{j=1}^n$ are positive and known constants. In this subsection, we first consider to test whether the Poisson components in $\mathbf{x}$ are independent; i.e., to test whether H_0: $\lambda_0 = 0$ is true. If H_0 is rejected, then, we need to confirm the status of the zero-adjusted distribution, i.e., whether it is zero-inflated or zero-deflated.

3.4.1 Hypothesis testing on $\lambda_0 = 0$

Suppose we want to test the null hypothesis

$$H_0\colon\ \lambda_0 = 0 \quad \text{against} \quad H_1\colon\ \lambda_0 > 0. \tag{3.16}$$

Under H_0, the likelihood ratio test statistic

$$T_1 = -2\{\ell(\hat\varphi_0, 0, \hat{\boldsymbol{\lambda}}_0|Y_{\text{obs}}) - \ell(\hat\varphi, \hat\lambda_0, \hat{\boldsymbol{\lambda}}|Y_{\text{obs}})\} \dot\sim \chi^2(1), \tag{3.17}$$

where $(\hat{\varphi}_0, \hat{\boldsymbol{\lambda}}_0)$ are the MLEs of $(\varphi, \boldsymbol{\lambda})$ under H_0, and $(\hat{\varphi}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}})$ are the unconstrained MLEs of $(\varphi, \lambda_0, \boldsymbol{\lambda})$. Since the null hypothesis corresponds to λ_0 being on the boundary of the parameter space, the reference distribution for T_1 should be an equal mixture of a χ_0^2 (a constant at zero) and a χ_1^2 distribution, and thus the corresponding p-value is

$$p_{\rm v1} = \frac{1}{2}\Pr(T_1 > t_1|H_0) = \frac{1}{2}\Pr(\chi^2(1) > t_1). \tag{3.18}$$

3.4.2 Hypothesis testing on inflation or deflation

In this subsubsection, we only consider the special case of $t_1 = \cdots = t_n = 1$; that is, we assume that $\mathbf{y}_1, \ldots, \mathbf{y}_n \stackrel{\rm iid}{\sim} \text{ZAP}(\varphi\,; \lambda_0, \lambda_1, \ldots, \lambda_m)$. We first consider to test whether $(Y_{1j}, \ldots, Y_{mj})^{\top} \sim \text{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$; i.e., to test the following hypotheses:

$$H_0\colon\ \varphi = {\rm e}^{-\lambda_0-\lambda_+} \quad \text{against} \quad H_1\colon\ \varphi \neq {\rm e}^{-\lambda_0-\lambda_+}. \tag{3.19}$$

Under H_0, the likelihood ratio test statistic

$$T_2 = -2\{\ell(\hat{\varphi}_0, \hat{\lambda}_{00}, \hat{\boldsymbol{\lambda}}_0|Y_{\rm obs}) - \ell(\hat{\varphi}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}}|Y_{\rm obs})\} \mathrel{\dot\sim} \chi^2(1), \tag{3.20}$$

where $(\hat{\varphi}_0, \hat{\lambda}_{00}, \hat{\boldsymbol{\lambda}}_0)$ are the MLEs of $(\varphi, \lambda_0, \boldsymbol{\lambda})$ under H_0, and $(\hat{\varphi}, \hat{\lambda}_0, \hat{\boldsymbol{\lambda}})$ are the unconstrained MLEs of $(\varphi, \lambda_0, \boldsymbol{\lambda})$. The corresponding p-value is given by

$$p_{\rm v2} = \Pr(T_2 > t_2|H_0). \tag{3.21}$$

If $p_{\rm v2} > \alpha$, we cannot reject the null hypothesis H_0 at the α level of significance. However, if H_0 specified by (3.19) is rejected, we should consider to test whether $\mathbf{y}$ follows a multivariate ZDP distribution, i.e., to test the following hypotheses:

$$H_0'\colon\ \varphi < {\rm e}^{-\lambda_0-\lambda_+} \quad \text{against} \quad H_1'\colon\ \varphi > {\rm e}^{-\lambda_0-\lambda_+}. \tag{3.22}$$

Let $\theta = \varphi - {\rm e}^{-\lambda_0-\lambda_+}$. Then, testing H_0' is equivalent to testing $H_0''\colon\ \theta < 0$. We can construct $100(1-\alpha)\%$ bootstrap CIs of θ and could use the nonnormal-based bootstrap CI $[\hat{\theta}_{\rm L},\ \hat{\theta}_{\rm U}]$ as the acceptable interval. If $\hat{\theta}_{\rm L} > 0$, then H_0'' should be rejected at the α level of significance, i.e., $\mathbf{y}$ follows a multivariate ZIP distribution; if $\hat{\theta}_{\rm U} < 0$, it cannot be rejected, i.e., $\mathbf{y}$ follows a multivariate ZDP distribution.

4. The multivariate ZAP regression model

4.1 Model formulation

We consider the following regression model

$$\begin{cases} \mathbf{y}_j & \stackrel{\text{ind}}{\sim} & \text{ZAP}(\varphi_j\,;\lambda_{0j},\lambda_{1j},\ldots,\lambda_{mj}), \qquad j=1,\ldots,n, \\ \log\left(\dfrac{\varphi_j}{1-\varphi_j}\right) & = & \boldsymbol{v}_{1j}^{\top}\boldsymbol{\beta}, \\ \log(\lambda_{0j}) & = & \boldsymbol{v}_{2j}^{\top}\boldsymbol{\eta}, \\ \log(\lambda_{ij}) & = & \boldsymbol{v}_{3j}^{\top}\boldsymbol{\gamma}_i, \qquad i=1,\ldots,m, \end{cases} \tag{4.1}$$

where $\boldsymbol{v}_{1j}=(1,v_{11j},\ldots,v_{1pj})^{\top}$, $\boldsymbol{v}_{2j}=(1,v_{21j},\ldots,v_{2sj})^{\top}$ and $\boldsymbol{v}_{3j}=(1,v_{31j},\ldots,v_{3qj})^{\top}$ are not necessarily identical covariate vectors associated with the subject j, $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_p)^{\top}$, $\boldsymbol{\eta}=(\eta_0,\eta_1,\ldots,\eta_s)^{\top}$ and $\boldsymbol{\gamma}_i=(\gamma_{i0},\gamma_{i1},\ldots,\gamma_{iq})^{\top}$ are regression coefficients. The primary purpose of this section is to estimate the parameter vector $\boldsymbol{\theta}=(\boldsymbol{\beta}^{\top},\boldsymbol{\eta}^{\top},\boldsymbol{\gamma}_1^{\top},\ldots,\boldsymbol{\gamma}_m^{\top})^{\top}$.

4.2 MLEs via the EM algorithm

4.2.1 Latent variables and complete-data likelihood function

Let $\boldsymbol{y}_j=(y_{1j},\ldots,y_{mj})^{\top}$ denote the realization of the random vector $\mathbf{y}_j$, and $Y_{\text{obs}}=\{\boldsymbol{y}_j\}_{j=1}^n$ be the observed data. To derive the MLEs of $\boldsymbol{\theta}$, we employ the EM algorithm again.

- For each $\boldsymbol{y}_j$ for $j\in\{1,\ldots,n\}$, based on (2.4), we first introduce latent variables $Z'_j \stackrel{\text{ind}}{\sim} \text{Bernoulli}(1-\varphi_j)$, latent vectors $\mathbf{w}_j \stackrel{\text{ind}}{\sim} \text{ZTP}(\lambda_{0j},\lambda_{1j},\ldots,\lambda_{mj})$ and $Z'_j \perp\!\!\!\perp \mathbf{w}_j$, such that $\boldsymbol{y}_j=z'_j\boldsymbol{w}_j$, where z'_j and $\boldsymbol{w}_j$ denote the realizations of Z'_j and $\mathbf{w}_j$, respectively. Furthermore, we define

$$\begin{aligned} \mathbb{J} &= \{j|\boldsymbol{y}_j=\mathbf{0},\ j=1,\ldots,n\}=\{j|z'_j=0,\ j=1,\ldots,n\} \quad \text{and} \\ \mathbb{J}^{\text{c}} &= \{j|\boldsymbol{y}_j\neq\mathbf{0},\ j=1,\ldots,n\}=\{j|z'_j=1,\ j=1,\ldots,n\}. \end{aligned} \tag{4.2}$$

- For each $\boldsymbol{w}_j$ for $j\in\mathbb{J}^{\text{c}}$, based on (2.1), we then introduce latent variables

$$U_j \stackrel{\text{ind}}{\sim} \text{Bernoulli}(1-\text{e}^{-\lambda_{0j}-\lambda_{+j}}) \quad \text{with} \quad \lambda_{+j}=\sum_{i=1}^m \lambda_{ij}, \tag{4.3}$$

$X_{0j}^* \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_{0j})$, $X_{ij}^* \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_{ij})$ for $i = 1, \ldots, m$, and $X_{0j}^* \perp\!\!\!\perp X_{ij}^*$, such that

$$(x_{0j}^* + x_{1j}^*, \ldots, x_{0j}^* + x_{mj}^*)^\top = u_j \boldsymbol{w}_j,$$

where u_j and x_{ij}^* denote the realizations of U_j and X_{ij}^*, respectively.

We denote the latent/missing data by $Y_{\text{mis}} = \{z_j'\}_{j=1}^n \cup \{\boldsymbol{w}_j, u_j, x_{0j}^*, x_{1j}^*, \ldots, x_{mj}^*\}_{j \in \mathbb{J}^c}$, so that the complete data are

$$\begin{aligned} Y_{\text{com}} &= Y_{\text{obs}} \cup Y_{\text{mis}} = \{\boldsymbol{y}_j, z_j'\}_{j=1}^n \cup \{\boldsymbol{w}_j, u_j, x_{0j}^*, x_{1j}^*, \ldots, x_{mj}^*\}_{j \in \mathbb{J}^c} \\ &= \{\boldsymbol{y}_j, z_j'\}_{j=1}^n \cup \{u_j, x_{0j}^*\}_{j \in \mathbb{J}^c}, \end{aligned}$$

where $x_{ij}^* = u_j y_{ij} - x_{0j}^*$ for $j \in \mathbb{J}^c$ and $i = 1, \ldots, m$. Therefore, the complete-data likelihood function is given by

$$\begin{aligned} L(\boldsymbol{\theta}|Y_{\text{com}}) &\propto \left[\prod_{j=1}^n \varphi_j^{1-z_j'}(1-\varphi_j)^{z_j'}\right] \cdot \left(\prod_{j \in \mathbb{J}^c} \lambda_{0j}^{x_{0j}^*} \mathrm{e}^{-\lambda_{0j}} \prod_{i=1}^m \lambda_{ij}^{u_j y_{ij} - x_{0j}^*} \mathrm{e}^{-\lambda_{ij}}\right) \\ &= \prod_{j=1}^n \varphi_j^{1-z_j'}(1-\varphi_j)^{z_j'} \left(\lambda_{0j}^{x_{0j}^*} \mathrm{e}^{-\lambda_{0j}} \textstyle\prod_{i=1}^m \lambda_{ij}^{u_j y_{ij} - x_{0j}^*} \mathrm{e}^{-\lambda_{ij}}\right)^{z_j'}, \end{aligned}$$

and the log-likelihood function can be decomposed into $2 + m$ parts:

$$\ell(\boldsymbol{\theta}|Y_{\text{com}}) = \ell_1(\boldsymbol{\beta}|Y_{\text{com}}) + \ell_2(\boldsymbol{\eta}|Y_{\text{com}}) + \sum_{i=1}^m \ell_{3i}(\boldsymbol{\gamma}_i|Y_{\text{com}}),$$

where

$$\begin{aligned} \ell_1(\boldsymbol{\beta}|Y_{\text{com}}) &= \sum_{j=1}^n [(1-z_j') \log \varphi_j + z_j' \log(1-\varphi_j)] \\ &= \sum_{j=1}^n \Big[(1-z_j')\boldsymbol{v}_{1j}^\top \boldsymbol{\beta} - \log(1 + \mathrm{e}^{\boldsymbol{v}_{1j}^\top \boldsymbol{\beta}})\Big], \\ \ell_2(\boldsymbol{\eta}|Y_{\text{com}}) &= \sum_{j=1}^n [-\lambda_{0j} z_j' + z_j' x_{0j}^* \log(\lambda_{0j})] = \sum_{j=1}^n z_j' \Big[-\mathrm{e}^{\boldsymbol{v}_{2j}^\top \boldsymbol{\eta}} + x_{0j}^* \boldsymbol{v}_{2j}^\top \boldsymbol{\eta}\Big], \\ \ell_{3i}(\boldsymbol{\gamma}_i|Y_{\text{com}}) &= \sum_{j=1}^n z_j' [-\lambda_{ij} + (u_j y_{ij} - x_{0j}^*) \log(\lambda_{ij})] \\ &= \sum_{j=1}^n z_j' \Big[-\mathrm{e}^{\boldsymbol{v}_{3j}^\top \boldsymbol{\gamma}_i} + (u_j y_{ij} - x_{0j}^*) \boldsymbol{v}_{3j}^\top \boldsymbol{\gamma}_i\Big]. \end{aligned}$$

