Ding *et al. Journal of Statistical Distributions and Applications* (2015) 2:5
DOI 10.1186/s40488-015-0029-5

Journal of Statistical Distributions and Applications
a SpringerOpen Journal

METHODOLOGY Open Access

Multivariate zero-truncated/adjusted Charlier series distributions with applications

Xiqian Ding, Da Ju and Guo-Liang Tian*

*Correspondence: gltian@hku.hk
Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, P. R. China

Abstract

Although the univariate Charlier series distribution (Biom. J. 30(8):1003–1009, 1988) and bivariate Charlier series distribution (Biom. J. 37(1):105–117, 1995; J. Appl. Stat. 30(1):63–77, 2003) can be easily generalized to the multivariate version via the method of *stochastic representation* (SR), the multivariate *zero-truncated Charlier series* (ZTCS) distribution is not available to date. The first aim of this paper is to propose the multivariate ZTCS distribution by developing its important distributional properties, and providing efficient likelihood-based inference methods via a novel data augmentation in the framework of the *expectation–maximization* (EM) algorithm. Since the joint marginal distribution of any r-dimensional sub-vector of the multivariate ZTCS random vector of dimension m is an r-dimensional *zero-deflated Charlier series* (ZDCS) distribution ($1 \le r < m$), it is the second objective of the paper to introduce a new family of multivariate *zero-adjusted Charlier series* (ZACS) distributions (including the multivariate ZDCS distribution as a special member) with a more flexible correlation structure by accounting for both inflation and deflation at zero. The corresponding distributional properties are explored and the associated maximum likelihood estimation method via EM algorithm is provided for analyzing correlated count data. Some simulation studies are performed and two real data sets are used to illustrate the proposed methods.

Mathematics subject classification primary: 62E15; Secondary 62F10

Keywords: EM algorithm; Multivariate zero-adjusted Charlier series distributions; Multivariate zero-truncated Charlier series distribution; Univariate Charlier series distribution

1 Introduction

The univariate *Charlier series* (CS) distribution was first introduced by Ong (1988) in the consideration of the conditional distribution of a bivariate Poisson distribution. The CS distribution is a convolution of a binomial variate and a Poisson variate. Let $X_0 \sim \text{Binomial}(K, \pi)$, $X_1 \sim \text{Poisson}(\lambda)$, and (X_0, X_1) be mutually independent (denoted by $X_0 \perp\!\!\!\perp X_1$). Then a discrete non-negative random variable X is said to follow the CS distribution with parameters $K \in \mathbb{N} \triangleq \{1, 2, \ldots, \infty\}$, $\pi \in [0, 1)$ and $\lambda \in \mathbb{R}_+$, denoted by $X \sim \text{CS}(K, \pi; \lambda)$, if it can be stochastically represented by $X = X_0 + X_1$. Its *probability mass function* (pmf) is given by

$$\Pr(X = x) = \sum_{k=0}^{\min(K,x)} \binom{K}{k} \pi^k (1-\pi)^{K-k} \cdot \frac{\lambda^{x-k} e^{-\lambda}}{(x-k)!}, \qquad x = 0, 1, \ldots, \infty. \tag{1.1}$$

The mean and variance of X are given by

$$E(X) = K\pi + \lambda \quad \text{and} \quad \text{Var}(X) = K\pi(1-\pi) + \lambda. \tag{1.2}$$

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{CS}(K, \pi; \lambda)$ and the observed data be $Y_{\text{obs}} = \{x_1, \ldots, x_n\}$, where $x_1, \ldots, x_n$ are the realizations of $X_1, \ldots, X_n$. Let $\bar{x}$ and s^2 be the sample mean and variance, respectively. Assuming K is known, Ong (1988) derived the moment estimates of the parameters in the univariate CS distribution as follows:

$$\hat{\pi} = \left(\frac{\bar{x} - s^2}{K}\right)^{1/2} \quad \text{and} \quad \hat{\lambda} = \bar{x} - K\hat{\pi}. \tag{1.3}$$

Next, Papageorgiou and Loukas (1995) proposed a bivariate CS distribution which arises as the conditional distribution from a trivariate Poisson distribution studied by Loukas and Papageorgiou (1991) and Loukas (1993). Let $X_0 \sim \text{Binomial}(K, \pi)$ and $X_{i0} \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_i)$, $i = 1, 2$ and define $X_i = X_0 + X_{i0}$, $i = 1, 2$. Then a discrete non-negative random vector $\mathbf{x} = (X_1, X_2)^\top$ is said to follow a bivariate CS distribution with parameters $K \in \mathbb{N}$, $\pi \in [0, 1)$ and $\lambda_i \in \mathbb{R}_+$, $i = 1, 2$. We denote it by $\mathbf{x} \sim \text{CS}(K, \pi; \lambda_1, \lambda_2)$. Its probability generating function, marginal means and the covariance are given by

$$G_{\mathbf{x}}(\boldsymbol{z}) = E\left(z_1^{X_1} z_2^{X_2}\right) = \exp\{\lambda_1(z_1 - 1) + \lambda_2(z_2 - 1)\}\left[(1-\pi) + \pi z_1 z_2\right]^K, \tag{1.4}$$

$$E(X_i) = K\pi + \lambda_i \quad \text{and} \quad \text{Cov}(X_1, X_2) = K\pi(1-\pi), \quad i = 1, 2. \tag{1.5}$$

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{ind}}{\sim} \text{CS}(K, \pi; \lambda_1, \lambda_2)$, where $\mathbf{x}_j = (X_{1j}, X_{2j})^\top$ for $j = 1, \ldots, n$ and the observed data be $Y_{\text{obs}} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are the realizations of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Let $\bar{x}_1, \bar{x}_2$ be the sample mean for X_1 and X_2 and m_{11} be the sample covariance, respectively. Assuming K is known, Papageorgiou and Loukas (1995) obtained the moment estimates of the three parameters as follows:

$$\hat{\pi} = \frac{1}{2} \pm \left(1 - \frac{4m_{11}}{K}\right)^{1/2}, \quad \hat{\lambda}_1 = \bar{x}_1 - K\hat{\pi} \quad \text{and} \quad \hat{\lambda}_2 = \bar{x}_2 - K\hat{\pi}. \tag{1.6}$$

In addition, Papageorgiou and Loukas (1995) also discussed the method of ratio of frequencies and the maximum likelihood estimate method.

Although the univariate Charlier series distribution (Ong 1988) and bivariate Charlier series distribution (Karlis 2003, Papageorgiou and Loukas 1995) can be easily generalized to the multivariate version via the method of *stochastic representation* (SR), the multivariate *zero-truncated Charlier series* (ZTCS) distribution is not available to date. The first aim of this paper is to propose the multivariate ZTCS distribution by developing its important distributional properties, and providing efficient likelihood-based inference methods via a novel data augmentation in the framework of the *expectation–maximization* (EM) algorithm. Since the joint marginal distribution of any r-dimensional sub-vector of the multivariate ZTCS random vector of dimension m is an r-dimensional *zero-deflated Charlier series* (ZDCS) distribution ($1 \le r < m$), it is the second objective of the paper to introduce a new family of multivariate *zero-adjusted Charlier series* (ZACS) distributions (including the multivariate ZDCS distribution as a special member) with a more flexible correlation structure by accounting for both inflation and deflation at zero. The corresponding distributional properties are explored and the associated maximum likelihood estimation method via EM algorithm is provided for analyzing correlated count data.

The rest of the paper is organized as follows. In Section 2, the multivariate ZTCS distribution is proposed and some important distributional properties are explored. In Section 3, the likelihood-based methods are developed for the multivariate ZTCS distribution. In Sections 4 and 5, we introduce the multivariate ZACS distribution, explore its distributional properties and provide associated likelihood-based methods for the case of without covariates. In Section 6, some simulation studies are performed to evaluate the proposed methods. In Section 7, two real data sets are used to illustrate the proposed methods. Section 8 provides some concluding remarks.

2 Multivariate zero-truncated Charlier series distribution

Let $X_{00} \sim \text{Binomial}(K, \pi)$, $\{X_{i0}\}_{i=1}^m \stackrel{\text{ind}}{\sim} \text{Poisson}\,(\lambda_i)$, $X_{00} \perp\!\!\!\perp \{X_{10}, \ldots, X_{m0}\}$ and define

$$X_i = X_{00} + X_{i0}, \qquad i = 1, \ldots, m.$$

A discrete non-negative random vector $\mathbf{x} = (X_1, \ldots, X_m)^\top$ is said to follow an m-dimensional CS distribution with parameters $K \in \mathbb{N} = \{1, 2, \ldots, \infty\}$, $\pi \in [0, 1)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}_+^m$, denoted by $\mathbf{x} \sim \text{CS}\,(K, \pi; \lambda_1, \ldots, \lambda_m)$ or $\mathbf{x} \sim \text{CS}_m(K, \pi; \boldsymbol{\lambda})$, accordingly. The joint pmf of $\mathbf{x}$ is

$$\Pr(\mathbf{x} = \boldsymbol{x}) = \sum_{k=0}^{\min(K,\boldsymbol{x})} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=1}^m \frac{\lambda_i^{x_i - k} \mathrm{e}^{-\lambda_i}}{(x_i - k)!} \hat{=} Q_{\boldsymbol{x}}(K, \pi, \boldsymbol{\lambda}), \tag{2.1}$$

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

In particular, as $K \to \infty$ and $K\pi$ remains finitely large (say, λ_0), the distribution of $\text{Binomial}(K, \pi)$ tends to the distribution of $\text{Poisson}(\lambda_0)$, so the above m-dimensional CS distribution approaches to the m-dimensional Poisson distribution $\text{MP}(\lambda_0, \lambda_1, \ldots, \lambda_m)$. Furthermore, if $\pi = 0$, then $\Pr(X_{00} = 0) = 1$ (i.e., X_{00} follows the degenerate distribution with all mass at zero, denoted by $X_{00} \sim \text{Degenerate}(0)$) and $\lambda_0 = 0$, so the m-dimensional CS distribution becomes the product of m independent $\text{Poisson}(\lambda_i)$ distributions.