4.2.2 M-step via the Newton–Raphson iteration

The first partial derivatives of the log-likelihood function are given by

$$\left\{\begin{array}{rcl} \dfrac{\partial \ell_1(\boldsymbol{\beta}|Y_{\rm com})}{\partial \boldsymbol{\beta}} &=& \displaystyle\sum_{j=1}^n (1 - z_j' - \varphi_j)\boldsymbol{v}_{1j} = \mathbf{V}_1^\top(\mathbf{1} - \boldsymbol{z}' - \boldsymbol{\varphi}), \\ \dfrac{\partial \ell_2(\boldsymbol{\eta}|Y_{\rm com})}{\partial \boldsymbol{\eta}} &=& \displaystyle\sum_{j=1}^n [z_j' x_{0j}^* - z_j'\lambda_{0j}]\boldsymbol{v}_{2j} = \mathbf{V}_2^\top[\boldsymbol{z}' \circ \boldsymbol{x}_0^* - \boldsymbol{z}' \circ \boldsymbol{\lambda}_{(0)}], \\ \dfrac{\partial \ell_{3i}(\boldsymbol{\gamma}_i|Y_{\rm com})}{\partial \boldsymbol{\gamma}_i} &=& \displaystyle\sum_{j=1}^n [(z_j' u_j) y_{ij} - z_j' x_{0j}^* - z_j'\lambda_{ij}]\boldsymbol{v}_{3j} \\ &=& \mathbf{V}_3^\top[(\boldsymbol{z}' \circ \boldsymbol{u}) \circ \boldsymbol{y}_{(i)} - \boldsymbol{z}' \circ \boldsymbol{x}_0^* - \boldsymbol{z}' \circ \boldsymbol{\lambda}_{(i)}], \qquad i = 1, \ldots, m, \end{array}\right. \tag{4.4}$$

where

$$\begin{array}{rclrclrcl} \mathbf{V}_1 &=& (\boldsymbol{v}_{11}, \ldots, \boldsymbol{v}_{1n})^\top, & \boldsymbol{z}' &=& (z_1', \ldots, z_n')^\top, & \boldsymbol{\varphi} &=& (\varphi_1, \ldots, \varphi_n)^\top, \\ \mathbf{V}_2 &=& (\boldsymbol{v}_{21}, \ldots, \boldsymbol{v}_{2n})^\top, & \boldsymbol{u} &=& (u_1, \ldots, u_n)^\top, & \boldsymbol{y}_{(i)} &=& (y_{i1}, \ldots, y_{in})^\top, \\ \mathbf{V}_3 &=& (\boldsymbol{v}_{31}, \ldots, \boldsymbol{v}_{3n})^\top, & \boldsymbol{x}_0^* &=& (x_{01}^*, \ldots, x_{0n}^*)^\top, & bla_{(0)} &=& (\lambda_{01}, \ldots, \lambda_{0n})^\top, \\ \boldsymbol{\lambda}_{(i)} &=& (\lambda_{i1}, \ldots, \lambda_{in})^\top, & i &=& 1, \ldots, m, \end{array}$$

and a new operator "$\circ$" is defined by $\boldsymbol{a} \circ \boldsymbol{b} = (a_1 b_1, \ldots, a_n b_n)^\top$.

From (4.4), closed-form MLEs of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m)$ based on the complete-data are not yet available. Since the complete-data observed information matrices are given by

$$\left\{\begin{array}{rcl} \mathbf{I}_{\rm com}(\boldsymbol{\beta}) &=& -\dfrac{\partial^2 \ell_1(\boldsymbol{\beta}|Y_{\rm com})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \mathbf{V}_1^\top \operatorname{diag}[\boldsymbol{\varphi} \circ (1 - \boldsymbol{\varphi})]\mathbf{V}_1, \\ \mathbf{I}_{\rm com}(\beta) &=& -\dfrac{\partial^2 \ell_2(\boldsymbol{\eta}|Y_{\rm com})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} = \mathbf{V}_2^\top \operatorname{diag}(\boldsymbol{z}' \circ \boldsymbol{\lambda}_{(0)})\mathbf{V}_2, \\ \mathbf{I}_{\rm com}(\boldsymbol{\gamma}_i) &=& -\dfrac{\partial^2 \ell_{3i}(\boldsymbol{\gamma}_i|Y_{\rm com})}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^\top} = \mathbf{V}_3^\top \operatorname{diag}(\boldsymbol{z}' \circ \boldsymbol{\lambda}_{(i)})\mathbf{V}_3, \qquad i = 1, \ldots, m, \end{array}\right. \tag{4.5}$$

the M-step is to separately calculate the MLEs of $(\boldsymbol{\beta}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_m)$ via the Newton–Raphson algorithm as follows:

$$\left\{\begin{array}{rcl} \boldsymbol{\beta}^{(t+1)} &=& \boldsymbol{\beta}^{(t)} + \mathbf{I}_{\rm com}^{-1}(\boldsymbol{\beta}^{(t)})\mathbf{V}_1^\top \left[\mathbf{1} - \boldsymbol{z}' - \boldsymbol{\varphi}^{(t)}\right], \\ \boldsymbol{\eta}^{(t+1)} &=& \boldsymbol{\eta}^{(t)} + \mathbf{I}_{\rm com}^{-1}(\boldsymbol{\eta}^{(t)})\mathbf{V}_2^\top \left[\boldsymbol{z}' \circ \boldsymbol{x}_0^* - \boldsymbol{z}' \circ \boldsymbol{\lambda}_{(0)}\right], \\ \boldsymbol{\gamma}_i^{(t+1)} &=& \boldsymbol{\gamma}_i^{(t)} + \mathbf{I}_{\rm com}^{-1}(\boldsymbol{\gamma}_i^{(t)})\mathbf{V}_3^\top \left[(\boldsymbol{z}' \circ \boldsymbol{u}) \circ \boldsymbol{y}_{(i)} - \boldsymbol{z}' \circ \boldsymbol{x}_0^* - \boldsymbol{z}' \circ \boldsymbol{\lambda}_{(i)}^{(t)})\right], \end{array}\right. \tag{4.6}$$

for $i = 1, \ldots, m$.

4.2.3 E-step with explicit expressions

When $j \in \mathbb{J}^c$, we have

$$E(U_j|Y_{\text{obs}}, \boldsymbol{\theta}) \overset{(4.3)}{=} 1 - \mathrm{e}^{-\lambda_{0j}-\lambda_{+j}} \tag{4.7}$$

and

$$\begin{aligned} E(X_{0j}^*|Y_{\text{obs}}, \boldsymbol{\theta}) &= E(X_{0j}^*|\mathbf{y}_j = \boldsymbol{y}_j, \boldsymbol{\theta}) \overset{(4.2)}{=} E(X_{0j}^*|\mathbf{w}_j = \boldsymbol{y}_j, \boldsymbol{\theta}) \\ &\overset{(C.23)}{=} (1 - \mathrm{e}^{-\lambda_{0j}-\lambda_{+j}}) \cdot \frac{\displaystyle\sum_{k=1}^{\min(\boldsymbol{y}_j)} \frac{\lambda_{0j}^k}{(k-1)!} \prod_{i=1}^m \frac{\lambda_{ij}^{y_{ij}-k}}{(y_{ij}-k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{y}_j)} \frac{\lambda_{0j}^k}{k!} \prod_{i=1}^m \frac{\lambda_{ij}^{y_{ij}-k}}{(y_{ij}-k)!}} \cdot I(\min(\boldsymbol{y}_j) \geqslant 1). \end{aligned} \tag{4.8}$$

Therefore, the E-step is to replace $\boldsymbol{z}'$, $\boldsymbol{z}' \circ \boldsymbol{u}$ and $\boldsymbol{z}' \circ \boldsymbol{x}_0^*$ in (4.6) by their conditional expectations

$$E(Z_j'|Y_{\text{obs}}, \boldsymbol{\theta}) = E(Z_j'|\mathbf{y}_j = \boldsymbol{y}_j) \overset{(D.9)}{=} I(\boldsymbol{y}_j \neq \mathbf{0}) \overset{(4.2)}{=} I(j \in \mathbb{J}^c), \tag{4.9}$$

$$\begin{aligned} E(Z_j'U_j|Y_{\text{obs}}, \boldsymbol{\theta}) &= E(Z_j'|Y_{\text{obs}}, \boldsymbol{\theta}) \cdot E(U_j|Y_{\text{obs}}, \boldsymbol{\theta}) \\ &\overset{(4.9)}{=} I(j \in \mathbb{J}^c) \cdot E(U_j|Y_{\text{obs}}, \boldsymbol{\theta}) \overset{(4.7)}{=} (1 - \mathrm{e}^{-\lambda_{0j}-\lambda_{+j}}) \cdot I(j \in \mathbb{J}^c), \end{aligned} \tag{4.10}$$

and

$$\begin{aligned} &E(Z_j'X_{0j}^*|Y_{\text{obs}}, \boldsymbol{\theta}) = E(Z_j'|Y_{\text{obs}}, \boldsymbol{\theta}) \cdot E(X_{0j}^*|Y_{\text{obs}}, \boldsymbol{\theta}) \\ \overset{(4.9)}{=}\ &I(j \in \mathbb{J}^c) \cdot E(X_{0j}^*|Y_{\text{obs}}, \boldsymbol{\theta}) \\ \overset{(4.8)}{=}\ &(1 - \mathrm{e}^{-\lambda_{0j}-\lambda_{+j}}) \frac{\displaystyle\sum_{k=1}^{\min(\boldsymbol{y}_j)} \frac{\lambda_{0j}^k}{(k-1)!} \prod_{i=1}^m \frac{\lambda_{ij}^{y_{ij}-k}}{(y_{ij}-k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{y}_j)} \frac{\lambda_{0j}^k}{k!} \prod_{i=1}^m \frac{\lambda_{ij}^{y_{ij}-k}}{(y_{ij}-k)!}} \cdot I(j \in \mathbb{J}^c,\, \min(\boldsymbol{y}_j) \geqslant 1). \end{aligned} \tag{4.11}$$

The CIs of the components of $\boldsymbol{\theta}$ can be constructed by the bootstrap method.

5. Real data analysis

In this section, a two-dimensional health center visit data set of California is studied by fitting the proposed bivariate ZTP model, while a three-dimensional physician office visit data set is analyzed by fitting the multivariate ZAP distribution without considering covariates and by fitting the corresponding regression model.

5.1 Health center visit data of California

Gurmu (1997) studied a doctor visit data set by using a semi-parametric hurdle regression model. This data set came from the Medicaid Consumer Survey which was sponsored by the Health Care Financing Administration. The survey conducted in 1986 was part of the data-collection activity of the Nationwide Evaluation of Medicaid Competition Demonstrations. The data items included in the survey are: beneficiary demographics, health-care utilization, health status, health habits, attitudes on access to and satisfaction with health services. Gurmu selected two sites in California—Santa Barbara and Ventura counties—to form his data set. The California survey was conducted in personal interviews with samples of demonstration enrollees in Santa Barbara county and a fee-for-service comparison group of non-enrollees from nearby Ventura county. The managed programme enrolled adults qualifying for Aid to Families with Dependent Children (AFDC) and non-institutionalized Supplementary Security Income (SSI) recipients. Gurmu (1997) treated the number of children in the household as an explanatory/independent variable, while viewing the number of the doctor office/clinic and health center visits as the response/dependent variable. Note that the number of children in the household is also a discrete variable, we fit the paired counts with a bivariate discrete model. In this subsection, we use Gurmu's AFDC data set by letting W_1 denote the number of the doctor office/clinic and health center visits during a period of 4 months (120 days) and W_2 the number of children in the household. A total of 243 enrollees and 242 non-enrollees were interviewed and the two samples are mixed. The sample mean, standard deviation, minimum and maximum for W_1 are 1.6103, 3.3468, 0 and 48, respectively, while those for W_2 are 2.2639, 1.3191, 1 and 9, respectively. Since the zero vector $(0,0)^\top$ does not occur in this data set shown in Table 1, we would like to fit the data set by the proposed bivariate ZTP model.

Table 1 The health center visit data of California in USA (Gurmu, 1997)

$W_1 \backslash W_2$	1	2	3	4	5	6	7	9	Total
0	78	73	44	27	11	4	3	1	241
1	33	25	23	12	3	0	0	0	96
2	26	15	9	2	2	0	1	0	55
3	9	9	5	4	3	0	0	0	30
4	5	5	2	1	0	1	0	0	14
5	3	3	1	1	0	0	0	0	8
6	3	5	3	1	0	0	0	0	12
7	3	3	1	0	0	0	0	0	7
8	0	0	4	0	0	1	0	0	5
9	5	1	1	1	0	0	0	0	8
10	1	2	0	0	0	0	0	0	3
11	1	0	0	0	0	0	0	0	1
12	1	0	0	0	0	0	0	0	1
15	1	0	0	0	0	0	0	0	1
16	0	1	0	0	0	0	0	0	1
24	1	0	0	0	0	0	0	0	1
48	1	0	0	0	0	0	0	0	1
Total	171	142	93	49	19	6	4	1	485

5.1.1 Likelihood-based inferences

Let $\mathbf{w}_j = (W_{1j}, W_{2j})^\top \overset{\text{iid}}{\sim} \text{ZTP}(\lambda_0, \lambda_1, \lambda_2)$, where W_{1j} and W_{2j} denote the number of the doctor office/clinic and health center visits during a period of 4 months (120 days) and the number of children in the household, respectively, for $j = 1, \dots, n \, (n = 485)$. To find the MLEs of λ_0 and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$, we choose $\lambda_0^{(0)} = 1$ and $\boldsymbol{\lambda}^{(0)} = \mathbf{1}_2$ as their initial values. The MLEs of $(\lambda_0, \boldsymbol{\lambda})$ converged to $\hat{\lambda}_0 = 8.3731 \times 10^{-10} \approx 0$ and $\hat{\boldsymbol{\lambda}} = (1.5738, 2.2126)^\top$ as shown in the second column of Table 2 in 177 iterations for the EM algorithm (3.5)–(3.7) with $m = 2$, $m_0 = 0$, $t_j = 1$ and $\boldsymbol{y}_j = \boldsymbol{w}_j$. Since the common Poisson parameter λ_0 is approximately equal to 0, it indicates that the proposed bivariate ZTP model reduces to the Type I bivariate ZTP model. In other words, the number of children in the household is independent of the number of the doctor office/clinic and health center visits under the bivariate Poisson structure but is correlated with the latter through the missing zero-vector. With $G = 6,000$ bootstrap replications, the means, standard deviations and two 95% bootstrap CIs of $(\lambda_0, \boldsymbol{\lambda})$ are showed in the other columns of Table 2. Table 2 shows that the bootstrap means are very close to the MLEs and the bootstrap standard deviations are relatively small, indicating that estimates from the parametric bootstrap methods in Section 3.2 are stable.