Motivated by the Type II multivariate *zero-truncated Poisson* (ZTP) distribution developed recently by Tian et al. (2014), we in this paper propose a new multivariate *zero-truncated Charlier series* (ZTCS) distribution, whose limiting form reduces to the Type II multivariate ZTP distribution.

Definition 1. *Let* $\mathbf{x} \sim CS\,(K, \pi; \lambda_1, \ldots, \lambda_m)$*. A discrete non-zero random vector* $\mathbf{w} = (W_1, \ldots, W_m)^\top$ *is said to have the multivariate ZTCS distribution with the parameters* (K, π) *and* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^\top$*, denoted by* $\mathbf{w} \sim ZTCS_m(K, \pi; \boldsymbol{\lambda})$ *or* $\mathbf{w} \sim ZTCS(K, \pi; \lambda_1, \ldots, \lambda_m)$*, if*

$$\mathbf{x} \stackrel{\mathrm{d}}{=} U\,\mathbf{w} = \begin{cases} \mathbf{0}, & \textit{with probability } \psi, \\ \mathbf{w}, & \textit{with probability } 1 - \psi, \end{cases} \tag{2.2}$$

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

Let $\mathbf{w} \sim \text{ZTCS}_m(K, \pi; \boldsymbol{\lambda})$, 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 $\mathbf{x}$ is specified in Definition 1. The SR (2.3) can be used to generate the ZTCS random vector $\mathbf{w}$ via the generation of the random vector $\mathbf{x}$ from the multivariate CS distribution, while the SR (2.2) is useful in deriving important distributional properties in the following subsections and in developing an EM algorithm in Section 3.1. Moreover, besides coming from the missing zero vector, the correlation between any two components of $\mathbf{w}$ may come from the common random variable $X_{00} \sim \text{Binomial}(K, \pi)$.

2.1 Joint probability mass function and mixed moments

From the SR (2.2), the joint pmf of $\mathbf{w} \sim \text{ZTCS}_m(K, \pi; \boldsymbol{\lambda})$ is

$$\begin{aligned} f(\boldsymbol{w}; K, \pi, \boldsymbol{\lambda}) &= \Pr(\mathbf{w} = \boldsymbol{w}) \overset{(2.2)}{=} \frac{\Pr(\mathbf{x} = \boldsymbol{w})}{\Pr(U = 1)} \\ &= \frac{1}{1 - (1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{k=0}^{\min(K, \boldsymbol{w})} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=1}^{m} \frac{\lambda_i^{w_i - k} \mathrm{e}^{-\lambda_i}}{(w_i - k)!}, \end{aligned} \tag{2.4}$$

where $\|\boldsymbol{w}\|_1 \neq 0$. From (2.2), it is easy to show that

$$\begin{cases} E(\mathbf{w}) &= \frac{\boldsymbol{\lambda} + K\pi \cdot \mathbf{1}}{1-\psi}, \\ E(\mathbf{w}\mathbf{w}^\top) &= \frac{\text{diag}(\boldsymbol{\lambda}) + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top + K\pi(\boldsymbol{\lambda}\mathbf{1}^\top + \mathbf{1}\boldsymbol{\lambda}^\top) + K\pi(1-\pi+K\pi)\cdot \mathbf{1}\mathbf{1}^\top}{1-\psi}, \\ \text{Var}(\mathbf{w}) &= \frac{1}{1-\psi} \Big\{ \text{diag}(\boldsymbol{\lambda}) + K\pi(1-\pi) \cdot \mathbf{1}\mathbf{1}^\top \\ & \quad - \frac{\psi}{1-\psi} \left[\boldsymbol{\lambda}\boldsymbol{\lambda}^\top + K\pi \left(\boldsymbol{\lambda}\mathbf{1}^\top + \mathbf{1}\boldsymbol{\lambda}^\top \right) + K^2\pi^2 \mathbf{1}\mathbf{1}^\top \right] \Big\}, \end{cases} \tag{2.5}$$

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

$$\text{Corr}(W_i, W_j) = \\ \frac{K\pi(1-\pi) - \frac{\psi}{1-\psi}(\lambda_i + K\pi)(\lambda_j + K\pi)}{\sqrt{\left[\lambda_i + K\pi(1-\pi) - \frac{\psi}{1-\psi}(\lambda_i + K\pi)^2\right]\left[\lambda_j + K\pi(1-\pi) - \frac{\psi}{1-\psi}(\lambda_j + K\pi)^2\right]}}, \tag{2.6}$$

for $i \neq j$. In particular, when $\pi = 0$, (2.6) becomes

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

In (2.6), let $\lambda_i = \lambda_j = \lambda$, we obtain

$$\text{Corr}(W_i, W_j) = \frac{K\pi(1-\pi) - \frac{\psi}{1-\psi}(\lambda + K\pi)^2}{\lambda + K\pi(1-\pi) - \frac{\psi}{1-\psi}(\lambda + K\pi)^2}, \qquad i \neq j.$$

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

$$E\left(\prod_{i=1}^{m} W_i^{r_i}\right) = (1-\psi)^{-1} E\left(\prod_{i=1}^{m} X_i^{r_i}\right) = (1-\psi)^{-1} E\left[\prod_{i=1}^{m} (X_{00} + X_{i0})^{r_i}\right]. \tag{2.7}$$

2.2 Moment generating function

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

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

Thus the mgf of $\mathbf{w} \sim \mathrm{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_m)$ is given by

$$\begin{aligned} M_{\mathbf{w}}(\boldsymbol{t}) &= \frac{M_{\mathbf{x}}(\boldsymbol{t}) - \psi}{1-\psi} \\ &= \frac{M_{X_{00}}(t_+)\prod_{i=1}^m M_{X_{i0}}(t_i) - \psi}{1-\psi} \\ &= \frac{\left(\pi \mathrm{e}^{t_+} + 1 - \pi\right)^K \exp\left(\sum_{i=1}^m \lambda_i \mathrm{e}^{t_i} - \lambda_+\right) - (1-\pi)^K \mathrm{e}^{-\lambda_+}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}}, \end{aligned}$$

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

2.3 Marginal distributions

2.3.1 Marginal distribution for each random component

Let $\mathbf{w} = (W_1, \ldots, W_m)^\top \sim \mathrm{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_m)$. We first derive the marginal distribution of W_i with realization w_i for $i = 1, \ldots, m$. If $w_i > 0$, then

$$\begin{aligned} \Pr(W_i = w_i) &= \sum_{w_1=0}^{\infty} \cdots \sum_{w_{i-1}=0}^{\infty} \sum_{w_{i+1}=0}^{\infty} \cdots \sum_{w_m=0}^{\infty} \Pr(\mathbf{w} = \boldsymbol{w}) \\ &= \frac{1}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{w_1=0}^{\infty} \cdots \sum_{w_{i-1}=0}^{\infty} \sum_{w_{i+1}=0}^{\infty} \cdots \sum_{w_m=0}^{\infty} \\ &\quad \times \sum_{k=0}^{\min\{K,w_i\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{j=1}^m \frac{\lambda_j^{w_j-k} \mathrm{e}^{-\lambda_j}}{(w_j-k)!} \cdot I(w_j - k > 0) \\ &= \frac{1}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{k=0}^{\min\{K,w_i\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \frac{\lambda_i^{w_i-k} \mathrm{e}^{-\lambda_i}}{(w_i-k)!} \\ &\quad \times \left[\sum_{w_1=0}^{\infty} \cdots \sum_{w_{i-1}=0}^{\infty} \sum_{w_{i+1}=0}^{\infty} \cdots \sum_{w_m=0}^{\infty} \prod_{j=1, j\neq i}^m \frac{\lambda_j^{w_j-k} \mathrm{e}^{-\lambda_j}}{(w_j-k)!} \cdot I(w_j - k > 0)\right] \\ &= \frac{1}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{k=0}^{\min\{K,w_i\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \frac{\lambda_i^{w_i-k} \mathrm{e}^{-\lambda_i}}{(w_i-k)!} \qquad (2.8) \\ &= \frac{1-\varphi_i}{1-(1-\pi)^K \mathrm{e}^{-\lambda_i}} \sum_{k=0}^{\min\{K,w_i\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \frac{\lambda_i^{w_i-k} \mathrm{e}^{-\lambda_i}}{(w_i-k)!}, \qquad (2.9) \end{aligned}$$

where

$$1 - \varphi_i = \frac{1-(1-\pi)^K \mathrm{e}^{-\lambda_i}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}}. \qquad (2.10)$$