Table 2 MLEs and CIs of parameters for the health center visit data of California in USA (Gurmu, 1997)

Parameter	MLE	Mean	std^{B}	95% bootstrap $\text{CI}^{\dagger}$	95% bootstrap $\text{CI}^{\ddagger}$
λ_0	0.0000	0.0006	0.0129	[−0.0253, 0.0253]	[5.3028×10^{-10}, 1.2781×10^{-9}]
λ_1	1.5738	1.5720	0.0589	[1.4583, 1.6893]	[1.4594, 1.6865]
λ_2	2.2126	2.2125	0.0696	[2.0761, 2.3490]	[2.0760, 2.3427]

std^{B}: The estimated standard deviation of the bootstrap samples, cf. (3.8).
$\text{CI}^{\dagger}$: Normal-based bootstrap CI, cf. (3.9).
$\text{CI}^{\ddagger}$: Non-normal-based bootstrap CI, cf. (3.10).

Based on the data in Table 1, we calculate the sample correlation coefficient matrix, which is given by

$$\mathbf{R} = \begin{pmatrix} 1.0000 & -0.0993 \\ -0.0993 & 1.0000 \end{pmatrix},$$

while the population correlation coefficient matrix $\boldsymbol{\rho}$, based on (C.2), is estimated to be

$$\hat{\boldsymbol{\rho}} = \begin{pmatrix} 1.0000 & -0.0453 \\ -0.0453 & 1.0000 \end{pmatrix}.$$

Note that $\hat{\boldsymbol{\rho}}$ is close to the sample correlation coefficient matrix $\mathbf{R}$. In addition, both the sample and the estimated population correlation between the number of the doctor office/clinic and health center visits and the number of children in the household are negative, indicating that the former decreases with the latter. This result is consistent with that obtained by Gurmu (1997).

5.1.2 Bayesian methods

In Bayesian analysis, we adopt independent Gamma$(1,1)$ as the prior distributions of $\{\lambda_i\}_{i=0}^2$, respectively. Using $\lambda_i^{(0)} = 1$ $(i = 0, 1, 2)$ as the initial values, the EM algorithm, specified by (3.13), (3.6) and (3.7) with $m = 2$, $m_0 = 0$, $t_j = 1$ and $\boldsymbol{y}_j = \boldsymbol{w}_j$, converged to the posterior modes in 181 iterations, which are listed in the second column of Table 3.

Based on (3.12), (3.14) and (3.15), we use the DA algorithm to generate $L = 60{,}000$ posterior samples of $(\lambda_0, \lambda_1, \lambda_2)$. 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 $(\lambda_0, \lambda_1, \lambda_2)$, which are given in Table 3. Figure 1 shows the corresponding posterior densities of $(\lambda_0, \lambda_1, \lambda_2)$ via a kernel density smoother and their histograms based on the second half posterior samples generated by the DA algorithm.

Table 3 Posterior estimates of parameters for the health center visit data of California (Gurmu, 1997)

Parameter	Posterior mode	Posterior mean	Posterior std	95% Bayesian credible interval
λ_0	0.0000	0.0161	0.0147	[0.0005, 0.0543]
λ_1	1.5702	1.5578	0.0635	[1.4317, 1.6814]
λ_2	2.2076	2.1947	0.0713	[2.0567, 2.3349]

[Insert Figure 1 here]

5.2 Physician office visit data

Deb and Trivedi (1997) used the physician office visit data to study the demand for medical care by elderly. Information of 4406 subjects (1778 male and 2628 female) in the United States aged between 66 and 109 were recorded. These individuals were covered by Medicare, a public insurance programme that offers substantial protection against health care costs. In their study, six mutually exclusive measures of utilization (visits of utilization, visits to a setting, visits to a non-physician in an office setting, visits to hospital outpatient setting, visits to a non-physician in an outpatient, visits to an emergency room) and the number of hospital stays were considered and were fitted by six finite mixture negative binomial count model, respectively. However, some of the number of health care visits may be correlated with each other, a multivariate discrete count model should be employed to fit them. In this subsection, let Y_1 denote the number of physician office visits, Y_2 denote the number of emergency room visits, Y_3 denote the number of hospitalizations. The sample means, standard deviations, minimums and maximums for Y_1, Y_2 and Y_3 are listed in Table 4. Note that the zero-vector $(0, 0, 0)^{\top}$ is not missing, it is reasonable to fit the data set by the the proposed multivariate ZAP model with dimension $m = 3$.

Table 4 Summary statistics for the physician office visit data (Deb and Trivedi, 1997)

Variable	Mean	Standard deviation	Minimum	Maximum
Y_1	5.7744	6.7592	0	89
Y_2	0.2635	0.7037	0	12
Y_3	0.2960	0.7564	0	8

5.2.1 Likelihood-based inferences

Let $\mathbf{y}_j = (Y_{1j}, Y_{2j}, Y_{3j})^{\top} \overset{\text{ind}}{\sim} \text{ZAP}(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, where Y_{1j}, Y_{2j}, Y_{3j} for $j = 1, \ldots, n\,(n = 4406)$. To find the MLEs of λ_0 and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^{\top}$, we choose $\lambda_0^{(0)} = 2$ and $\boldsymbol{\lambda}^{(0)} = 2 \cdot \mathbf{1}_3$ as their initial values. The MLEs of $(\lambda_0, \boldsymbol{\lambda})$ converged to $(\hat{\lambda}_0, \hat{\boldsymbol{\lambda}})$ as shown in the second column of Table 5 in 17 iterations for the EM algorithm (3.5)–(3.7) with $m = 3$ and $m_0 = 581$. With $G = 6{,}000$ bootstrap replications, the two 95% bootstrap CIs of $(\lambda_0, \boldsymbol{\lambda})$ are showed in the last two columns of Table 5. From the Table 5, we can see that bootstrap means are very close to the MLEs and the bootstrap standard deviations are relatively small, showing that estimates from the parametric bootstrap methods described in Section 3.2 are stable.

Based on the data, we calculate the sample correlation coefficient matrix, which is given by

$$\mathbf{R} = \begin{pmatrix} 1.0000 & 0.1587 & 0.2408 \\ 0.1587 & 1.0000 & 0.4761 \\ 0.2408 & 0.4761 & 1.0000 \end{pmatrix},$$

while the population correlation coefficient matrix $\boldsymbol{\rho}$, based on (C.2), is estimated to be

$$\hat{\boldsymbol{\rho}} = \begin{pmatrix} 1.0000 & 0.1754 & 0.1792 \\ 0.1754 & 1.0000 & 0.3603 \\ 0.1792 & 0.3603 & 1.0000 \end{pmatrix}.$$

Table 5 MLEs and CIs of parameters for the physician office visit data (Deb and Trivedi, 1997)

Parameter	MLE	Mean	std$^{\mathrm{B}}$	95% bootstrap CI†	95% bootstrap CI‡
φ	0.1319	0.1318	0.0051	[0.1219, 0.1418]	[0.1219, 0.1416]
λ_0	0.0937	0.0937	0.0058	[0.0824, 0.1051]	[0.0824, 0.1054]
λ_1	5.6123	5.6122	0.0391	[5.5356, 5.6891]	[5.5352, 5.6899]
λ_2	0.1671	0.1670	0.0073	[0.1528, 0.1814]	[0.1529, 0.1817]
λ_3	0.2026	0.2029	0.0078	[0.1873, 0.2179]	[0.1873, 0.2181]

std$^{\mathrm{B}}$: The estimated standard deviation of the bootstrap samples, cf. (3.8).
CI†: Normal-based bootstrap CI, cf. (3.9).
CI‡: Non-normal-based bootstrap CI, cf. (3.10).

First, we want to test whether the physician office visits data follow the proposed multivariate ZAP model or the Type I multivariate ZAP model, which is equivalent to testing the null hypothesis H_0: $\lambda_0 = 0$ against the alternative hypothesis H_1: $\lambda_0 > 0$. According to (3.17), we calculate the value of the likelihood ratio test statistic T_1, which is given by $t_1 = 1165.829$. Hence, from (3.18), we obtain $p_{\mathrm{v}1} = 0.0000 \ll \alpha = 0.05$. Thus, we should reject H_0 at the 0.05 significant level. In other words, $\mathbf{y}_j = (Y_{1j}, Y_{2j}, Y_{3j})^{\top}$ follows the proposed multivariate ZAP model, i.e., ZAP$(\varphi; \lambda_0, \lambda_1, \lambda_2, \lambda_3)$, rather than the Type I multivariate ZAP model, i.e., ZAP$^{(\mathrm{I})}(\varphi; \lambda_1, \lambda_2, \lambda_3)$. In other words, using the proposed multivariate ZAP model to fit this data set is reasonable and satisfactory.

Furthermore, we would like to see whether the physician office visits data set is zero-inflated or zero-deflated, that is to test the null hypothesis H_0: $\varphi = \mathrm{e}^{-\lambda_0 - \lambda_+}$ against the alternative H_1: $\varphi \neq e^{-\lambda_0-\lambda_+}$. According to (3.20), we calculate the value of the likelihood ratio test statistic T_2, and obtain $t_2 = 4086.0514$. Then from (3.21), we have $p_{v2} = 0.0000 \ll 0.01$. Thus, we should reject the H_0 at the significance level of 0.01. We further calculate the 95% nonnormal-based bootstrap CI of $\theta = \varphi - e^{-\lambda_0-\lambda_+}$, which is given by [0.1219, 0.1419]. Since the lower bound is larger than 0, we can conclude that $\mathbf{y}$ follows a multivariate ZIP model rather than a multivariate ZDP model.

5.2.2 Bayesian methods

In Bayesian analysis, we adopt Beta(1, 1) as the prior distribution of φ and independent Gamma(1, 1) as the prior distributions of $\{\lambda_i\}_{i=0}^4$, respectively. Using $\lambda_i^{(0)} = 2$ ($i = 0, 1, 2, 3$) as the initial values, the EM algorithm, specified by (3.13), (3.6) and (3.7) with $m = 3$ and $m_0 = 581$, converged to the posterior modes in 17 iterations, which are listed in the second column of Table 6.

Table 6 Posterior estimates of parameters for the physician office visit data (Deb and Trivedi, 1997)

Parameter	Posterior mode	Posterior mean	Posterior std	95% Bayesian credible interval
φ	0.131866	0.132014	0.0050809	[0.1223, 0.1421]
λ_0	0.093726	0.093940	0.0055216	[0.0834, 0.1050]
λ_1	5.610835	5.611122	0.0391270	[5.5345, 5.6876]
λ_2	0.167042	0.167355	0.0070389	[0.1538, 0.1814]
λ_3	0.202507	0.202718	0.0076845	[0.1879, 0.2181]

Based on (3.12), (3.14) and (3.15), we use the DA algorithm to generate $L = 60{,}000$ posterior samples of $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$. 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 $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$, which are given in Table 6. Figure 2 shows the corresponding posterior densities of $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$ via a kernel density smoother and their histograms based on the second half posterior samples generated by the DA algorithm.

[Insert Figure 2 here]

5.2.3 Model selection

We therefore concentrate on the comparison between the Type I and our proposed multivariate ZAP models by the AIC and BIC, based on the full likelihood function, in Table 7. We could see that both the AIC and BIC of the proposed multivariate ZAP model are less than those of the Type I multivariate ZAP model, indicating that fitting the physician office visits data with the proposed multivariate ZAP model is more appropriate than with the Type I multivariate ZAP model. This conclusion is consistent with that obtained by the first hypothesis testing in Section 5.2.1.