Hence,

$$\begin{aligned}
\Pr(W_i = 0) &= 1 - \sum_{w_i=1}^{\infty} \Pr(W_i = w_i) \\
&\overset{(2.8)}{=} 1 - \frac{1}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{w_i=1}^{\infty} \sum_{k=0}^{\min\{K,w_i\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \frac{\lambda_i^{w_i-k} \mathrm{e}^{-\lambda_i}}{(w_i-k)!} \\
&= 1 - \frac{1-(1-\pi)^K \mathrm{e}^{-\lambda_i}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \overset{(2.10)}{=} \varphi_i \qquad (2.11) \\
&= \frac{(1-\pi)^K (\mathrm{e}^{-\lambda_i} - \mathrm{e}^{-\lambda_+})}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \in \left(0, (1-\pi)^K \mathrm{e}^{-\lambda_i}\right) \subset (0,1).
\end{aligned}$$

By combining (2.11) with (2.9) and noting that a ZDCS distribution is a special case of a ZACS distribution (4.2), we obtain

$$W_i \sim \mathrm{ZDCS}(\varphi_i; K, \pi, \lambda_i). \qquad (2.12)$$

2.3.2 Marginal distribution for an arbitrary random sub-vector

Second, the marginal distribution for an arbitrary random sub-vector will be considered. Before that, a so-called *multivariate zero-adjusted Charlier series* distribution is needed to be introduced. We will give the definition of this distribution in Definition 2 in Section 4. We now consider the marginal distributions of $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$, where

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

Furthermore in Section 4, we will introduce *multivariate zero-adjusted Charlier series* distribution and it can be shown that

$$\mathbf{w}^{(1)} \sim \mathrm{ZDCS}(\varphi^{(1)}; K, \pi, \lambda_1, \ldots, \lambda_r) \quad \text{and} \quad \mathbf{w}^{(2)} \sim \mathrm{ZDCS}(\varphi^{(2)}; K, \pi, \lambda_{r+1}, \ldots, \lambda_m), \qquad (2.13)$$

where

$$\varphi^{(i)} = \frac{(1-\pi)^K \left(\mathrm{e}^{-\lambda_+^{(i)}} - \mathrm{e}^{-\lambda_+}\right)}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \in \left(0,\ (1-\pi)^K \mathrm{e}^{-\lambda_+^{(i)}}\right) \subset (0,1), \quad i = 1,2, \qquad (2.14)$$

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

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

$$\begin{pmatrix} W_{i_1} \\ \vdots \\ W_{i_r} \end{pmatrix} \sim \mathrm{ZDCS}(\varphi^*; K, \pi, \lambda_{i_1}, \ldots, \lambda_{i_r}), \qquad (2.15)$$

where

$$\varphi^* = \frac{(1-\pi)^K \left[\mathrm{e}^{-(\lambda_{i_1}+\cdots+\lambda_{i_r})} - \mathrm{e}^{-\lambda_+}\right]}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \in \left(0,\ (1-\pi)^K \mathrm{e}^{-(\lambda_{i_1}+\cdots+\lambda_{i_r})}\right) \subset (0,1). \qquad (2.16)$$

2.4 Conditional distributions

2.4.1 Conditional distribution of $\mathbf{w}^{(1)}|\mathbf{w}^{(2)}$

From (2.4), (2.13) and (4.4), the conditional distribution of $\mathbf{w}^{(1)}|\mathbf{w}^{(2)}$ is given by

$$\Pr(\mathbf{w}^{(1)} = \boldsymbol{w}^{(1)}|\mathbf{w}^{(2)} = \boldsymbol{w}^{(2)}) = \frac{f(\boldsymbol{w}; K, \pi, \boldsymbol{\lambda})}{\Pr(\mathbf{w}^{(2)} = \boldsymbol{w}^{(2)})} \tag{2.17}$$

$$= \frac{\frac{1}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \cdot Q_{\boldsymbol{w}}(K, \pi, \boldsymbol{\lambda})}{\varphi^{(2)} I(\boldsymbol{w}^{(2)} = \mathbf{0}) + \left[\frac{1-\varphi^{(2)}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+^{(2)}}} \cdot Q_{\boldsymbol{w}^{(2)}}(K, \pi, \boldsymbol{\lambda}^{(2)})\right] I(\boldsymbol{w}^{(2)} \neq \mathbf{0})},$$

where $\boldsymbol{w}^{(2)} = (w_{r+1}, \ldots, w_m)^\top$, $\lambda^{(2)} = (\lambda_{r+1}, \ldots, \lambda_m)^\top$ and

$$Q_{\boldsymbol{w}}(K, \pi, \boldsymbol{\lambda}) = \sum_{k=0}^{\min\{K, \boldsymbol{w}\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{j=1}^{m} \frac{\lambda_j^{w_j - k} \mathrm{e}^{-\lambda_j}}{(w_j - k)!},$$

$$Q_{\boldsymbol{w}^{(2)}}\left(K, \pi, \boldsymbol{\lambda}^{(2)}\right) = \sum_{l=0}^{\min\{K, \boldsymbol{w}^{(2)}\}} \binom{K}{l} \pi^l (1-\pi)^{K-l} \prod_{p=r+1}^{m} \frac{\lambda_p^{w_p - l} \mathrm{e}^{-\lambda_p}}{(w_p - p)!}.$$

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 (2.17), it is easy to obtain

$$\Pr\left(\mathbf{w}^{(1)} = \boldsymbol{w}^{(1)}|\mathbf{w}^{(2)} = \boldsymbol{w}^{(2)}\right) = \frac{\mathrm{e}^{-\lambda_+^{(1)}} \sum\limits_{k=0}^{\min\{K, \boldsymbol{w}\}} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod\limits_{j=1}^{m} \frac{\lambda_j^{w_j - k}}{(w_j - k)!}}{\sum\limits_{l=0}^{\min\{K, \boldsymbol{w}^{(2)}\}} \binom{K}{l} \pi^l (1-\pi)^{K-l} \prod\limits_{p=r+1}^{m} \frac{\lambda_p^{w_p - l}}{(w_p - l)!}}. \tag{2.18}$$

Case Ⅱ: $\boldsymbol{w}^{(2)} = \mathbf{0}$. Under Case Ⅱ, it is obviously that $\boldsymbol{w}^{(1)} \neq \mathbf{0}$ and the sharing binomial variable equals to zero. Thus we have

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

This implies

$$\mathbf{w}^{(1)}|\left(\mathbf{w}^{(2)} = \mathbf{0}\right) \sim \mathrm{ZTP}^{(\mathrm{I})}(\lambda_1, \ldots, \lambda_r). \tag{2.19}$$

2.4.2 Conditional distribution of $X_0^*|(\boldsymbol{w}, U)$

The stochastic representation (2.2) can be rewritten as

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

where $X_0^* \sim \mathrm{Binomial}(K, \pi)$ and $\{X_i^*\}_{i=1}^m \stackrel{\mathrm{ind}}{\sim} \mathrm{Poisson}(\lambda_i)$. To obtain the conditional distribution of $X_0^*|(\mathbf{w}, U)$, we consider two cases: $U = 1$ and $U = 0$. When $U = 1$, the conditional distribution of $X_0^*|(\mathbf{w}, U)$ is given by

$$\begin{aligned}
\Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 1) &= \frac{\Pr(X_0^* = l, X_1^* = w_1 - l, \ldots, X_m^* = w_m - l)}{\Pr(X_1 = w_1, \ldots, X_m = w_m)} \\
&= \frac{\binom{K}{l} \pi^l (1-\pi)^{K-l} \prod\limits_{i=1}^{m} \frac{\lambda_i^{w_i - l}}{(w_i - l)!}}{\sum\limits_{k=0}^{\min(K, \boldsymbol{w})} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod\limits_{i=1}^{m} \frac{\lambda_i^{w_i - k}}{(w_i - k)!}} \\
&\mathrel{\hat{=}} q_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda}),
\end{aligned} \tag{2.20}$$

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

$$X_0^*|(\mathbf{w} = \boldsymbol{w}, U = 1) \sim \text{Finite}(l, q_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(\boldsymbol{w})). \tag{2.21}$$

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

$$X_0^*|(\mathbf{w} = \boldsymbol{w}, U = 0) \sim \text{Degenerate}(0). \tag{2.22}$$

Hence, for any l, we have

$$\Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 0) = I(l = 0). \tag{2.23}$$

Thus, we have the conditional distribution of $X_0^*|(\mathbf{w}, U)$, which is given by the following:

$$X_0^*|(\mathbf{w}, U) \sim \begin{cases} \text{Finite}(l, q_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(\boldsymbol{w})), & \text{if } U = 1, \\ \text{Degenerate}(0), & \text{if } U = 0, \end{cases} \tag{2.24}$$

where $q_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda})$ is defined by (2.20).