Table 7 Comparisons of AIC and BIC for the two multivariate ZAP models

Model	Criterion	
	AIC	BIC
Type I multivariate ZAP model	48577.43	48596.60
The proposed multivariate ZAP model	47413.60	47439.16

5.2.4 Regression analysis

In this subsection, we try to fit the data set by the multivariate ZAP regression model presented in Section 4. Three covariates (i.e., sex, the number of chronic conditions and age in years) are considered. Let V_1 denote sex, where $V_1 = 1$ for female and $V_1 = 0$ for male. Let V_2 denote the number of chronic conditions and V_3 denote age in years, which is divided by 100. The sample mean, standard deviation, minimum and maximum for the number of chronic conditions of the respondents are 1.5420, 1.3496, 0 and 8, respectively, while the mean age for such individuals are 74.0241. Let $\mathbf{v}_1 = \mathbf{v}_2 = (1, V_1, V_2)^\top$ and $\mathbf{v}_3 = (1, V_1, V_2, V_3)^\top$. Let $\mathbf{y}_j \overset{\text{ind}}{\sim} \text{ZAP}(\varphi_j\,; \lambda_0, \lambda_{1j}, \lambda_{2j}, \lambda_{3j})$ and $\boldsymbol{y}_j = (y_{1j}, y_{2j}, y_{3j})^\top$ denote the observed values of $\mathbf{y}_j = (Y_{1j}, Y_{2j}, Y_{3j})^\top$ for $j = 1, \ldots, n\,(n = 4406)$, where

$$\begin{cases} \log\left(\dfrac{\varphi_j}{1-\varphi_j}\right) = \boldsymbol{v}_{1j}^\top \boldsymbol{\beta}, \\ \log(\lambda_{0j}) = \boldsymbol{v}_{2j}^\top \boldsymbol{\eta}, \\ \log(\lambda_{ij}) = \boldsymbol{v}_{3j}^\top \boldsymbol{\gamma}_i, \qquad i = 1, \ldots, m\,(m = 3). \end{cases}$$

Table 8 MLEs and bootstrap CIs of regression coefficients for the physician office visit data (Deb and Trivedi, 1997)

Para	Coefficient	MLE	std$^{\rm B}$	95% bootstrap CI†	95% bootstrap CI‡	t-Statistic
φ	Constant	−0.8978	0.0780	[−1.0506, −0.7450]	[−1.0538, −0.7465]	−11.5168
	Sex(F)	−0.3888	0.0915	[−0.5681, −0.2094]	[−0.5660, −0.2061]	−4.2481
	Chronic	−0.6513	0.0483	[−0.7460, −0.5566]	[−0.7536, −0.5679]	−13.4839
λ_0	Constant	−2.9431	0.1360	[−3.2096, −2.6767]	[−3.2297, −2.6903]	−21.6485
	Sex(F)	−0.2139	0.1398	[−0.4880, 0.0601]	[−0.5005, 0.0671]	−1.5303
	Chronic	0.3322	0.0440	[0.2459, 0.4185]	[0.2436, 0.4147]	7.5464
λ_1	Constant	2.0482	0.0764	[1.8986, 2.1979]	[1.8920, 2.1935]	26.8245
	Sex(F)	0.0604	0.0133	[0.0344, 0.0864]	[0.0336, 0.0864]	4.5478
	Chronic	0.1505	0.0042	[0.1423, 0.1587]	[0.1424, 0.1584]	35.9358
	Age	−0.6427	0.1030	[−0.8446, −0.4408]	[−0.8444, −0.4359]	−6.2386
λ_2	Constant	−3.6622	0.4298	[−4.5046, −2.8198]	[−4.5646, −2.8095]	−8.5207
	Sex(F)	0.0916	0.0833	[−0.0717, 0.2548]	[−0.0683, 0.2605]	1.0992
	Chronic	0.2567	0.0269	[0.2040, 0.3093]	[0.2031, 0.3087]	9.5515
	Age	2.0920	0.5747	[0.9656, 3.2184]	[0.9546, 3.2737]	3.6401
λ_3	Constant	−3.7460	0.3766	[−4.4841, −3.0079]	[−4.4661, −2.9763]	−9.9469
	Sex(F)	−0.1182	0.0724	[−0.2602, 0.0238]	[−0.2575, 0.0266]	−1.6317
	Chronic	0.2897	0.0236	[0.2434, 0.3360]	[0.2433, 0.3329]	12.2529
	Age	2.4881	0.5001	[1.5079, 3.4682]	[1.4612, 3.4587]	4.9753

std$^{\rm B}$: The estimated standard deviation of the bootstrap samples, cf. (3.8).
CI†: Normal-based bootstrap CI, cf. (3.9).
CI‡: Non-normal-based bootstrap CI, cf. (3.10).

To calculate the MLEs of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$, we choose $(\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}) = (1, -1, -1)$, $(\eta_0^{(0)}, \eta_1^{(0)}, \eta_2^{(0)}) = (2, -1, 1)$ and $\boldsymbol{\gamma}_1^{(0)} = \boldsymbol{\gamma}_2^{(0)} = \boldsymbol{\gamma}_3^{(0)} = (-1, -1, 1, 1)^{\top}$ as their initial values. The MLEs of $(\boldsymbol{\beta}, \mu, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ converged to $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2, \hat{\boldsymbol{\gamma}}_3)$ as shown in the third column of Table 8 in 33 iterations for the EM algorithm specified by (4.6), (4.9)–(4.11). The two 95% bootstrap CIs of $(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$ are listed in the fifth and sixth columns of Table 8. The last column of Table 8 shows the corresponding t-statistics. From Table 8, we can see that (for $\alpha = 0.05$): (i) both sex and the number of chronic conditions are significant and negatively correlated with φ; (ii) sex does not have a significant effect on the common parameter for the multivariate Poisson part λ_0 while the number of chronic conditions does and when it increases one unit, λ_0 increases 33.22%; (iii) λ_1 increases with sex but decreases with age, both relationships are significant, while the number of chronic conditions has a significant positive effect on λ_1; (iv) all the three covariaes—sex, number of chronic conditions and age in years—are positively correlated with λ_2, however, only sex is insignificant; (v) except for the insignificant and negatively correlated covariate sex, the others are significant and are non-negatively correlated with λ_3.

6. Discussion

By employing the SR method, we in this paper introduced a new multivariate ZAP distribution based on a multivariate Poisson random vector. This new distribution is useful because it possesses the following advantages: (a) Some existing multivariate zero-inflated distributions are too complicated so that it is very difficult to apply them to the cases of dimension larger than 3. For instance, the multivariate zero-inflated Poisson model proposed by Li *et al.* (1999) contains too many parameters and the corresponding computing method for MLEs is not easy to implement for cases of dimension larger than 3. Furthermore, duo to its complexity, Li *et al.*'s multivariate ZIP distribution does not possess helpful distributional properties (see, Liu and Tian, 2015). However, by using the SR, our proposed multivariate ZAP distribution (i) has a simple structure relationship and is easy to apply to the high-dimensional cases; (ii) is easy to derive its distributional properties (see Appendices C & D). In addition, we provided efficient statistical methods including the calculation of the MLEs of parameters via the EM algorithm with explicit M- and E-step. (b) To our best knowledge, the majority of the existing references only focus on some kind of specific count data, e.g., univariate zero-truncated data, univariate zero-inflated data, or multivariate zero-inflated data. However, the proposed models in this paper aim at the construction of a general framework for fitting a class of zero-adjusted multivariate discrete data, with the zero-truncated, zero-deflated, zero-inflated data as special cases and thus has a wider application range. (c) It allows a more flexible correlation structure than the Type I multivariate ZAP model for some of the correlation coefficients could be positive while others could be negative. We have further provided systemic unified and efficient statistical inference methods including bootstrap CI construction, Bayesian calculation, and hypothesis testing on independency and whether truncation, deflation or inflation.

In this paper, we have also considered the multivariate ZAP regression model with covariates. In the real data analysis of physician office visit data, the number of chronic conditions and age in years are significantly contribute to all these three outcomes—the number of physician office visits, the number of emergency room visits, and the number of hospitalizations, while sex only has non-ignorable effect on the number of physician office visits. Note that the proposed regression model only available for fixed effects, one of the future studies is to consider the multivariate ZAP regression model with random/mixed effects.

Acknowledgments

The authors would like to thank an Associate Editor and the Editor for their helpful suggestions which led to some improvement over the earlier version of the paper. Yin LIU's research was fully supported by two grants from the National Natural Science Foundation of China (No. 11601524 and No. 61773401) and a grant from the Young Teachers Innovation Project of Zhongnan University of Economics and Law (Project No. 31541711204). Guo-Liang TIAN's research was fully supported by a grant from the National Natural Science Foundation of China (No. 11771199). The work of Man-Lai TANG was supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region (Project No. UGC/FDS14/P01/16). Kam Chuen Yuen's research was supported by the Seeding Funding Programme for Basic Research of the University of Hong Kong.

Appendix A. Three multivariate discrete distributions

A.1 Multivariate Poisson distribution

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

$$\Pr(\mathbf{x} = \boldsymbol{x}) = \mathrm{e}^{-(\lambda_0+\lambda_1+\cdots+\lambda_m)} \sum_{k=0}^{\min(\boldsymbol{x})} \frac{\lambda_0^k}{k!} \prod_{i=1}^m \frac{\lambda_i^{x_i-k}}{(x_i-k)!}, \tag{A.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$ is the realization of $\mathbf{x}$ and $\min(\boldsymbol{x}) \hat{=} \min(x_1, \ldots, x_m)$.

A.2 Conditional multivariate Poisson distribution

Let $\mathbf{x} = (X_1, \ldots, X_m)^\top \sim \text{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$. Partition the random vector $\mathbf{x}$ into two random sub-vectors $\mathbf{x}^{(1)} = (X_1, \ldots, X_r)^\top$ and $\mathbf{x}^{(2)} = (X_{r+1}, \ldots, X_m)^\top$. Given $\mathbf{x}^{(2)} = \boldsymbol{x}^{(2)}$, $\mathbf{x}^{(1)}$ is said to have an r-dimensional conditional Poisson distribution with parameters $\lambda_0 \geqslant 0$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{x}^{(1)}|\mathbf{x}^{(2)} \sim \text{CP}_r(\lambda_0, \lambda_1, \ldots, \lambda_m)$ with pmf

$$\Pr(\mathbf{x}^{(1)} = \boldsymbol{x}^{(1)}|\mathbf{x}^{(2)} = \boldsymbol{x}^{(2)}) = \frac{\mathrm{e}^{-(\lambda_1+\cdots+\lambda_r)} \displaystyle\sum_{k=0}^{\min(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})} \frac{\lambda_0^k}{k!} \prod_{i=1}^m \frac{\lambda_i^{x_i-k}}{(x_i-k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{x}^{(2)})} \frac{\lambda_0^k}{k!} \prod_{i=r+1}^m \frac{\lambda_i^{x_i-k}}{(x_i-k)!}}, \tag{A.2}$$

where $\boldsymbol{x}^{(1)} = (x_1, \ldots, x_r)^\top$ and $\boldsymbol{x}^{(2)} = (x_{r+1}, \ldots, x_m)^\top$.

A.3 Type I multivariate ZTP distribution

Definition 3 (Tian *et al.*, 2017). Let $\mathbf{x} = (X_1, \ldots, X_m)^\top$, where $X_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda_i)$ for $i = 1, \ldots, m$. A discrete random vector $\mathbf{w} = (W_1, \ldots, W_m)^\top$ is said to follow the Type I multivariate ZTP distribution with the parameter vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{w} \sim \text{ZTP}^{(\text{I})}(\lambda_1, \ldots, \lambda_m)$ or $\mathbf{w} \sim \text{ZTP}_m^{(\text{I})}(\boldsymbol{\lambda})$, if

$$\mathbf{x} \stackrel{\mathrm{d}}{=} U\,\mathbf{w} = \begin{cases} \mathbf{0}, & \text{with probability } \psi, \\ \mathbf{w}, & \text{with probability } 1-\psi, \end{cases} \tag{A.3}$$

where $U \sim \text{Bernoulli}\,(1-\psi)$ with $\psi = \mathrm{e}^{-\lambda_+}$, $\lambda_+ = \sum_{i=1}^m \lambda_i \mathrel{\hat{=}} \|\boldsymbol{\lambda}\|_1$, and $U \perp\!\!\!\perp \mathbf{w}$. ¶

Given $U = 1$, from (A.3), we can see that $\mathbf{x}$ and $\mathbf{w}$ have the same distribution; i.e., $\mathbf{w} \stackrel{\mathrm{d}}{=} \mathbf{x}|(U = 1)$, indicating that the components of $\mathbf{w}$ are "partially" (in the sense only giving $U = 1$ rather than giving the whole U) conditionally independent. Tian *et al.* (2017) have shown that

$$\mathrm{Corr}(W_i, W_j) = -\sqrt{\frac{\lambda_i\lambda_j}{(\mathrm{e}^{\lambda_+} - 1 - \lambda_i)(\mathrm{e}^{\lambda_+} - 1 - \lambda_j)}}, \qquad i \neq j. \tag{A.4}$$

Therefore, the unique source resulting in the correlation between any two components of $\mathbf{w}$ comes from the missing zero vector.

Appendix B. Univariate zero-adjusted Poisson distribution

A non-negative discrete random variable Y is said to have a ZAP distribution with parameters $\varphi \in [0, 1)$ and $\lambda > 0$, denoted by $Y \sim \mathrm{ZAP}(\varphi, \lambda)$, if

$$Y \stackrel{\mathrm{d}}{=} Z'W, \tag{B.1}$$

where $Z' \sim \text{Bernoulli}(1 - \varphi)$, $W \sim \mathrm{ZTP}(\lambda)$, and $Z' \perp\!\!\!\perp W$. It is clear that the pmf of $Y \sim \mathrm{ZAP}(\varphi, \lambda)$ is given by

$$\Pr(Y = y) = \varphi I(y = 0) + \left[(1 - \varphi)\frac{\lambda^y \mathrm{e}^{-\lambda}}{(1 - \mathrm{e}^{-\lambda})y!}\right] I(y \neq 0). \tag{B.2}$$

Several important special cases of (B.1) or (B.2) include

(a) If $\varphi = 0$, then $Y \stackrel{\mathrm{d}}{=} W \sim \mathrm{ZTP}(\lambda)$;

(b) If $\varphi \in (0, \mathrm{e}^{-\lambda})$, then Y follows the ZDP distribution with parameters φ and $\lambda > 0$, denoted by $Y \sim \mathrm{ZDP}(\varphi, \lambda)$;

(c) If $\varphi = \mathrm{e}^{-\lambda}$, then $Y \sim \mathrm{Poisson}(\lambda)$; and

(d) If $\varphi \in (\mathrm{e}^{-\lambda}, 1)$, then Y follows the ZIP distribution with parameters $\phi \hat{=} (\varphi - \mathrm{e}^{-\lambda})/(1-\mathrm{e}^{-\lambda})$ and $\lambda > 0$, denoted by $Y \sim \mathrm{ZIP}(\phi, \lambda)$. The SR of $Y \sim \mathrm{ZIP}(\phi, \lambda)$ is (Meng, 1997)

$$Y \stackrel{\mathrm{d}}{=} ZX,$$

where $Z \sim \mathrm{Bernoulli}(1-\phi)$, $X \sim \mathrm{Poisson}(\lambda)$, and $Z \perp\!\!\!\perp X$. The corresponding pmf is given by (Lambert, 1992)

$$\Pr(Y=y) = [\phi + (1-\phi)\mathrm{e}^{-\lambda}]I(y=0) + \left[(1-\phi)\frac{\lambda^y \mathrm{e}^{-\lambda}}{y!}\right] I(y \neq 0).$$

Appendix C. Properties of the proposed multivariate ZTP distribution

C.1 Moments and moment generating function

From the SR (2.1), it is easy to show that

$$\begin{cases} E(\mathbf{w}) & = & \dfrac{\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda}}{1-\psi}, \\ E(\mathbf{w}\mathbf{w}^\top) & = & \dfrac{\lambda_0 \cdot \mathbf{1}\mathbf{1}^\top + \mathrm{diag}(\boldsymbol{\lambda}) + (\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})^\top}{1-\psi}, \\ \mathrm{Var}(\mathbf{w}) & = & \dfrac{1}{1-\psi}\left[\lambda_0 \cdot \mathbf{1}\mathbf{1}^\top + \mathrm{diag}(\boldsymbol{\lambda}) - \dfrac{\psi}{1-\psi}(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})(\lambda_0 \cdot \mathbf{1} + \boldsymbol{\lambda})^\top\right], \end{cases} \tag{C.1}$$

where $\mathbf{1} = \mathbf{1}_m = (1, \ldots, 1)^\top$. Thus, for $i \neq j$, we have

$$\mathrm{Corr}(W_i, W_j) = \frac{\lambda_0 - (\lambda_0+\lambda_i)(\lambda_0+\lambda_j)\psi/(1-\psi)}{\sqrt{[\lambda_0+\lambda_i-(\lambda_0+\lambda_i)^2\psi/(1-\psi)]\,[\lambda_0+\lambda_j-(\lambda_0+\lambda_j)^2\psi/(1-\psi)]}}, \tag{C.2}$$

i.e., the correlation coefficient between W_i and W_j could be positive or negative depending on the values of the parameters λ_0 and $\boldsymbol{\lambda}$. In particular, if $\lambda_0 = 0$, we obtain (A.4) again. Furthermore, if $\lambda_i = \lambda_j = \lambda$, then $\mathrm{Corr}(W_i, W_j) = \lambda/(\lambda + 1 - \psi^{-1})$.

For any $r_1, \ldots, r_m \geqslant 0$, the mixed moments of $\mathbf{w}$ are given by

$$E\left(\prod_{i=1}^m W_i^{r_i}\right) = (1-\psi)^{-1} E\left(\prod_{i=1}^m X_i^{r_i}\right) = (1-\psi)^{-1} E\left[\prod_{i=1}^m (X_i^* + X_0^*)^{r_i}\right]. \tag{C.3}$$

Using the identity of $E(\xi) = E[E(\xi|U)]$, the mgf of $\mathbf{x}$ is

$$\begin{aligned} M_{\mathbf{x}}(\boldsymbol{t}) &= E[\exp(\boldsymbol{t}^\top \mathbf{x})] = E[\exp(U \cdot \boldsymbol{t}^\top \mathbf{w})] = E\Big\{E[\exp(U\boldsymbol{t}^\top \mathbf{w})|U]\Big\} \\ &= E[M_{\mathbf{w}}(U\boldsymbol{t})] = \psi M_{\mathbf{w}}(\mathbf{0}) + (1-\psi)M_{\mathbf{w}}(\boldsymbol{t}) = \psi + (1-\psi)M_{\mathbf{w}}(\boldsymbol{t}). \end{aligned}$$

Thus, the mgf of $\mathbf{w} \sim \mathrm{ZTP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$ is given by

$$\begin{aligned} M_{\mathbf{w}}(\boldsymbol{t}) &= \frac{M_{\mathbf{x}}(\boldsymbol{t}) - \psi}{1-\psi} = \frac{M_{X_0^*}(t_+)\prod_{i=1}^m M_{X_i^*}(t_i) - \psi}{1-\psi} \\ &= \frac{\exp(\sum_{i=1}^m \lambda_i \mathrm{e}^{t_i} + \lambda_0 \mathrm{e}^{t_+} - \lambda_+ - \lambda_0) - \mathrm{e}^{-\lambda_0 - \lambda_+}}{1 - \mathrm{e}^{-\lambda_0-\lambda_+}}, \end{aligned}$$

where $t_+ = \sum_{i=1}^m t_i$.

C.2 Marginal distributions

We first consider the marginal distribution of each component for the random vector following the proposed multivariate ZTP distribution. These results are showed in Theorem 1 below, indicating that each random component follows a ZDP distribution which is a special case of a ZAP distribution (B.2).

Theorem 1 (Marginal distribution of W_i). Let $\mathbf{w} = (W_1, \ldots, W_m)^\top \sim \mathrm{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$, then, the marginal distributions of W_i is

$$W_i \sim \mathrm{ZDP}(\varphi_i, \lambda_0 + \lambda_i), \qquad i = 1, \ldots, m, \tag{C.4}$$

where

$$\varphi_i = \frac{\mathrm{e}^{-\lambda_0-\lambda_i} - \mathrm{e}^{-\lambda_0-\lambda_+}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}}. \tag{C.5}$$

Proof of Theorem 1. If $w_i > 0$, then

$$\begin{aligned}
\Pr(W_i = w_i) &= \sum_{w_1=0}^{\infty} \cdots \sum_{w_{i-1}=0}^{\infty} \sum_{w_{i+1}=0}^{\infty} \cdots \sum_{w_m=0}^{\infty} \Pr(\mathbf{w} = \boldsymbol{w}) \\
&\overset{(2.2)}{=} \frac{(\lambda_0+\lambda_i)^{w_i} \mathrm{e}^{-\lambda_0-\lambda_i}}{(1-\mathrm{e}^{-\lambda_0-\lambda_+}) w_i!} \sum_{w_1=0}^{\infty} \cdots \sum_{w_{i-1}=0}^{\infty} \sum_{w_{i+1}=0}^{\infty} \cdots \sum_{w_m=0}^{\infty} f_i(\boldsymbol{w}) \\
&= \frac{(\lambda_0+\lambda_i)^{w_i} \mathrm{e}^{-\lambda_0-\lambda_i}}{(1-\mathrm{e}^{-\lambda_0-\lambda_+}) w_i!} \sum_{k=0}^{w_i} \binom{w_i}{k} \left(\frac{\lambda_0}{\lambda_0+\lambda_i}\right)^k \left(\frac{\lambda_i}{\lambda_0+\lambda_i}\right)^{w_i-k} \\
&= \frac{1-\mathrm{e}^{-\lambda_0-\lambda_i}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \cdot \frac{(\lambda_0+\lambda_i)^{w_i} \mathrm{e}^{-\lambda_0-\lambda_i}}{(1-\mathrm{e}^{-\lambda_0-\lambda_i}) w_i!}, \qquad \text{(C.6)}
\end{aligned}$$

where

$$f_i(\boldsymbol{w}) = \sum_{k=0}^{\min(\boldsymbol{w})} \binom{w_i}{k} \left(\frac{\lambda_0}{\lambda_0+\lambda_i}\right)^k \left(\frac{\lambda_i}{\lambda_0+\lambda_i}\right)^{w_i-k} \prod_{j=1, j\neq i}^{m} \frac{\lambda_j^{w_j-k} \mathrm{e}^{-\lambda_j}}{(w_j-k)!}.$$

Hence,

$$\begin{aligned}
\Pr(W_i = 0) &= 1 - \sum_{w_i=1}^{\infty} \Pr(W_i = w_i) \\
&\overset{(C.6)}{=} 1 - \frac{1}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \sum_{w_i=1}^{\infty} \frac{(\lambda_0+\lambda_i)^{w_i} \mathrm{e}^{-\lambda_0-\lambda_i}}{w_i!} \\
&= 1 - \frac{1-\mathrm{e}^{-\lambda_0-\lambda_i}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} = \frac{\mathrm{e}^{-\lambda_0-\lambda_i} - \mathrm{e}^{-\lambda_0-\lambda_+}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \\
&\hat{=} \varphi_i \in [0, \mathrm{e}^{-\lambda_0-\lambda_i}) \subset (0,1). \qquad \text{(C.7)}
\end{aligned}$$

By combining (C.7) with (C.6) and noting that a ZDP distribution is a special case of a ZAP distribution (B.2), we obtain $W_i \sim \mathrm{ZDP}(\varphi_i, \lambda_0 + \lambda_i), \quad i = 1, \ldots, m.$ □

Next, we consider the marginal distributions of $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$, where

$$\mathbf{w}^{(1)} = \begin{pmatrix} W_1 \\ \vdots \\ W_r \end{pmatrix}, \quad \mathbf{w}^{(2)} = \begin{pmatrix} W_{r+1} \\ \vdots \\ W_m \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \end{pmatrix}.$$

Similar to the proof of (C.4), we can obtain

$$\mathbf{w}^{(1)} \sim \mathrm{ZDP}(\varphi^{(1)}; \lambda_0, \lambda_1, \ldots, \lambda_r) \quad \text{and} \quad \mathbf{w}^{(2)} \sim \mathrm{ZDP}(\varphi^{(2)}; \lambda_0, \lambda_{r+1}, \ldots, \lambda_m), \qquad \text{(C.8)}$$

where

$$\varphi^{(k)} = \frac{\mathrm{e}^{-\lambda_0-\lambda_+^{(k)}} - \mathrm{e}^{-\lambda_0-\lambda_+}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \in \left(0,\ \mathrm{e}^{-\lambda_0-\lambda_+^{(k)}}\right) \subset (0,1), \quad k=1,2, \tag{C.9}$$

$\lambda_+^{(1)} = \sum_{i=1}^{r}\lambda_i$ and $\lambda_+^{(2)} = \sum_{i=r+1}^{m}\lambda_i$.

In fact, for any positive integers $i_1,\ldots,i_r$ satisfying $1 \leqslant i_1 < \cdots < i_r \leqslant m$, we have

$$\begin{pmatrix} W_{i_1} \\ \vdots \\ W_{i_r} \end{pmatrix} \sim \mathrm{ZDP}(\varphi^*; \lambda_0, \lambda_{i_1}, \ldots, \lambda_{i_r}), \tag{C.10}$$

where

$$\varphi^* = \frac{\mathrm{e}^{-\lambda_0-(\lambda_{i_1}+\cdots+\lambda_{i_r})} - \mathrm{e}^{-\lambda_0-\lambda_+}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \in \left(0,\ \mathrm{e}^{-\lambda_0-(\lambda_{i_1}+\cdots+\lambda_{i_r})}\right) \subset (0,1). \tag{C.11}$$

C.3 Conditional distributions

We first consider the conditional distribution of $\mathbf{w}^{(1)}|\mathbf{w}^{(2)}$. These results are summarized in Theorem 2 below, indicating that given $\boldsymbol{w}^{(2)} \neq \mathbf{0}_{m-r}$, $\mathbf{w}^{(1)}$ follows the r-dimensional conditional Poisson distribution $\mathrm{CP}_r(\lambda_0, \lambda_1, \ldots, \lambda_m)$, which is defined by (A.2).

Theorem 2 (Conditional distribution of $\mathbf{w}^{(1)}|\mathbf{w}^{(2)}$). Let $\mathbf{w} = (\mathbf{w}^{(1)\top}, \mathbf{w}^{(2)\top})^\top \sim \mathrm{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$, then, the conditional distribution of $\mathbf{w}^{(1)}|\mathbf{w}^{(2)}$ is given by

$$\mathbf{w}^{(1)}|(\mathbf{w}^{(2)} = \boldsymbol{w}^{(2)}) \begin{cases} \stackrel{\mathrm{d}}{=} & \mathbf{x}^{(1)}|\mathbf{x}^{(2)} \sim \mathrm{CP}_r(\lambda_0, \lambda_1, \ldots, \lambda_m), \quad \text{if } \boldsymbol{w}^{(2)} \neq \mathbf{0}_{m-r}, \\ \sim & \mathrm{ZTP}^{(\mathrm{I})}(\lambda_1, \ldots, \lambda_r), \quad \text{if } \boldsymbol{w}^{(2)} = \mathbf{0}_{m-r}, \end{cases} \tag{C.12}$$

where $\mathbf{x} = (\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top})^\top \sim \mathrm{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$.