2.4.3 Conditional distribution of $X_0^*|\mathbf{w}$

By using (2.24), the conditional distribution of $X_0^*|\mathbf{w}$ is

$$\begin{aligned} \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}) &= \sum_{u=0}^{1} \Pr(X_0^* = l, U = u|\mathbf{w} = \boldsymbol{w}) \\ &= \sum_{u=0}^{1} \Pr(U = u|\mathbf{w} = \boldsymbol{w}) \cdot \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = u) \\ &= \Pr(U = 0) \cdot \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 0) \\ &\quad + \Pr(U = 1) \cdot \Pr(X_0^* = l|\mathbf{w} = \boldsymbol{w}, U = 1) \\ &\overset{(2.24)}{=} \mathrm{e}^{-\lambda_+}(1-\pi)^K \cdot I(l = 0) + [1 - \mathrm{e}^{-\lambda_+}(1-\pi)^K] \cdot q_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda}) \\ &\hat{=} p_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda}), \end{aligned} \tag{2.25}$$

for $l = 0, 1, \ldots, \min(K, \boldsymbol{w})$, where $q_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda})$ is defined by (2.20). Thus,

$$X_0^*|(\mathbf{w} = \boldsymbol{w}) \sim \text{Finite}(l, p_l(\boldsymbol{w}, K, \pi, \boldsymbol{\lambda});\ l = 0, 1, \ldots, \min(K, \boldsymbol{w})). \tag{2.26}$$

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

$$\begin{aligned} E(X_0^*|\mathbf{w} = \boldsymbol{w}) &= \left[1 - \mathrm{e}^{-\lambda_+}(1-\pi)^K\right] \\ &\quad \times \frac{\sum_{k=1}^{\min(K,\boldsymbol{w})} k\binom{K}{k}\pi^k(1-\pi)^{K-k} \prod_{i=1}^{m} \frac{\lambda_i^{w_i-k}}{(w_i-k)!}}{\sum_{k=0}^{\min(K,\boldsymbol{w})} \binom{K}{k}\pi^k(1-\pi)^{K-k} \prod_{i=1}^{m} \frac{\lambda_i^{w_i-k}}{(w_i-k)!}} \cdot I(\min(\boldsymbol{w}) \geq 1). \end{aligned} \tag{2.27}$$

3 Likelihood-based methods for the multivariate ZTCS distribution

Suppose that $\mathbf{w}_j \overset{\text{ind}}{\sim} \text{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_m)$, where $\mathbf{w}_j = (W_{1j}, \ldots, W_{mj})^\top$ for $j = 1, \ldots, n$. Let $\boldsymbol{w}_j = (w_{1j}, \ldots, w_{mj})^\top$ denote the realization of the random vector $\mathbf{w}_j$, and $Y_{\text{obs}} =$

$\{\boldsymbol{w}_j\}_{j=1}^n$ be the observed data. We consider K as a known positive integer. Then, the observed-data likelihood function for $(\pi, \boldsymbol{\lambda})$ is

$$L(\pi, \boldsymbol{\lambda}|Y_{\rm obs}) = \prod_{j=1}^n \frac{\mathrm{e}^{-\lambda_+}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{k=0}^{\min(K, \boldsymbol{w}_j)} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=1}^m \frac{\lambda_i^{w_{ij}-k}}{(w_{ij}-k)!},$$

so that the log-likelihood function is

$$\begin{aligned}\ell(\pi, \boldsymbol{\lambda}|Y_{\rm obs}) &= -n \log\left[\mathrm{e}^{\lambda_+} - (1-\pi)^K\right] \\ &\quad + \sum_{j=1}^n \log\left[\sum_{k=0}^{\min(K, \boldsymbol{w}_j)} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=1}^m \frac{\lambda_i^{w_{ij}-k}}{(w_{ij}-k)!}\right]. \qquad (3.1)\end{aligned}$$

3.1 MLEs via the EM algorithm

The SR (2.2) can motivate a novel EM algorithm, where some latent variables are independent of the observed variables. For each $\boldsymbol{w}_j = (w_{1j}, \ldots, w_{mj})^\top$, we introduce latent variables $U_j \overset{\text{iid}}{\sim} \text{Bernoulli}(1-\psi)$ with $\psi = (1-\pi)^K \mathrm{e}^{-\lambda_+}$, $X_{0j}^* \overset{\text{iid}}{\sim} \text{Binomial}(K, \pi)$, $X_{ij}^* \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_i)$ for $i = 1, \ldots, m$, and $X_{0j}^* \perp\!\!\!\perp X_{ij}^*$, such that

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

where u_j and x_{ij}^* denote the realizations of U_j and X_{ij}^*, respectively. We denote the latent/missing data by $Y_{\rm mis} = \left\{u_j, x_{0j}^*, x_{1j}^*, \ldots, x_{mj}^*\right\}_{j=1}^n$, so that the complete data are

$$\begin{aligned}Y_{\rm com} &= Y_{\rm obs} \cup Y_{\rm mis} = \left\{\boldsymbol{w}_j, u_j, x_{0j}^*, x_{1j}^*, \ldots, x_{mj}^*\right\}_{j=1}^n \\ &= \left\{x_{0j}^*, x_{1j}^*, \ldots, x_{mj}^*\right\}_{j=1}^n = \left\{x_{0j}^*, u_j, \boldsymbol{w}_j\right\}_{j=1}^n,\end{aligned}$$

where $x_{ij}^* = u_j w_{ij} - x_{0j}^*$ for $j = 1, \ldots, n$ and $i = 1, \ldots, m$. Thus, the complete-data likelihood function is given by

$$\begin{aligned}L(\pi, \boldsymbol{\lambda}|Y_{\rm com}) &= \prod_{j=1}^n \left[\binom{K}{x_{0j}^*} \pi^{x_{0j}^*} (1-\pi)^{K-x_{0j}^*} \prod_{i=1}^m \frac{\lambda_i^{x_{ij}^*} \mathrm{e}^{-\lambda_i}}{x_{ij}^*!}\right] \\ &= \prod_{j=1}^n \left[\binom{K}{x_{0j}^*} \pi^{x_{0j}^*} (1-\pi)^{K-x_{0j}^*} \prod_{i=1}^m \frac{\lambda_i^{u_j w_{ij} - x_{0j}^*} \mathrm{e}^{-\lambda_i}}{\left(u_j w_{ij} - x_{0j}^*\right)!}\right] \\ &\propto \pi^{n\bar{x}_0^*} (1-\pi)^{nK - n\bar{x}_0^*} \prod_{i=1}^m \lambda_i^{\sum_{j=1}^n u_j w_{ij} - n\bar{x}_0^*} \mathrm{e}^{-n\lambda_i}, \qquad (3.2)\end{aligned}$$

where $\bar{x}_0^* = (1/n) \sum_{j=1}^n x_{0j}^*$. The complete-data log-likelihood function is

$$\ell(\pi, \boldsymbol{\lambda}|Y_{\rm com}) = n\bar{x}_0^* \log \pi + \left(nK - n\bar{x}_0^*\right) \log(1-\pi) + \sum_{i=1}^m \left[\left(\sum_{j=1}^n u_j w_{ij} - n\bar{x}_0^*\right) \log \lambda_i - n\lambda_i\right].$$

The M-step is to calculate the complete-data *maximum likelihood estimates* (MLEs):

$$\hat{\pi} = \frac{\bar{x}_0^*}{K} \quad \text{and} \quad \hat{\lambda}_i = \frac{\sum_{j=1}^n u_j w_{ij}}{n} - K\hat{\pi}, \qquad i = 1, \ldots, m, \qquad (3.3)$$

and the E-step is to replace $\{u_j\}_{j=1}^n$ and $\left\{x_{0j}^*\right\}_{j=1}^n$ in (3.3) by their conditional expectations:

$$E(U_j|Y_{\rm obs},\pi,\boldsymbol{\lambda}) \;=\; E(U_j)=1-(1-\pi)^K\mathrm{e}^{-\lambda_+}, \quad \text{and} \tag{3.4}$$

$$E(X_{0j}^*|Y_{\rm obs},\pi,\boldsymbol{\lambda}) \overset{(2.27)}{=} \frac{[1-(1-\pi)^K\mathrm{e}^{-\lambda_+}]\sum\limits_{k_j=1}^{\min(K,\boldsymbol{w}_j)} k_j\binom{K}{k_j}\pi^{k_j}(1-\pi)^{K-k_j}\prod\limits_{i=1}^{m}\frac{\lambda_i^{w_{ij}-k_j}}{(w_{ij}-k_j)!}}{\sum\limits_{k_j=0}^{\min(K,\boldsymbol{w}_j)}\binom{K}{k_j}\pi^{k_j}(1-\pi)^{K-k_j}\prod\limits_{i=1}^{m}\frac{\lambda_i^{w_{ij}-k_j}}{(w_{ij}-k_j)!}} \times I(\min(\boldsymbol{w}_j)\ge 1), \tag{3.5}$$

respectively. An important feature of this EM algorithm is that the latent variables $\{U_j\}_{j=1}^n$ are independent of the observed variables $\{\mathbf{w}_j\}_{j=1}^n$.

Also note that here we assume that K is a known positive integer. In practice, since $K\in\{1,2,\ldots,N\}$, say $N=100$. For a given K, we first use the EM iteration (3.3)–(3.5) to find the MLEs of π and $\boldsymbol{\lambda}$, denoted by $\hat\pi$ and $\hat{\boldsymbol{\lambda}}$. Then, we can calculate $\ell(\hat\pi,\hat{\boldsymbol{\lambda}}|Y_{\rm obs})$ and choose the K that maximizes $\ell(\hat\pi,\hat{\boldsymbol{\lambda}}|Y_{\rm obs})$.

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 $(\pi,\boldsymbol{\lambda})$, say, $\vartheta=h(\pi,\boldsymbol{\lambda})$. Let $(\hat\pi,\hat{\boldsymbol{\lambda}})$ be the MLEs of $(\pi,\boldsymbol{\lambda})$ calculated by the EM algorithm (3.3)–(3.5), then $\hat\vartheta=h(\hat\pi,\hat{\boldsymbol{\lambda}})$ is the MLE of ϑ. Based on $(\hat\pi,\hat{\boldsymbol{\lambda}})$, we can generate $\mathbf{w}_j^*\overset{\rm iid}{\sim}\mathrm{ZTCS}(K,\hat\pi,\hat\lambda_1,\ldots,\hat\lambda_m)$ via the SR (2.2) for $j=1,\ldots,n$. Having obtained $Y_{\rm obs}^*=\left\{\boldsymbol{w}_1^*,\ldots,\boldsymbol{w}_n^*\right\}$, we can calculate the bootstrap replication $(\hat\pi^*,\hat{\boldsymbol{\lambda}}^*)$ and get $\hat\vartheta^*=h(\hat\pi^*,\hat{\boldsymbol{\lambda}}^*)$. Independently repeating this process G times, we obtain G bootstrap replications $\left\{\hat\vartheta_g^*\right\}_{g=1}^G$. Consequently, the standard error, $\mathrm{se}(\hat\vartheta)$, of $\hat\vartheta$ can be estimated by the sample standard deviation of the G replications, i.e.,

$$\widehat{\mathrm{se}}(\hat\vartheta)=\left\{\frac{1}{G-1}\sum_{g=1}^G\left[\hat\vartheta_g^*-(\hat\vartheta_1^*+\cdots+\hat\vartheta_G^*)/G\right]^2\right\}^{1/2}. \tag{3.6}$$

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

$$\left[\hat\vartheta-z_{\alpha/2}\cdot\widehat{\mathrm{se}}(\hat\vartheta),\ \hat\vartheta+z_{\alpha/2}\cdot\widehat{\mathrm{se}}(\hat\vartheta)\right]. \tag{3.7}$$

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

$$[\,\hat\vartheta_{\rm L},\ \hat\vartheta_{\rm U}\,], \tag{3.8}$$

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

4 Multivariate zero-adjusted Charlier series distribution

To introduce the multivariate *zero-adjusted Charlier series* (ZACS) distribution, we first define the univariate ZACS distribution. A non-negative discrete random variable Y is said to have a ZACS distribution with parameters $\varphi\in[0,1)$ and $\lambda>0$, denoted by $Y\sim\mathrm{ZACS}(\varphi,K,\pi,\lambda)$, if

$$Y\overset{\rm d}{=}Z'W, \tag{4.1}$$

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

$$\Pr(Y = y) = \varphi I(y = 0) + \left[\frac{1-\varphi}{1-(1-\pi)^K \mathrm{e}^{-\lambda}} \cdot Q_y(K, \pi, \lambda)\right] I(y \neq 0), \tag{4.2}$$

where

$$Q_y(K, \pi, \lambda) = \sum_{k=0}^{\min(K,y)} \binom{K}{k} \pi^k (1-\pi)^{K-k} \frac{\lambda^{y-k} \mathrm{e}^{-\lambda}}{(y-k)!}.$$

Motivated by (4.1), naturally, we have the following multivariate generalization.

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

$$\mathbf{y} \stackrel{\mathrm{d}}{=} Z'\mathbf{w} = \begin{cases} \mathbf{0}, & \text{with probability } \varphi, \\ \mathbf{w}, & \text{with probability } 1-\varphi, \end{cases} \tag{4.3}$$

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

It is easy to show that the joint pmf of $\mathbf{y} \sim \text{ZACS}(\varphi; K, \pi, \lambda_1, \ldots, \lambda_m)$ is

$$\Pr(\mathbf{y} = \boldsymbol{y}) = \varphi I(\boldsymbol{y} = \mathbf{0}) + \left[\frac{1-\varphi}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \cdot Q_{\boldsymbol{y}}(K, \pi, \boldsymbol{\lambda})\right] I(\boldsymbol{y} \neq \mathbf{0}), \tag{4.4}$$

where

$$Q_{\boldsymbol{y}}(K, \pi, \boldsymbol{\lambda}) = \sum_{k=0}^{\min(K,\boldsymbol{y})} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=1}^{m} \frac{\lambda_i^{y_i-k} \mathrm{e}^{-\lambda_i}}{(y_i-k)!}.$$

We consider several special cases of (4.3) or (4.4):

(i) If $\varphi = 0$, then $\mathbf{y} \stackrel{\mathrm{d}}{=} \mathbf{w} \sim \text{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_m)$, i.e., the multivariate ZTCS distribution is a special member of the family of the multivariate ZACS distributions. Thus, we can see that studying the multivariate ZTCS distribution is a basis for studying the multivariate ZACS distribution;

(ii) If $\varphi \in (0,\ (1-\pi)^K \mathrm{e}^{-\lambda_+})$, then $\mathbf{y}$ follows the multivariate *zero-deflated Charlier series* (ZDCS) distribution with parameters $(\varphi, K, \pi, \boldsymbol{\lambda})$, denoted by $\mathbf{y} \sim \text{ZDCS}_m(\varphi; K, \pi, \boldsymbol{\lambda})$ or $\mathbf{y} \sim \text{ZDCS}(\varphi; K, \pi, \lambda_1, \ldots, \lambda_m)$;

(iii) If $\varphi = (1-\pi)^K \mathrm{e}^{-\lambda_+}$, then $\mathbf{y} \sim \text{CS}_m(K, \pi; \boldsymbol{\lambda})$;

(iv) If $\varphi \in ((1-\pi)^K \mathrm{e}^{-\lambda_+},\ 1)$, then $\mathbf{y}$ follows the multivariate *zero-inflated Charlier series* (ZICS) distribution with parameters $(\varphi, K, \pi, \boldsymbol{\lambda})$, denoted by $\mathbf{y} \sim \text{ZICS}_m(\varphi; K, \pi, \boldsymbol{\lambda})$ or $\mathbf{y} \sim \text{ZICS}(\varphi; K, \pi, \lambda_1, \ldots, \lambda_m)$.

4.1 Mixed moments and moment generating function

From (4.1) and (2.2), we immediately have

$$\begin{cases} E(\mathbf{y}) &= \frac{1-\varphi}{1-\psi}(\boldsymbol{\lambda} + K\pi\cdot\mathbf{1}), \\ E(\mathbf{y}\mathbf{y}^\top) &= \frac{1-\varphi}{1-\psi}\left[\mathrm{diag}(\boldsymbol{\lambda}) + \boldsymbol{\lambda}\boldsymbol{\lambda}^\top + K\pi(\boldsymbol{\lambda}\mathbf{1}^\top + \mathbf{1}\boldsymbol{\lambda}^\top) + K\pi(1-\pi+K\pi)\cdot\mathbf{1}\mathbf{1}^\top\right], \\ \mathrm{Var}(\mathbf{y}) &= \frac{1-\varphi}{1-\psi}\Big\{\mathrm{diag}(\boldsymbol{\lambda}) + K\pi(1-\pi)\cdot\mathbf{1}\mathbf{1}^\top \\ & \quad - \frac{\psi-\varphi}{1-\psi}\left[\boldsymbol{\lambda}\boldsymbol{\lambda}^\top + K\pi(\lambda\mathbf{1}^\top + \mathbf{1}\boldsymbol{\lambda}^\top) + K^2\pi^2\mathbf{1}\mathbf{1}^\top\right]\Big\}. \end{cases} \tag{4.5}$$

Thus, we have

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

for $i \neq j$. In particular, if $\pi = 0$, 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

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

Clearly, $\mathrm{Corr}(Y_i, Y_j)$ could be either positive or negative, which depend on the values of φ, K, π and $\boldsymbol{\lambda}$.

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

$$E\left(\prod_{i=1}^m Y_i^{r_i}\right) = (1-\varphi)E\left(\prod_{i=1}^m W_i^{r_i}\right) = \frac{1-\varphi}{1-\psi}E\left(\prod_{i=1}^m X_i^{r_i}\right). \tag{4.6}$$

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

$$\begin{aligned} M_{\mathbf{y}}(\boldsymbol{t}) &= E[\exp(\boldsymbol{t}^\top\mathbf{y})] = E[\exp(Z'\cdot\boldsymbol{t}^\top\mathbf{w})] = E\Big\{E[\exp(Z'\boldsymbol{t}^\top\mathbf{w})|Z']\Big\} \\ &= E[M_{\mathbf{w}}(Z'\boldsymbol{t})] = \varphi M_{\mathbf{w}}(\mathbf{0}) + (1-\varphi)M_{\mathbf{w}}(\boldsymbol{t}) = \varphi + (1-\varphi)M_{\mathbf{w}}(\boldsymbol{t}) \\ &= \varphi + \frac{1-\varphi}{1-\psi}\left[(\pi e^{t_+} + 1 - \pi)^K \exp\left(\sum_{i=1}^m \lambda_i e^{t_i} - \lambda_+\right) - (1-\pi)^K e^{-\lambda_+}\right], \end{aligned} \tag{4.7}$$

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

4.2 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 (4.1) and (2.13), 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 (2.14), $\boldsymbol{\xi}^{(1)} \sim \text{ZTCS}(K,\pi;\lambda_1,\ldots,\lambda_r)$ and $\boldsymbol{\xi}^{(2)} \sim \text{ZTCS}(K,\pi;\lambda_{r+1},\ldots,\lambda_m)$. 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 (4.3), we can obtain

$$\mathbf{y}^{(1)} \sim \text{ZACS}(\nu^{(1)};K,\pi,\lambda_1,\ldots,\lambda_r) \quad \text{and} \quad \mathbf{y}^{(2)} \sim \text{ZACS}(\nu^{(2)};K,\pi,\lambda_{r+1},\ldots,\lambda_m), \tag{4.8}$$

where

$$\nu^{(k)} = 1-(1-\varphi)(1-\varphi^{(k)}) = 1-(1-\varphi)\frac{1-(1-\pi)^K \mathrm{e}^{-\lambda_+^{(k)}}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \in (0,1), \quad k=1,2, \tag{4.9}$$