Proof of Theorem 2. From (2.2), (C.8) and (2.5), the conditional distribution of $\mathbf{w}^{(1)}|\mathbf{w}^{(2)}$ is

$$\Pr\{\mathbf{w}^{(1)} = \boldsymbol{w}^{(1)}|\mathbf{w}^{(2)} = \boldsymbol{w}^{(2)}\} = \frac{f(\boldsymbol{w}; \lambda_0, \boldsymbol{\lambda})}{\Pr\{\mathbf{w}^{(2)} = \boldsymbol{w}^{(2)}\}}$$

$$= \frac{\dfrac{\mathrm{e}^{-\lambda_0-\lambda_+}}{1-\mathrm{e}^{-\lambda_0-\lambda_+}} \displaystyle\sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i-k}}{(w_i-k)!}}{\varphi^{(2)} I(\boldsymbol{w}^{(2)} = \mathbf{0}) + \left[\dfrac{1-\varphi^{(2)}}{1-\mathrm{e}^{-\lambda_0-\lambda_+^{(2)}}} \displaystyle\sum_{k=0}^{\min(\boldsymbol{w}^{(2)})} \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \prod_{i=r+1}^{m} \frac{\lambda_i^{w_i-k}\mathrm{e}^{-\lambda_i}}{(w_i-k)!}\right] I(\boldsymbol{w}^{(2)} \neq \mathbf{0})}. \tag{C.13}$$

We first consider Case I: $\boldsymbol{w}^{(2)} \neq \mathbf{0}$. Under Case I, it is possible that $\boldsymbol{w}^{(1)} = \mathbf{0}$ or $\boldsymbol{w}^{(1)} \neq \mathbf{0}$. From (C.13), it is easy to show that

$$\Pr(\mathbf{w}^{(1)} = \boldsymbol{w}^{(1)} | \mathbf{w}^{(2)} = \boldsymbol{w}^{(2)}) = \frac{\mathrm{e}^{-\lambda_+^{(1)}} \displaystyle\sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{w}^{(2)})} \frac{\lambda_0^k}{k!} \prod_{i=r+1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}},$$

indicating

$$\mathbf{w}^{(1)} | (\mathbf{w}^{(2)} \neq \mathbf{0}) \ \stackrel{\mathrm{d}}{=}\ \mathbf{x}^{(1)} | \mathbf{x}^{(2)} \sim \mathrm{CP}_r(\lambda_0, \lambda_1, \ldots, \lambda_m),$$

where $\mathbf{x} = (\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top})^\top \sim \mathrm{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$, see (A.2).

Case II: $\boldsymbol{w}^{(2)} = \mathbf{0}$. Under Case II, it is obviously that $\boldsymbol{w}^{(1)} \neq \mathbf{0}$. From (C.13), we have

$$\Pr\{\mathbf{w}^{(1)} = \boldsymbol{w}^{(1)} | \mathbf{w}^{(2)} = \mathbf{0}\} = \frac{1}{1 - \mathrm{e}^{-\lambda_+^{(1)}}} \prod_{i=1}^{r} \frac{\lambda_i^{w_i} \mathrm{e}^{-\lambda_i}}{w_i!},$$

implying $\mathbf{w}^{(1)} | (\mathbf{w}^{(2)} = \mathbf{0}) \sim \mathrm{ZTP}^{(\mathrm{I})}(\lambda_1, \ldots, \lambda_r)$. □

Next, we are interested in deriving the conditional distribution of $X_0^* | \mathbf{w}$, which will be used in the E-step of the EM algorithm (see, Section 3.1) and the I-step of the DA algorithm (see, Section 3.3.2). Note that (2.1) can be rewritten as

$$\mathbf{x} = (X_1, \ldots, X_m)^\top = (X_0^* + X_1^*, \ldots, X_0^* + X_m^*)^\top \stackrel{\mathrm{d}}{=} U\mathbf{w}, \tag{C.14}$$

where $\{X_i^*\}_{i=0}^m \stackrel{\mathrm{ind}}{\sim} \mathrm{Poisson}(\lambda_i)$. To derive the conditional distribution of $X_0^* | \mathbf{w}$ (which is summarized in Theorem 4 below), we first need to derive the conditional distribution of $X_0^* | (\mathbf{w}, U)$, which is given in Theorem 3 below.

Theorem 3 (Conditional distribution of $X_0^* | (\mathbf{w}, U)$). Let $(X_0^*, U, \mathbf{w})$ be specified by (C.14) and $\mathbf{w} \sim \mathrm{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$. Then, the conditional distribution of $X_0^* | (\mathbf{w}, U)$ is given by

$$X_0^* | (\mathbf{w} = \boldsymbol{w}, U = u) \sim \begin{cases} \mathrm{Finite}(l, q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(\boldsymbol{w})), & \text{if } u \neq 1, \\ \mathrm{Degenerate}(0), & \text{if } u = 0, \end{cases} \tag{C.15}$$

where $\text{Finite}(x_k, p_k; k = 0, 1, \ldots, K)$ denotes the general finite distribution and

$$q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}) = \frac{\dfrac{\lambda_0^l}{l!} \displaystyle\prod_{i=1}^{m} \frac{\lambda_i^{w_i - l}}{(w_i - l)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}} \tag{C.16}$$

for $l = 0, 1, \ldots, \min(\boldsymbol{w})$.

Proof of Theorem 3. When $U = 1$, the conditional distribution of $X_0^*|(\mathbf{w}, U)$ is given by

$$\begin{aligned} \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 1) &= \frac{\Pr(X_0^* = l, X_1^* = w_1 - l, \ldots, X_m^* = w_m - l)}{\Pr(X_1 = w_1, \ldots, X_m = w_m)} \\ &\overset{(A.1)}{=} \frac{\dfrac{\lambda_0^l}{l!} \displaystyle\prod_{i=1}^{m} \frac{\lambda_i^{w_i - l}}{(w_i - l)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}} \\ &= q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}), \end{aligned} \tag{C.17}$$

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

$$X_0^*|(\mathbf{w} = \boldsymbol{w}, U = 1) \sim \text{Finite}(l, q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(\boldsymbol{w})). \tag{C.18}$$

When $U = 0$, we obtain $\Pr(X_0^* = 0|\mathbf{w} = \boldsymbol{w}, U = 0) = 1$, i.e.,

$$X_0^*|(\mathbf{w} = \boldsymbol{w}, U = 0) \sim \text{Degenerate}(0). \tag{C.19}$$

Hence, for any l, we have

$$\Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 0) = I(l = 0). \tag{C.20}$$

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

Theorem 4 (Conditional distribution of $X_0^*|\mathbf{w}$). Let $(X_0^*, \mathbf{w})$ be specified by (C.14) and $\mathbf{w} \sim \mathrm{ZTP}_m(\lambda_0, \boldsymbol{\lambda})$. Then, the conditional distribution of $X_0^*|\mathbf{w}$ is given by

$$X_0^*|(\mathbf{w} = \boldsymbol{w}) \sim \mathrm{Finite}(l, p_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(\boldsymbol{w})), \tag{C.21}$$

where

$$p_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}) = \mathrm{e}^{-\lambda_0-\lambda_+} \cdot I(l = 0) + (1 - \mathrm{e}^{-\lambda_0-\lambda_+}) \cdot q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}), \tag{C.22}$$

$I(\cdot)$ is the indicator function and $q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda})$ is defined by (C.16). And more importantly, the conditional expectation of $X_0^*|\mathbf{w}$ is given by

$$E(X_0^*|\mathbf{w} = \boldsymbol{w}) = (1 - \mathrm{e}^{-\lambda_0-\lambda_+}) \cdot \frac{\displaystyle\sum_{k=1}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{(k-1)!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i-k}}{(w_i-k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i-k}}{(w_i-k)!}} \cdot I(\min(\boldsymbol{w}) \geqslant 1). \tag{C.23}$$

Proof of Theorem 4. By using (C.19) and (C.18), the conditional distribution of $X_0^*|\mathbf{w}$ is

$$\begin{aligned}
\Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}) &= \sum_{u=0}^{1} \Pr(X_0^* = l, U = u|\mathbf{w} = \boldsymbol{w}) \\
&= \sum_{u=0}^{1} \Pr(U = u|\mathbf{w} = \boldsymbol{w}) \cdot \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = u) \\
&= \Pr(U = 0) \cdot \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 0) \\
&\quad + \Pr(U = 1) \cdot \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 1) \\
&\overset{(C.20)}{=} \mathrm{e}^{-\lambda_0-\lambda_+} \cdot I(l = 0) + (1 - \mathrm{e}^{-\lambda_0-\lambda_+}) \cdot q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}) \\
&= p_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}),
\end{aligned} \tag{C.24}$$

for $l = 0, 1, \ldots, \min(\boldsymbol{w})$, where $q_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda})$ is defined by (C.4). Thus,

$$X_0^*|(\mathbf{w} = \boldsymbol{w}) \sim \mathrm{Finite}(l, p_l(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(\boldsymbol{w})). \tag{C.25}$$

Especially, when $\min(\boldsymbol{w}) = 0$, we have $X_0^*|(\mathbf{w} = \boldsymbol{w}) \sim \text{Degenerate}(0)$. Thus, the conditional expectation of $X_0^*|\mathbf{w}$ is given by

$$E(X_0^*|\mathbf{w} = \boldsymbol{w}) = (1 - \mathrm{e}^{-\lambda_0 - \lambda_+}) \cdot \frac{\displaystyle\sum_{k=1}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{(k-1)!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{w})} \frac{\lambda_0^k}{k!} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}} \cdot I(\min(\boldsymbol{w}) \geqslant 1). \qquad \square$$

Appendix D. Properties of the proposed multivariate ZAP distribution

D.1 Marginal distributions

Now we consider the marginal distributions of $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$, where

$$\mathbf{y}^{(1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_r \end{pmatrix}, \quad \mathbf{y}^{(2)} = \begin{pmatrix} Y_{r+1} \\ \vdots \\ Y_m \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix}.$$

Based on (2.4) and (C.8), we have

$$\mathbf{y}^{(k)} \stackrel{\mathrm{d}}{=} Z'\mathbf{w}^{(k)} \stackrel{\mathrm{d}}{=} Z'Z^{(k)}\boldsymbol{\xi}^{(k)}, \quad k = 1, 2,$$

where $Z' \sim \text{Bernoulli}(1 - \varphi)$, $Z^{(k)} \sim \text{Bernoulli}(1 - \varphi^{(k)})$, $\varphi^{(k)}$ is given by (C.9), $\boldsymbol{\xi}^{(1)} \sim \text{ZTP}(\lambda_0, \lambda_1, \ldots, \lambda_r)$ and $\boldsymbol{\xi}^{(2)} \sim \text{ZTP}(\lambda_0, \lambda_{r+1}, \ldots, \lambda_m)$. Note that $Z'Z^{(k)} \perp\!\!\!\perp \boldsymbol{\xi}^{(k)}$ and $Z'Z^{(k)} \sim \text{Bernoulli}((1 - \varphi)(1 - \varphi^{(k)}))$. According to the SR (2.4), we can obtain

$$\mathbf{y}^{(1)} \sim \text{ZAP}(\nu^{(1)}; \lambda_0, \lambda_1, \ldots, \lambda_r) \quad \text{and} \quad \mathbf{y}^{(2)} \sim \text{ZAP}(\nu^{(2)}; \lambda_0, \lambda_{r+1}, \ldots, \lambda_m), \tag{D.1}$$

where

$$\nu^{(k)} = 1 - (1 - \varphi)(1 - \varphi^{(k)}) = 1 - (1 - \varphi)\frac{1 - \mathrm{e}^{-\lambda_0 - \lambda_+^{(k)}}}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+}} \in (0, 1), \quad k = 1, 2, \tag{D.2}$$

$\lambda_+^{(1)} = \sum_{i=1}^{r} \lambda_i$ and $\lambda_+^{(2)} = \sum_{i=r+1}^{m} \lambda_i$.

In fact, for any positive integers $i_1, \ldots, i_r$ satisfying $1 \leqslant i_1 < \cdots < i_r \leqslant m$, we have

$$\begin{pmatrix} Y_{i_1} \\ \vdots \\ Y_{i_r} \end{pmatrix} \sim \text{ZAP}(\nu^*; \lambda_0, \lambda_{i_1}, \ldots, \lambda_{i_r}), \tag{D.3}$$

where φ^* is given by (C.11) and

$$\nu^* = 1 - (1-\varphi)(1-\varphi^*) = 1 - (1-\varphi)\frac{1 - \mathrm{e}^{-\lambda_0 - (\lambda_{i_1} + \cdots + \lambda_{i_r})}}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+}} \in (0, 1). \tag{D.4}$$

D.2 Conditional distributions

D.2.1 Conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$

Theorem 5 (Conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$). Let $\mathbf{y} = (\mathbf{y}^{(1)\top}, \mathbf{y}^{(2)\top})^\top \sim \mathrm{ZAP}(\varphi; \lambda_0, \lambda_1, \ldots, \lambda_m)$, then, the conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$ is given by

$$\mathbf{y}^{(1)}|(\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}) \begin{cases} \stackrel{\mathrm{d}}{=} & \mathbf{x}^{(1)}|\mathbf{x}^{(2)} \sim \mathrm{CP}_r(\lambda_0, \lambda_1, \ldots, \lambda_m), \quad \text{if } \boldsymbol{y}^{(2)} \neq \mathbf{0}_{m-r}, \\ \sim & \mathrm{ZAP}^{(\mathrm{I})}(\tau^{(2)}; \lambda_1, \ldots, \lambda_r), \qquad\qquad \text{if } \boldsymbol{y}^{(2)} = \mathbf{0}_{m-r}, \end{cases} \tag{D.5}$$

where $\mathbf{x} = (\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top})^\top \sim \mathrm{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$, $\tau^{(2)} \hat{=} \frac{\varphi}{\nu^{(2)}}$, and $\nu^{(2)}$ is defined by (D.2).