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

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

$$\begin{pmatrix} Y_{i_1} \\ \vdots \\ Y_{i_r} \end{pmatrix} \sim \text{ZACS}(\nu^*;K,\pi,\lambda_{i_1},\ldots,\lambda_{i_r}), \tag{4.10}$$

where φ^* is given by (2.16) and

$$\nu^* = 1-(1-\varphi)(1-\varphi^*) = 1-(1-\varphi)\frac{1-(1-\pi)^K \mathrm{e}^{-(\lambda_{i_1}+\cdots+\lambda_{i_r})}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \in (0,1). \tag{4.11}$$

4.3 Conditional distributions

4.3.1 Conditional distribution of $\boldsymbol{y}^{(1)}|\boldsymbol{y}^{(2)}$

From (4.4) and (4.8), the conditional distribution of $\mathbf{y}^{(1)}|\mathbf{y}^{(2)}$ is given by

$$\begin{aligned} \Pr(\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}) &= \frac{\Pr(\mathbf{y}=\boldsymbol{y})}{\Pr(\mathbf{y}^{(2)}=\boldsymbol{y}^{(2)})} \\ &= \frac{\varphi I(\boldsymbol{y}=\mathbf{0}) + R(\boldsymbol{y},K,\pi,\boldsymbol{\lambda},\varphi)I(\boldsymbol{y}\neq\mathbf{0})}{\nu^{(2)}I(\boldsymbol{y}^{(2)}=\mathbf{0}) + S(\boldsymbol{y}^{(2)},K,\pi,\lambda,\nu^{(2)})I(\boldsymbol{y}^{(2)}\neq\mathbf{0})}. \end{aligned} \tag{4.12}$$

where

$$R(\boldsymbol{y},K,\pi,\boldsymbol{\lambda},\varphi) = \frac{1-\varphi}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+}} \sum_{k=0}^{\min(K,\boldsymbol{y})} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=1}^m \frac{\lambda_i^{y_i-k}\mathrm{e}^{-\lambda_i}}{(y_i-k)!} \text{ and}$$

$$S\left(\boldsymbol{y}^{(2)},K,\pi,\boldsymbol{\lambda},\nu^{(2)}\right) = \frac{1-\nu^{(2)}}{1-(1-\pi)^K \mathrm{e}^{-\lambda_+^{(2)}}} \sum_{k=0}^{\min(K,\boldsymbol{y}^{(2)})} \binom{K}{k} \pi^k (1-\pi)^{K-k} \prod_{i=r+1}^m \frac{\lambda_i^{y_i-k}\mathrm{e}^{-\lambda_i}}{(y_i-k)!}.$$

We first consider Case I: $\boldsymbol{y}^{(2)} \neq \mathbf{0}$. Under Case I, it is clear that $\boldsymbol{y} \neq \mathbf{0}$. From (4.12), it is easy to obtain

$$\Pr\left(\mathbf{y}^{(1)} = \boldsymbol{y}^{(1)}|\mathbf{y}^{(2)} = \boldsymbol{y}^{(2)}\right) = \frac{\mathrm{e}^{-\lambda_+^{(1)}} \sum_{k=0}^{\min\{K,\boldsymbol{y}\}} \binom{K}{k}\pi^k(1-\pi)^{K-k} \prod_{j=1}^m \frac{\lambda_j^{y_j-k}}{(y_j-k)!}}{\sum_{l=0}^{\min\{K,\boldsymbol{y}^{(2)}\}} \binom{K}{l}\pi^l(1-\pi)^{K-l} \prod_{p=r+1}^m \frac{\lambda_p^{y_p-l}}{(y_p-l)!}}.$$

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 (4.12), we obtain

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

When $\boldsymbol{y}^{(1)} \neq \mathbf{0}$, from (4.12), we have

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

4.3.2 Conditional distribution of $Z'|\boldsymbol{y}$

Since $Z' \sim \text{Bernoulli}(1-\varphi)$, 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,

$$Z'|(\mathbf{y} = \boldsymbol{y}) \sim \begin{cases} \text{Degenerate}(0), & \text{if } \boldsymbol{y} = \mathbf{0}, \\ \text{Degenerate}(1), & \text{if } \boldsymbol{y} \neq \mathbf{0}, \end{cases} \tag{4.13}$$

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

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

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

$$\mathbf{w}|(\mathbf{y} = \boldsymbol{y} \neq \mathbf{0}) \sim \text{Degenerate}(\boldsymbol{y}). \tag{4.14}$$

5 Likelihood-based methods for multivariate ZACS distribution without covariates

Suppose that $\mathbf{y}_j \overset{\text{iid}}{\sim} \text{ZACS}(\varphi; K, \pi, \lambda_1, \ldots, \lambda_m)$, where $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{mj})^\top$ for $j = 1, \ldots, n$. 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. 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 $\mathbb{J}$. We assume that K is a known positive integer. Therefore, the observed-data likelihood function is proportional to

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

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

$$\ell(\varphi, \pi, \boldsymbol{\lambda}|Y_{\text{obs}}) = \ell_1(\varphi|Y_{\text{obs}}) + \ell_2(\pi, \boldsymbol{\lambda}|Y_{\text{obs}}), \tag{5.1}$$

where

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

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

$$\hat{\varphi} = \frac{m_0}{n}, \tag{5.2}$$

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

5.1 MLEs via the EM algorithm and bootstrap CIs

The objective of this section is to find the MLEs of $(\pi, \boldsymbol{\lambda})$ based on (5.1). For the log-likelihood function (3.1), the corresponding EM iteration for finding the MLEs of $(\pi, \boldsymbol{\lambda})$ is defined by (3.3)–(3.5). By comparing (3.1) with (5.1), if we replace $(\sum_{j=1}^{n} w_{ij})$ in (3.1) with $(\sum_{j\notin\mathbb{J}} y_{ij})$, we promptly obtain the MLEs of $(\pi, \boldsymbol{\lambda})$ by using the EM algorithm. The M-step is to calculate the complete-data MLEs:

$$\hat{\pi} = \frac{\sum_{j\notin\mathbb{J}} x_{0j}^*}{(n-m_0)K} \quad \text{and} \quad \hat{\lambda}_i = \frac{\sum_{j\notin\mathbb{J}} u_j y_{ij}}{(n-m_0)} - K\hat{\pi}, \qquad i = 1, \ldots, m, \tag{5.3}$$

and the E-step is to replace $\{u_j\}_{j\notin\mathbb{J}}$ and $\{x_{0j}^*\}_{j\notin\mathbb{J}}$ in (5.3) by their conditional expectations:

$$E(U_j|Y_{\text{obs}}, \pi, \boldsymbol{\lambda}) = E(U_j) = 1 - (1-\pi)^K \mathrm{e}^{-\lambda_+}, \quad \text{and} \tag{5.4}$$

$$E(X_{0j}^*|Y_{\text{obs}}, \pi, \boldsymbol{\lambda}) = \frac{[1-(1-\pi)^K \mathrm{e}^{-\lambda_+}] \sum\limits_{k_j=1}^{\min(K,\boldsymbol{y}_j)} \binom{K}{k_j} \pi^{k_j} (1-\pi)^{K-k_j} \prod\limits_{i=1}^{m} \frac{\lambda_i^{y_{ij}-k_j}}{(y_{ij}-k_j)!}}{\sum\limits_{k_j=0}^{\min(K,\boldsymbol{y}_j)} \binom{K}{k_j} \pi^{k_j} (1-\pi)^{K-k_j} \prod\limits_{i=1}^{m} \frac{\lambda_i^{y_{ij}-k_j}}{(y_{ij}-k_j)!}} \times I(\min(\boldsymbol{y}_j) \geq 1), \tag{5.5}$$

respectively.

The procedure of constructing bootstrap CIs for an arbitrary function of $(\varphi, \pi, \boldsymbol{\lambda})$, say $\vartheta = h(\varphi, \pi, \boldsymbol{\lambda})$, is very similar to that presented in Section 3.2.

6 Simulation studies

To evaluate the performance of the proposed methods in Section 3, we investigate the accuracy of MLEs and confidence interval estimators of the parameters in the multivariate ZTCS distribution. We consider two cases for the dimension with $m = 2$ and $m = 3$.

6.1 Experiment 1: $m = 2$

When $m = 2$, the parameters $(K, \pi; \lambda_1, \lambda_2)$ are set to be (5, 0.5; 3, 5). We generate $\{\mathbf{w}_j\}_{j=1}^{n} \overset{\text{iid}}{\sim} \text{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_m)$ with $n = 200$. Based on this simulated data set, for different K values we first calculate the MLEs of π and (λ_1, λ_2) by using the EM algorithm (3.3)–(3.5) and then calculate the estimated log-likelihood. We choose $K = 5$ that maximizes the log-likelihood among all K values. These results are reported in Table 1.