Proof of Theorem 5. From (2.5) and (D.1), the conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$ is

$$\Pr\{\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}\} = \frac{\Pr(\mathbf{y} = \boldsymbol{y})}{\Pr\{\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}\}}$$

$$= \frac{\varphi I(\boldsymbol{y} = \mathbf{0}) + \left[\dfrac{1-\varphi}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+}} \displaystyle\sum_{k=0}^{\min(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})} \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \prod_{i=1}^m \frac{\lambda_i^{y_i - k} \mathrm{e}^{-\lambda_i}}{(y_i - k)!}\right] I(\boldsymbol{y} \neq \mathbf{0})}{\nu^{(2)} I(\boldsymbol{y}^{(2)} = \mathbf{0}) + \left[\dfrac{1 - \nu^{(2)}}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+^{(2)}}} \displaystyle\sum_{k=0}^{\min(\boldsymbol{y}^{(2)})} \frac{\lambda_0^k \mathrm{e}^{-\lambda_0}}{k!} \prod_{i=r+1}^m \frac{\lambda_i^{y_i - k} \mathrm{e}^{-\lambda_i}}{(y_i - k)!}\right] I(\boldsymbol{y}^{(2)} \neq \mathbf{0})}. \tag{D.6}$$

We first consider Case I: $\boldsymbol{y}^{(2)} \neq \mathbf{0}$. Under Case I, it is clear that $\boldsymbol{y} \neq \mathbf{0}$. From (D.6), it is easy to obtain

$$\Pr\{\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}\} = \frac{\mathrm{e}^{-\lambda_+^{(1)}} \displaystyle\sum_{k=0}^{\min(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})} \frac{\lambda_0^k}{k!} \prod_{i=1}^m \frac{\lambda_i^{y_i - k}}{(y_i - k)!}}{\displaystyle\sum_{k=0}^{\min(\boldsymbol{y}^{(2)})} \frac{\lambda_0^k}{k!} \prod_{i=r+1}^m \frac{\lambda_i^{y_i - k}}{(y_i - k)!}},$$

indicating

$$\mathbf{y}^{(1)}|(\mathbf{y}^{(2)} \neq \mathbf{0}) \stackrel{\mathrm{d}}{=} \mathbf{x}^{(1)}|\mathbf{x}^{(2)} \sim \mathrm{CP}_r(\lambda_0, \lambda_1, \ldots, \lambda_m),$$

where $\mathbf{x} = (\mathbf{x}^{(1)\top}, \mathbf{x}^{(2)\top})^\top \sim \mathrm{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$, see (A.2).

Case II: $\boldsymbol{y}^{(2)} = \mathbf{0}$. Under Case II, it is possible that $\boldsymbol{y}^{(1)} = \mathbf{0}$ or $\boldsymbol{y}^{(1)} \neq \mathbf{0}$. When $\boldsymbol{y}^{(1)} = \mathbf{0}$, from (D.6), we obtain

$$\Pr\{\mathbf{y}^{(1)} = \mathbf{0} | \mathbf{y}^{(2)} = \mathbf{0}\} = \frac{\varphi}{\nu^{(2)}}. \tag{D.7}$$

When $\boldsymbol{y}^{(1)} \neq \mathbf{0}$, from (D.6), we have

$$\Pr\{\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)} | \mathbf{y}^{(2)} = \mathbf{0}\} = \frac{(1-\varphi)\mathrm{e}^{-\lambda_0 - \lambda_+^{(2)}}}{\nu^{(2)}(1 - \mathrm{e}^{-\lambda_0 - \lambda_+})} \prod_{i=1}^{r} \frac{\lambda_i^{y_i} \mathrm{e}^{-\lambda_i}}{y_i!} = \frac{1 - \varphi/\nu^{(2)}}{1 - \mathrm{e}^{-\lambda_+^{(1)}}} \prod_{i=1}^{r} \frac{\lambda_i^{y_i} \mathrm{e}^{-\lambda_i}}{y_i!}. \tag{D.8}$$

By combining (D.7) with (D.8), we obtain

$$\mathbf{y}^{(1)} | (\mathbf{y}^{(2)} = \mathbf{0}) \sim \mathrm{ZAP}^{(\mathrm{I})}(\tau^{(2)}; \lambda_1, \ldots, \lambda_r),$$

where $\tau^{(2)} \hat{=} \varphi/\nu^{(2)}$. □

D.2.2 Conditional distribution of $Z'|\mathbf{y}$

Since $Z' \sim \mathrm{Bernoulli}(1-\varphi)$, Z' only takes the value 0 or 1. Note that $\boldsymbol{y} = \mathbf{0}$ is equivalent to $Z' = 0$. Thus, $\Pr(Z' = 0 | \mathbf{y} = \mathbf{0}) = \Pr(Z' = 0)/\Pr(\mathbf{y} = \mathbf{0}) = 1$. And when $\boldsymbol{y} \neq \mathbf{0}$, we have $\Pr(Z' = 1 | \mathbf{y} = \boldsymbol{y}) = \Pr(Z' = 1, \mathbf{w} = \boldsymbol{y})/\Pr(\mathbf{y} = \boldsymbol{y}) = 1$. Therefore,

$$Z' | (\mathbf{y} = \boldsymbol{y}) \sim \begin{cases} \mathrm{Degenerate}(0), & \text{if } \boldsymbol{y} = \mathbf{0}, \\ \mathrm{Degenerate}(1), & \text{if } \boldsymbol{y} \neq \mathbf{0}, \end{cases} \tag{D.9}$$

i.e., $Z' | (\mathbf{y} = \boldsymbol{y}) \sim \mathrm{Degenerate}(I(\boldsymbol{y} \neq \mathbf{0}))$.

D.2.3 Conditional distribution of $\mathbf{w} | (\mathbf{y} = \boldsymbol{y} \neq \mathbf{0})$

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

$$\Pr(\mathbf{w} = \boldsymbol{w} | \mathbf{y} = \boldsymbol{y}) = \frac{\Pr(\mathbf{w} = \boldsymbol{w}, \mathbf{y} = \boldsymbol{y})}{\Pr(\mathbf{y} = \boldsymbol{y})} = \frac{\Pr(\mathbf{w} = \boldsymbol{y}, Z' = 1)}{\Pr(\mathbf{y} = \boldsymbol{y})} = I(\boldsymbol{w} = \boldsymbol{y}).$$

Thus, given $\mathbf{y} = \boldsymbol{y} \neq \mathbf{0}$, we have

$$\mathbf{w} | (\mathbf{y} = \boldsymbol{y} \neq \mathbf{0}) \sim \mathrm{Degenerate}(\boldsymbol{y}). \tag{D.10}$$

D.2.4 Conditional distribution of $W_i|(Y_i = y_i = 0),\ i = 1, \ldots, m$

Theorem 6 (Conditional distribution of $W_i|(Y_i = y_i = 0)$). The conditional distribution of $W_i|(Y_i = y_i = 0)$ is given by

$$W_i|(Y_i = 0) \sim \mathrm{ZAP}(\tau_i, \lambda_0 + \lambda_i), \qquad i = 1, \ldots, m, \tag{D.11}$$

where $\tau_i \ \hat{=}\ \varphi_i/\nu_i$, φ_i is defined by (C.5), and

$$\nu_i \,\hat{=}\, 1 - (1-\varphi)(1-\varphi_i) \overset{(C.5)}{=} 1 - (1-\varphi)\frac{1 - \mathrm{e}^{-\lambda_0 - \lambda_i}}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+}}. \tag{D.12}$$

Proof of Theorem 6. For $i = 1, \ldots, m$, from (2.4), we know the relationship between Y_i and W_i is $Y_i \overset{\mathrm{d}}{=} Z'W_i$, where $W_i \sim \mathrm{ZDP}(\varphi_i, \lambda_0 + \lambda_i)$ with

$$\varphi_i = (\mathrm{e}^{-\lambda_0 - \lambda_i} - \mathrm{e}^{-\lambda_0 - \lambda_+})/(1 - \mathrm{e}^{-\lambda_0 - \lambda_+}),$$

see (C.13) and (C.5), and $Y_i \sim \mathrm{ZAP}(\nu_i, \lambda_0 + \lambda_i)$ with

$$\nu_i = 1 - (1-\varphi)\frac{1 - \mathrm{e}^{-\lambda_0 - \lambda_i}}{1 - \mathrm{e}^{-\lambda_0 - \lambda_+}},$$

see (D.3) and (D.12). Thus,

$$\begin{aligned}
\Pr(W_i = w_i | Y_i = 0) &= \frac{\Pr(W_i = w_i, Y_i = 0)}{\Pr(Y_i = 0)} \\
&= \frac{\Pr(W_i = 0, Y_i = 0)}{\nu_i} I(w_i = 0) + \frac{\Pr(W_i = w_i, Z' = 0)}{\nu_i} I(w_i \neq 0) \\
&= \frac{\Pr(W_i = 0)}{\nu_i} I(w_i = 0) + \frac{\varphi \Pr(W_i = w_i)}{\nu_i} I(w_i \neq 0) \\
&= \frac{\varphi_i}{\nu_i} I(w_i = 0) + \frac{\varphi(1-\varphi_i)}{\nu_i} \frac{(\lambda_0 + \lambda_i)^{w_i} \mathrm{e}^{-\lambda_0 - \lambda_i}}{w_i!(1 - \mathrm{e}^{-\lambda_0 - \lambda_i})} I(w_i \neq 0) \\
&= \tau_i I(w_i = 0) + (1 - \tau_i) \frac{(\lambda_0 + \lambda_i)^{w_i} \mathrm{e}^{-\lambda_0 - \lambda_i}}{w_i!(1 - \mathrm{e}^{-\lambda_0 - \lambda_i})} I(w_i \neq 0),
\end{aligned}$$

i.e., $W_i|(Y_i = 0) \sim \mathrm{ZAP}(\tau_i, \lambda_0 + \lambda_i)$, where $\tau_i \ \hat{=}\ \varphi_i/\nu_i$. □

D.2.5 Conditional distribution of $W_i|(Y_i = y_i > 0),\ i = 1, \ldots, m$

From (D.3), we have $Y_i \sim \text{ZAP}(\nu_i, \lambda_0 + \lambda_i)$, where ν_i is defined by (D.12). Since

$$\Pr(W_i = w_i|Y_i = y_i > 0) = \frac{\Pr(W_i = w_i, Y_i = y_i)}{\Pr(Y_i = y_i)} = \frac{\Pr(W_i = y_i, Z' = 1)}{\Pr(Y_i = y_i)} = I(w_i = y_i).$$

we obtain $W_i|(Y_i = y_i > 0) \sim \text{Degenerate}(y_i)$.

References

Aitchison, J. and Ho, C.H. (1989). The multivariate Poisson-log normal distribution. *Biometrika* **76**: 643–653.

Alfò, M. and Aitkin, M. (2006). Variance component models for longitudinal count data with baseline information: Epilepsy data revisited. *Statistics and Computing* **16**, 231–238.

Alfò, M. and Maruotti, A. (2010). Two-part regression models for longitudinal zero-inflated count data. *The Canadian Journal of Statistics* **38**, 197–216.

Alfò, M. and Trovato, G. (2004). Semiparametric mixture models for multivariate count data, with application. *The Econometrics Journal* **7**, 1–29.

Alfò, M., Maruotti, A. and Trovato, G. (2011). A finite mixture model for multivariate counts under endogenous selectivity. *Statistics and Computing* **21**, 185–202.

Angers, J.F. and Biawas, A. (2003). A Bayesian analysis of zero-inflated generalized Poisson model. *Computational Statistics & Data Analysis* **42**, 37–46.

Bago d'Uva, T. (2006). Latent class model for utilisation of health care. *Health Economics* **15**, 329–343.

Belloc, F., Bernardi, M., Maruotti, A. and Petrella, L. (2013) A dynamic hurdle model for zero-inflated panel count data. *Applied Economics Letters* **20**, 837–841.

Best, D.J., Rayner, J.C.W. and Thas, O. (2007). Goodness of fit for the zero-truncated Poisson distribution. *Journal of Statistical Computation and Simulation* **77**(7), 585–591.

Böhning, D., Dietz, E., Schlattmann, P., Mendonca, L. and Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A* **162**, 195–209.

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

Campbell, J.T. (1934). The Poisson correlation function. *Proceedings of the Edinburgh Mathematical Society* (Ser. 2) **4**, 18–26.

Charalambides, C.A. (1984). Minimum variance unbiased estimation for the zero class truncated bivariate poisson and logarithmic series distributions. *Metrika* **31**(1), 115–123.

Cheung, Y. (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine* **21**, 1461–1469.

Cohen, A.C., Jr. (1954). Estimation of the Poisson parameter from truncated samples and from censored samples. *Journal of the American Statistical Association* **49**, 158–168.

Cohen, A.C., Jr. (1960a). Estimating the parameter in a conditional Poisson distribution. *Biometrics* **16**(2), 203–211.

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

Consul, P.C. (1989). *Generalized Poisson Distributions.* Marcel Dekker, New York.

Cui, Y. and Yang, W. (2009). Zero-inflated generalized Poisson regression mixture model for mapping quantitative trait loci underlying count trait with many zeros. *Journal of theoretical biology* **256**, 276–285.

Dahiya, R.C. (1977). Estimation in a truncated bivariate Poisson distribution. *Communications in Statistics—Theory and Methods* **6**(2), 113–120.

Dahiya, R.C. and Gross, A.J. (1973). Estimating the zero class from a truncated Poisson sample. *Journal of the American Statistical Association* **68**, 731–733.

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

Deb, P. and Trivedi P.K.(1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics* **12**: 313–326.

Deng, D., Paul, S. (2000). Score tests for zero inflation in generalized linear models. *Canadian Journal of Statistics* **27**, 563–570.

Deng, D., Paul, S. (2005). Score tests for zero-inflation and over-dispersion in generalized linear models. *Statistica Sinica* **15**, 257–276.

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.

Famoye, F., Singh, K. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science* **4**, 117–130.