Table 1 Finding the value of K by maximizing the log-likelihood function for $m = 2$

K	$\hat{\pi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	Log-likelihood
3	0.66742	3.4739056	5.4464959	−884.6629
4	0.58617	3.1314781	5.1040629	−883.4288
5	0.51145	2.9188539	4.8914319	−883.0151
6	0.44308	2.8176227	4.7901937	−883.1228

For this fixed value of $K = 5$, we first calculate the MLEs of $(\pi, \lambda_1, \lambda_2)$ by using the EM algorithm (3.3)–(3.5), the bootstrap *standard deviations* (stds) of these MLEs, the corresponding *mean square errors* (MSEs) and two 95 % bootstrap *confidence intervals* (CIs) of these parameters with $G = 1000$ by the bootstrap method presented in Section 3.2. Then, we independently repeat the above process 1000 times. The resulting average MLE, std, MSE and two *coverage probabilities* (CPs) based on the normal-based and non-normal-based bootstrap samples, respectively, are displayed in Table 2.

From Table 2, we can see that the average MSE of $\hat{\pi}$ is very small while the average MSEs of $(\hat{\lambda}_1, \hat{\lambda}_2)$ are reasonably small. The two bootstrap coverage probabilities are close to but less than 0.95.

6.2 Experiment 2: $m = 3$

When $m = 3$, the parameters $(K, \pi; \lambda_1, \lambda_2, \lambda_3)$ are set to be (4, 0.3; 2, 4, 6). We generate $\{\mathbf{w}_j\}_{j=1}^n \overset{\text{iid}}{\sim} \text{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_m)$ with $n = 200$. Based on this simulated data set, for different K values we first calculate the MLEs of π and $(\lambda_1, \lambda_2, \lambda_3)$ by using the EM algorithm (3.3)–(3.5) and then calculate the estimated log-likelihood. We choose $K = 4$ that maximizes the log-likelihood among all K values. These results are reported in Table 3.

For this fixed value of $K = 4$, we first calculate the MLEs of $(\pi, \lambda_1, \lambda_2, \lambda_3)$ by using the EM algorithm (3.3)–(3.5), the bootstrap stds of these MLEs, the corresponding MSEs and two 95 % bootstrap CIs of these parameters with $G = 1000$ by the bootstrap method presented in Section 3.2. Then, we independently repeat the above process 1000 times. The resulting average MLE, std, MSE and two CPs based on the normal-based and non-normal-based bootstrap samples, respectively, are displayed in Table 4.

From Table 4, we can see that the average MSEs of $\hat{\pi}$ and $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ are very small. The two bootstrap coverage probabilities are close to 0.95.

7 Two real examples

7.1 Students' absenteeism data

In this section, we use the data set on the number of absences of 113 students from a lecture course in two successive semesters reported by Karlis (2003) to illustrate the proposed statistical methods for the multivariate ZTCS distribution. Let W_1 denote the number of absences in the first semester and W_2 denote the number of absences in the second semester. The data are displayed in Table 5 below.

For the purpose of illustration, we artificially remove the $(0, 0)$ cell counts from Table 5 and the updated data are shown in Table 6.

Let $\mathbf{w}_j = (W_{1j}, W_{2j})^\top \overset{\text{iid}}{\sim} \text{ZTCS}(K, \pi; \lambda_1, \lambda_2)$ for $j = 1, \ldots, n$ with $n = 98$. Let $\boldsymbol{w}_j = (w_{1j}, w_{2j})^\top$ denote the realization of the random vector $\mathbf{w}_j$, and $Y_{\text{obs}} = \{\boldsymbol{w}_j\}_{j=1}^n$ be the observed data. The parameter K of the binomial distribution is considered unknown

Table 2 The average MLE, std, MSE and two CPs of $(\pi, \lambda_1, \lambda_2)$ for $m = 2$ and $K = 5$

Parameter	True value	Average MLE	Average std^B	Average MSE	CP†	CP‡
π	0.5	0.511457	0.06417	0.004209	0.930	0.932
λ_1	3	2.918853	0.33051	0.114730	0.927	0.932
λ_2	5	4.891431	0.35926	0.139568	0.921	0.928

std^B: The sample standard deviation for the bootstrap samples
CP†: Normal-based bootstrap CP
CP‡: Non-normal-based bootstrap CP

Table 3 Finding the value of K by maximizing the log-likelihood function for $m = 3$

K	$\hat{\pi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	Log-likelihood
3	0.4037377	1.9887834	4.2987810	5.9287793	−650.5100604
4	0.3203847	1.9184570	4.2284540	5.8584519	−649.8799727
5	0.2583191	1.9084000	4.2183967	5.8483944	−650.1021029
6	0.2137641	1.9174109	4.2274075	5.8574051	−650.1044541

Table 4 The average MLE, std, MSE and two CPs of $(\pi, \lambda_1, \lambda_2)$ for $m = 3$ and $K = 4$

Parameter	True value	Average MLE	Average std^B	Average MSE	CP†	CP‡
π	0.3	0.320384	0.0541052	0.00295	0.937	0.939
λ_1	2	1.918457	0.222460	0.05689	0.925	0.932
λ_2	4	4.228454	0.242082	0.06476	0.921	0.925
λ_3	6	5.858451	0.255456	0.09402	0.954	0.948

std^B: The sample standard deviation for the bootstrap samples
CP†: Normal-based bootstrap CP
CP‡: Non-normal-based bootstrap CP

Table 5 Cross-tabulation of the students' absenteeism data (Karlis 2003)

$W_1 \backslash W_2$	0	1	2	3	4	5	6	7	8	Total
0	15	10	4	4	2	0	0	0	0	35
1	6	11	9	4	2	0	0	0	0	32
2	5	7	6	5	0	0	0	0	0	23
3	1	3	2	4	3	1	0	0	0	14
4	1	0	2	0	1	0	0	0	0	4
5	0	0	0	0	0	1	1	0	0	2
6	0	0	0	0	0	0	2	0	0	2
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	1	1
Total	29	31	23	17	8	2	3	0	0	113

Table 6 The number of absences of 113 students from a course in two successive semesters without the (0, 0) cell counts (Karlis 2003)

$W_1 \backslash W_2$	0	1	2	3	4	5	6	7	8	Total
0	–	10	4	4	2	0	0	0	0	20
1	6	11	9	4	2	0	0	0	0	32
2	5	7	6	5	0	0	0	0	0	23
3	1	3	2	4	3	1	0	0	0	14
4	1	0	2	0	1	0	0	0	0	4
5	0	0	0	0	0	1	1	0	0	2
6	0	0	0	0	0	0	2	0	0	2
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	1
9	1	0	0	0	0	0	0	0	1	1
Total	14	31	23	17	8	2	3	0	0	98

and it is attempted to estimate this. Based on the data in Table 6, for different K values we first calculate the MLEs of π and (λ_1, λ_2) by using the EM algorithm (3.3)–(3.5) and then calculate the estimated values of the log-likelihood function. These results are reported in Table 7.

We should choose the K that maximizes the log-likelihood among all K values. From Table 7, we observed that the values of log-likelihood monotonically increase as $K \to \infty$. On the other hand, K must be larger than or equal to $\max(W_1, W_2)$. From Table 6, we have $\max(W_1, W_2) = 9$. To illustrate how to obtain the confidence intervals of the parameters, it seems reasonable to choose $K = 10$. With $G = 6000$ bootstrap replications, we calculate the bootstrap average MLEs, the bootstrap stds of $(\hat{\pi}, \hat{\lambda}_1, \hat{\lambda}_2)$ and two 95 % bootstrap CIs of $(\pi, \lambda_1, \lambda_2)$. These results are listed in Table 8.

7.2 Road accident data of Athens

The number of accidents in 24 roads of Athens for the period 1987–1991 were reported and analyzed by Karlis (2003) with a multivariate Poisson distribution. Since only accidents that caused injuries are included as shown in Table 9, we want to fit the data set by the multivariate ZTCS model.

Let $\mathbf{w}_j = (W_{1j}, \ldots, W_{5j})^\top \overset{\text{iid}}{\sim} \text{ZTCS}(K, \pi; \lambda_1, \ldots, \lambda_5)$, where $W_{1j}, \ldots, W_{5j}$ denote the average numbers of accidents reported in the j-th road per kilometer from 1987 to 1991, respectively, for $j = 1, \ldots, n$ $(n = 24)$. For example, when $j = 1$, we have $t_j = t_1 = 1.2$ and

$$(W_{11}, \ldots, W_{51})^\top = (11, 33, 25, 23, 6)^\top / 1.2.$$

The unknown parameter K is assumed to be an positive integer. Based on the data in Table 9, for different K values we first calculate the MLEs of π and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_5)^\top$ by using the EM algorithm (3.3)–(3.5) and then calculate the estimated values of the log-likelihood function. These results are reported in Table 10.