Finney, D.J. (1949). The truncated binomial distribution. *Annals of Eugenics* **14**, 319–328.

Finney, D.J. and Varley, G.C. (1955). An example of the truncated Poisson distribution. *Biometrics* **11**(3), 387–394.

Goraski, A. (1977). Distribution z-Poisson. *Publications de lInstitut de statistique de lUniversité de Paris* **12**, 45–53.

Grogger, J.T. and Carson, R.T. (1991). Models for truncated counts. *Journal of Applied Econometrics* **6**(3), 225–238.

Gupta, R.C. (1974). Distribution of the sum of independent decapitated generalized negative binomial variables. *Sankhyā: The Indian Journal of Statistics, Series B* **36**(1), 67–69.

Gurmu, S. (1991). Tests for detecting overdispersion in the positive Poisson regression model. *Journal of Business & Economic Statistics* **9**(2), 215–222.

Gurmu, S. (1997). Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *Journal of Applied Econometrics* **12**(3), 225–242.

Hall, D. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics* **56**, 1030–39.

Hall, D. and Zhang, Z. (2004). Marginal models for zero inflated clustered data. *Statistical Modelling* **4**, 161–180.

Hamdan, M.A. (1972). Estimation in the truncated bivariate Poisson distribution. *Tech-*

nometrics **14**(1), 37–45.

Hartley, H.J. (1958). Maximum likelihood estimation from incomplete data. *Biometrics* **14**(2), 174–194.

Irwin, J.O. (1959). On the estimation of the mean of a Poisson distribution from a sample with the zero class missing. *Biometrics* **15**(2), 324–326.

Johnson, N., Kotz, S., 1969. *Distributions in Statistics: Discrete Distributions.* Houghton Mifflim, Boston.

Jung, B.C., Han, S.M. and Lee, J. (2007). Score tests for testing independence in the zero-truncated bivariate Poisson models. *Communications in Statistics—Theory and Methods* **36**, 599–611.

Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics* **30**(1), 63–77.

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.

Kemp, A.W. (1986). Weighted discrepancies and maximum likelihood estimation for discrete distribution. *Communications in Statistics—Theory and Methods* **15**, 783–803.

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

Lee, A. H., Wang, K., Scott, J. A, Yau, K. K. W., and McLachlan, G. J. (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research* **15**, 47–61.

Li, C.S. (2012). Score tests for semiparametric zero-inflated Poisson models. *International Journal of Statistics and Probability* **1**(2), 1–7.

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

Martin, D., Katti, S. (1965). Fitting of some contagious distributions to some available data by the maximum likelihood method. *Biometrics* **21**, 34–48.

Maruotti, A. and Raponi, V. (2014). On baseline conditions for zero-inflated longitudinal count data. *Communications in StatisticsSimulation and Computation* **43**, 1–18.

Medhi, J. (1975). On the convolutions of left-truncated generalised negative binomial and Poisson variables. *Sankhyā: The Indian Journal of Statistics, Series B* **37**(3), 293–299.

Meng, X.L. (1997). The EM algorithm and medical studies: A historical linik. *Statistical Methods in Medical Research* **6**, 3–23.

Min, Y. and Agresti, A. (2005).Random effect models for repeatedmeasures of zero-inflated count data. *Statistical Modelling* **5**, 1–19.

Min, A. and Czado, C. (2010). Testing for zero-modification in count regression models. *Statistica Sinica* **20**, 323–341.

Moore, P.G. (1952). The estimation of the Poisson parameter from a truncated distribution. *Biometrika* **39**, 247–251.

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

Neelon, B.H., OMalley, A.J. and Normand, S.T. (2010). A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical Modelling* **10**(4), 421–439.

Piperigou, V.E. and Papageorgiou, H. (2003). On truncated bivariate discrete distributions: A unified treatment. *Metrika* **58**, 221–233.

Rao, C. and Chakravarti, I.M. (1956). Some small sample tests of significance for a Poisson distribution. *Biometrics* **12**(3), 264–282.

Rider, P.R. (1953). Truncated Poisson distributions. *Journal of the American Statistical Association* **48**, 826–830.

Rider, P.R. (1955). Truncated binomial and negative binomial distributions. *Journal of the American Statistical Association* **50**, 877–833.

Ridout, M., Hinde, J. and DeméAtrio, C.G.B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* **57**(1), 219–223.

Sampford, M.R. (1955). The truncated negative binomial distribution. *Biometrika* **42**(1/2), 58–69.

Shanmugam, R. (1985). An intervened Poisson distribution and its medical application. *Biometrics* **41**, 1025–1029.

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

Tian, G.L., Liu, Y., Tang, M.L. and Jiang, X.J. (2017). Type I multivariate zero-truncated /adjusted Poisson distributions with applications. Submitted to *Journal of Computational and applied Mathematics*.

Wang, L. (2010). IRT-ZIP modeling for multivariate zero-inflated count data. *Journal of Educational and Behavioral Statistics* **35**(6), 671–692.

Xie, F.C., Wei, B.C. and Lin, J.G. (2009). Score tests for zero-inflated generalized Poisson mixed regression models. *Computational Statistics & Data Analysis* **53**(9) 3478–3489.

Yau, K.K.W., Yang, K. and Lee, A.H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal* **45**(4), 437–452.

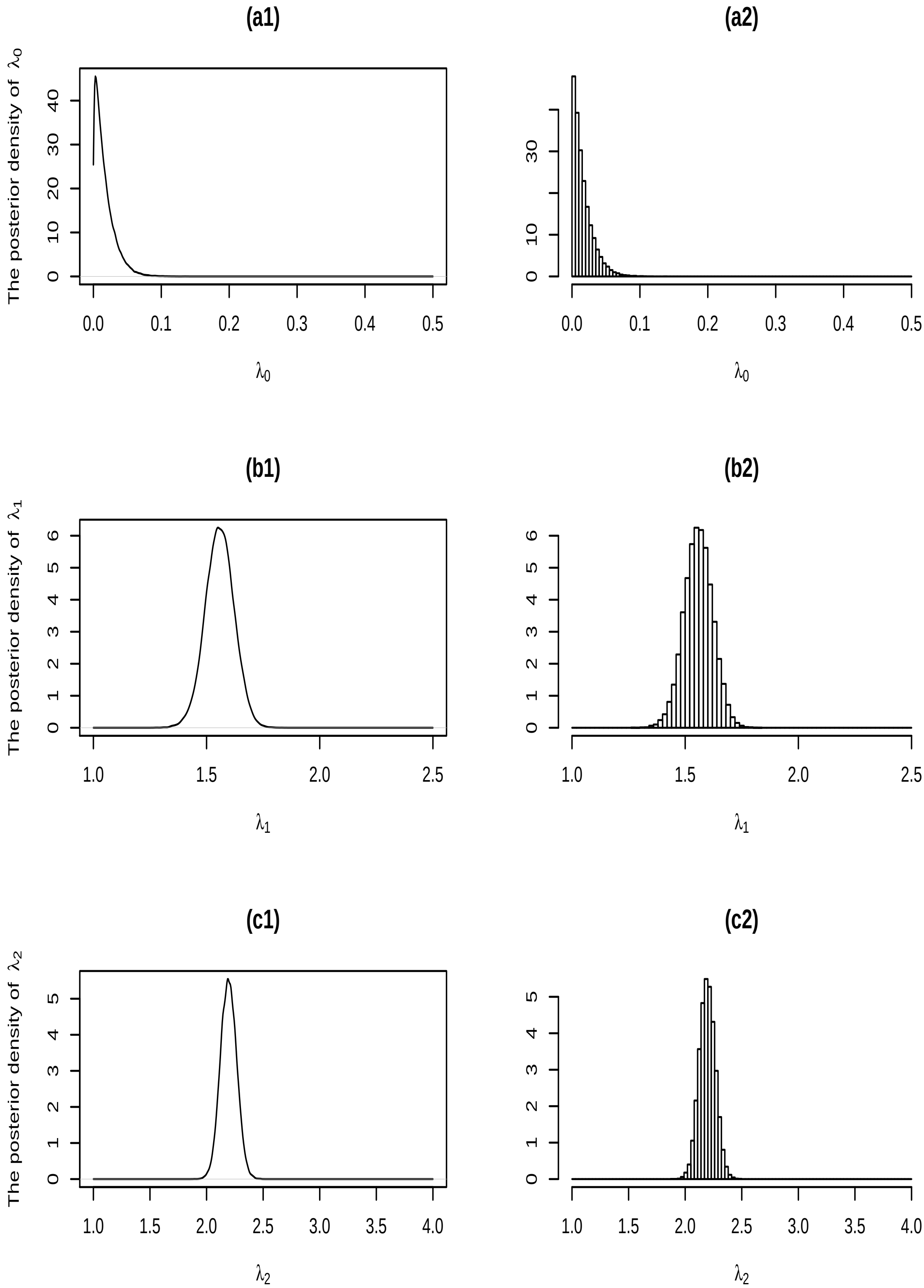

Figure 1 Posterior densities of $(\lambda_0, \lambda_1, \lambda_2)$ via a kernel density smoother based on $L = 60{,}000$ i.i.d. posterior samples generated by the DA algorithm with independent Gamma$(1, 1)$ as the prior distributions of $(\lambda_0, \lambda_1, \lambda_2)$. (a1)–(c1) The posterior densities of $(\lambda_0, \lambda_1, \lambda_2)$; (a2)–(c2) the histograms of $(\lambda_0, \lambda_1, \lambda_2)$.

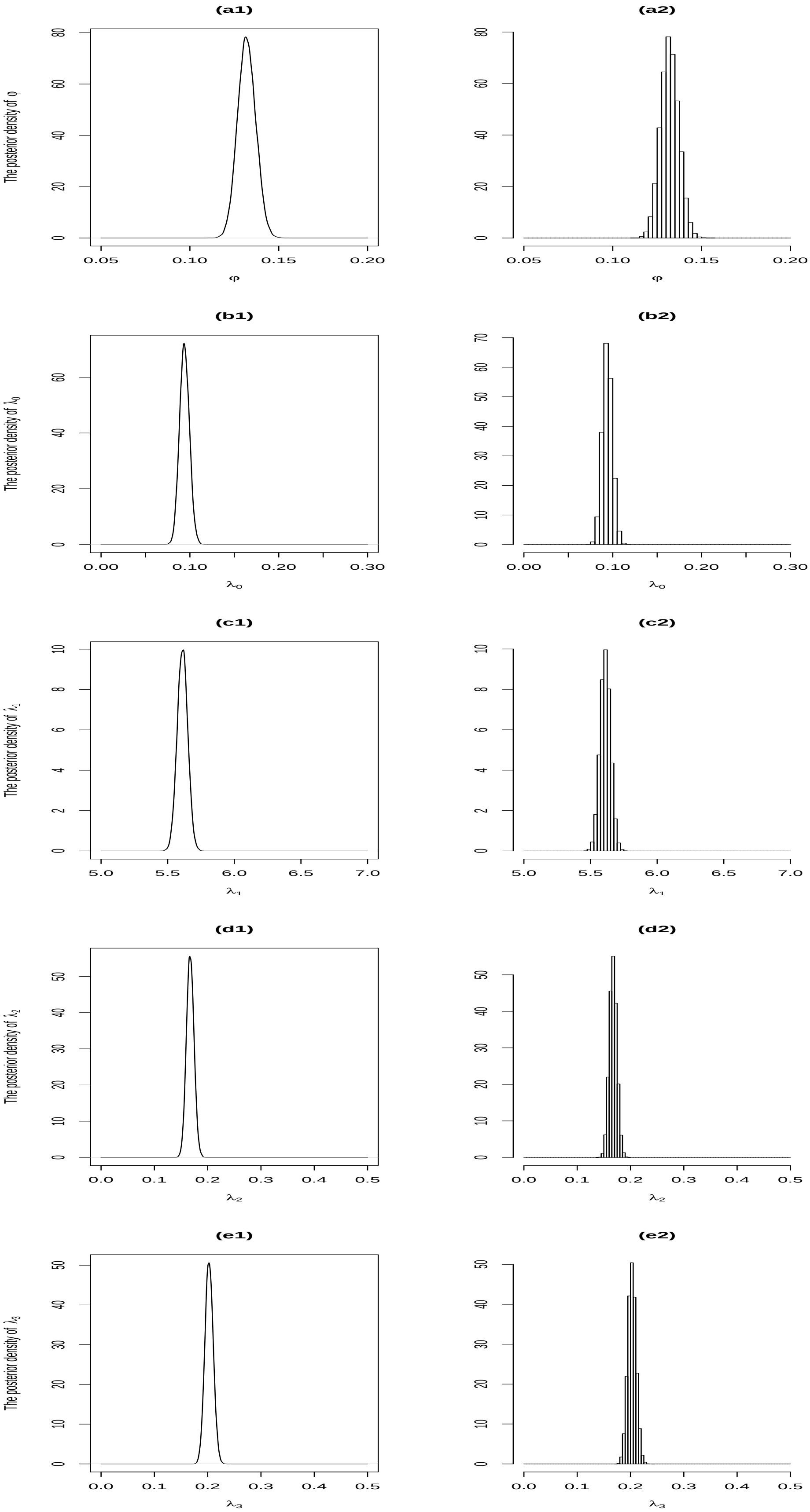

Figure 2 Posterior densities of $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$ via a kernel density smoother based on $L =$ 60,000 i.i.d. posterior samples generated by the DA algorithm with independent Gamma$(1, 1)$ as the prior distributions of $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$. (a1)–(e1) The posterior densities of $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$; (a2)–(e2) the histograms of $(\varphi, \lambda_0, \lambda_1, \lambda_2, \lambda_3)$.