Table 7 Finding the value of K by maximizing the log-likelihood function for fitting the data of Table 6 by the multivariate ZTCS distribution

K	$\hat{\pi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	Log-likelihood
2	0.1220034	1.394314	1.600330	−328.8195
3	0.1024329	1.326026	1.531414	−328.1680
4	0.0869704	1.281892	1.486834	−327.7241
5	0.0748624	1.252965	1.457593	−327.4137
8	0.0517887	1.208834	1.412942	−326.8897
9	0.0468272	1.200906	1.404915	−326.7862
10	0.0427041	1.194675	1.398604	−326.7022
14	0.0314939	1.179167	1.382890	−326.4824
15	0.0295422	1.176680	1.380369	−326.4453
20	0.0225358	1.168154	1.371725	−326.3146
30	0.0152633	1.160049	1.363504	−326.1831
50	0.0092682	1.153806	1.357169	−326.0775
75	0.0062141	1.150805	1.354123	−326.0247
100	0.0046735	1.149330	1.352626	−325.9982
150	0.0031242	1.147871	1.351145	−325.9718
250	0.0018786	1.146717	1.349972	−325.9507
350	0.0013431	1.146225	1.349473	−325.9416

Table 8 MLEs and confidence intervals of parameters for the students' absenteeism data

Parameter	MLE^B	std^B	95 % bootstrap CI†	95 % bootstrap CI‡
π	0.043961	0.017672	[0.009325, 0.078598]	[0.007857, 0.078343]
λ_1	1.175550	0.210425	[0.763118, 1.587982]	[0.785992, 1.606149]
λ_2	1.377584	0.218819	[0.948699, 1.806468]	[0.965662, 1.824147]

MLE^B: The average MLE for the bootstrap samples
std^B: The sample standard deviation for the bootstrap samples
CI†: Normal-based bootstrap CI
CI‡: Non-normal-based bootstrap CI

We should choose the K that maximizes the log-likelihood among all K values. From Table 10, we observed that the values of log-likelihood monotonically increase as $K \to \infty$. On the other hand, K must be larger than or equal to $\max\{W_{ij}: 1 \le i \le 5,\ 1 \le j \le 24\}$. From Table 9, we have $\max\{W_{ij}: 1 \le i \le 5,\ 1 \le j \le 24\} = 52.7$. To illustrate how to obtain the confidence intervals of the parameters, it seems reasonable to choose $K = 53$. With $G = 6000$ bootstrap replications, we calculate the bootstrap average MLEs, the bootstrap stds of $(\hat{\pi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_5)$ and two 95 % bootstrap CIs of $(\pi, \lambda_1, \ldots, \lambda_5)$. These results are reported in Table 11.

Based on the data in Table 9, we calculate the sample correlation coefficient matrix, which is given by

$$\mathbf{R} = \begin{pmatrix} 1.0000 & 0.8038 & 0.7643 & 0.8089 & 0.5746 \\ 0.8038 & 1.0000 & 0.8326 & 0.8297 & 0.4084 \\ 0.7643 & 0.8326 & 1.0000 & 0.9058 & 0.5768 \\ 0.8089 & 0.8297 & 0.9058 & 1.0000 & 0.6557 \\ 0.5746 & 0.4084 & 0.5768 & 0.6557 & 1.0000 \end{pmatrix},$$

Table 9 Accident data of 24 roads in Athens for the period 1987–1991 (Karlis 2003)

Road	j	Year					Length(km)
		1987	1988	1989	1990	1991	t_j
Akadimias	1	11	33	25	23	6	1.2
Alexandras	2	41	63	91	77	29	2.6
Amfitheas	3	5	35	44	21	13	2.4
Aharnon	4	44	79	91	88	33	5.5
Vas. Olgas	5	5	3	4	4	0	0.5
Vas. Konstantinou	6	8	15	26	13	7	1.3
Vas. Sofias	7	34	63	81	67	23	2.6
Vouliagmenis	8	17	16	24	24	4	2.1
G' Septemvriou	9	16	24	30	30	13	1.7
Galatsioy	10	13	13	15	17	9	1.1
Iera Odos	11	7	15	20	19	8	2.7
Kalirois	12	15	24	39	32	7	2.6
Katehaki	13	2	3	27	24	7	1.4
Kifisias	14	22	23	38	22	11	1.4
Kifisou	15	38	48	60	53	24	7.9
Leof. Kavalas	16	4	6	12	9	3	2.0
Lenorman	17	19	30	37	48	22	2.0
Leof. Athinon	18	15	11	16	21	28	6.1
Mesogeion	19	20	30	33	28	9	1.5
P. Ralli	20	13	14	13	17	9	2.6
Panepistimiou	21	24	58	40	36	5	1.1
Patision	22	80	108	114	113	86	4.1
Peiraios	23	86	89	109	90	49	8.0
Sigrou	24	60	61	87	86	29	4.8

Table 10 Finding the value of K by maximizing the log-likelihood function for fitting the data of Table 9 by the multivariate ZTCS distribution

K	$\hat{\pi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	Log-likelihood
2	0.396	8.405	13.520	16.667	14.513	5.262	−498.992
3	0.383	8.046	13.161	16.308	14.154	4.903	−492.749
4	0.372	7.707	12.822	15.969	13.815	4.564	−487.185
5	0.349	7.452	12.567	15.714	13.559	4.309	−482.820
8	0.292	6.858	11.973	15.120	12.966	3.715	−473.529
9	0.276	6.710	11.825	14.972	12.818	3.567	−471.336
10	0.261	6.585	11.700	14.847	12.693	3.442	−469.490
14	0.209	6.271	11.387	14.533	12.379	3.128	−464.604
15	0.198	6.226	11.341	14.488	12.333	3.083	−463.810
20	0.155	6.092	11.207	14.354	12.200	2.949	−461.170
30	0.106	5.997	11.112	14.259	12.105	2.854	−458.806
50	0.065	5.944	11.059	14.206	12.051	2.800	−457.125
51	0.064	5.942	11.057	14.204	12.050	2.799	−457.078
52	0.062	5.941	11.056	14.203	12.049	2.798	−457.033
53	0.061	5.940	11.055	14.202	12.047	2.797	−456.990
54	0.060	5.939	11.054	14.201	12.046	2.795	−456.949
75	0.043	5.923	11.038	14.185	12.030	2.779	−456.350
100	0.032	5.913	11.028	14.175	12.021	2.770	−455.978
150	0.021	5.905	11.020	14.167	12.012	2.762	−455.617
250	0.013	5.898	11.014	14.160	12.006	2.755	−455.334
350	0.009	5.896	11.011	14.158	12.003	2.753	−455.215

while the population correlation coefficient matrix $\boldsymbol{\rho}$, based on (2.6) is estimated to be

$$\hat{\boldsymbol{\rho}} = \begin{pmatrix} 1.0000 & 0.2703 & 0.2444 & 0.2612 & 0.4199 \\ 0.2703 & 1.0000 & 0.1951 & 0.2085 & 0.3352 \\ 0.2444 & 0.1951 & 1.0000 & 0.1886 & 0.3031 \\ 0.2612 & 0.2085 & 0.1886 & 1.0000 & 0.3240 \\ 0.4199 & 0.3352 & 0.3031 & 0.324 & 1.0000 \end{pmatrix}.$$

it can be easily seen that $\hat{\boldsymbol{\rho}}$ is very close to $\mathbf{R}$.

Table 11 MLEs and confidence intervals of parameters for the road accident data of Athens

Parameter	MLE^{B}	std^{B}	95 % bootstrap $\text{CI}^{\dagger}$	95 % bootstrap $\text{CI}^{\ddagger}$
π	0.0663	0.017	[0.0327, 0.1000]	[0.0356, 0.1027]
λ_1	5.8610	1.217	[3.4750, 8.2470]	[3.8330, 8.4930]
λ_2	10.926	2.347	[6.3260, 15.526]	[7.1450, 16.313]
λ_3	14.118	1.824	[10.543, 17.694]	[10.844, 18.022]
λ_4	11.954	1.624	[8.7700, 15.138]	[9.2400, 15.478]
λ_5	2.7533	0.604	[1.5676, 3.9390]	[1.7988, 4.0889]

MLE^{B}: The average MLE for the bootstrap samples
std^{B}: The sample standard deviation for the bootstrap samples
$\text{CI}^{\dagger}$: Normal-based bootstrap CI
$\text{CI}^{\ddagger}$: Non-normal-based bootstrap CI

8 Concluding remarks

In this paper, we first proposed the multivariate ZTCS distribution and studied its distributional properties. Since the joint marginal distribution of any r-dimensional sub-vector of the multivariate ZTCS random vector of m-dimensional has certain probability mass function, we then proposed the multivariate ZACS distribution. It is noted that the multivariate ZTCS distribution is a special case of the multivariate ZACS distribution. The EM algorithm is used to obtain the MLEs of the parameters in the multivariate ZACS distribution. The multivariate ZTCS distribution can be used when other distributions, like multivariate zero-truncated Poisson distribution is not a good fit to some real data sets. Meanwhile, the multivariate ZACS distribution, as a more general form, can be used in a much wider range. It can be a good substitute for the Type II multivariate ZTP distribution (Tian et al. 2014).

Endnote

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

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
XD contributed to the most of the work, DJ to the simulation studies (Section 6) and the second application (Section 7.2), and GLT was the supervisor and reviewed all work from the initial idea, through preparation of the manuscript until the final version. All authors read and approved the final manuscript.

Received: 15 May 2015 Accepted: 20 July 2015
Published online: 04 August 2015

References
Karlis, D: An EM algorithm for multivariate Poisson distribution and related models. J. Appl. Stat. **30**(1), 63–77 (2003)
Loukas, S: Some methods of estimation for a trivariate Poisson distribution. Zastoswania Math. **21**, 503–510 (1993)
Loukas, S, Papageorgiou, H: On a trivariate Poisson distribution. Appl. Math. **36**, 432–439 (1991)
Ong, SH: A discrete Charlier series distribution. Biom. J. **30**(8), 1003–1009 (1988)
Papageorgiou, H, Loukas, S: A bivariate discrete Charlier Series distribution. Biom. J. **37**(1), 105–117 (1995)
Tian, GL, Liu, Y, Tan, MT: Type II multivariate zero-truncated/adjusted Poisson models and their applications. Technical Report at Department of Statistics and Actuarial Science. The University of Hong Kong (2014)