Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Statistical inference for zero-and-one-inflated poisson models

Yincai Tang, Wenchen Liu & Ancha Xu

STATISTICAL THEORY AND RELATED FIELDS, 2017
VOL. 1, NO. 2, 216–226
https://doi.org/10.1080/24754269.2017.1400419

Statistical inference for zero-and-one-inflated poisson models

Yincai Tang[a], Wenchen Liu[a] and Ancha Xu[b]

[a]School of Statistics, East China Normal University, Shanghai, China; [b]College of Mathematics and Information Science, Wenzhou University, Zhejiang, China

ABSTRACT

In this paper, a zero-and-one-inflated Poisson (ZOIP) model is studied. The maximum likelihood estimation and the Bayesian estimation of the model parameters are obtained based on data augmentation method. A simulation study based on proposed sampling algorithm is conducted to assess the performance of the proposed estimation for various sample sizes. Finally, two real data-sets are analysed to illustrate the practicability of the proposed method.

ARTICLE HISTORY
Received 22 July 2017
Revised 30 October 2017
Accepted 31 October 2017

KEYWORDS
Zero-inflated Poisson model; zero-and-one-inflated Poisson model; MLE; Bayesian estimate; EM algorithm; latent variable; Gibbs sampling

1. Introduction

In data analysis and modelling, the problem of excess zeros has been massively investigated. Neyman (1939) and Feller (1943) introduced the concept of zero inflation to address the problem of excess zeros. Zero-inflated models have successfully been used in manufacturing defects (Lambert, 1992), medicine (Cheung, 2002), patent applications (Crepon & Duguet, 1997), species abundance (Faddy, 1998; Welsh, Cunningham, Donnelly, & Lindenmayer, 1996), Legionellosis infection (Xu, Xie, & Goh, 2014) and in many other fields. One of their main interesting features is that they adjust well to data issued from a particular mixture of two populations: one in which one has only zero counts and another in which the counts are the realisations of a discrete distribution. Various methods have been developed to address this issue, in which the zero-inflated Possion (ZIP) model proposed by Lambert (1992) plays an important part. ZIP model with Bayesian techniques has been considered by Ghosh, Mukhopadhyay, and Lu (2006), Chen (2009), Dagne (2010), and Musio, Sauleau, and Buemi (2010). For modelling complete female fertility, Melkersson and Rooth (2000) proposed a zero-and-two-inflated count data model, which accounts for a relative excess of both zero and two children. However, in many cases, count data may contain excess zeros and ones simultaneously. For example, it is most probable that in a shopping trip one does not buy anything or just buys one item at a clothing store; and one may be infected by some virus for at most one time due to the generation of corresponding antibodies once after the infection. The Legionellosis infection count data in Xu et al. (2014) contains many zeros and ones, and we have noted that the count of one is nearly half underestimated by the ZIP model. Motivated by this example, we studied the zero-and-one-inflated Poisson (ZOIP) model for count data with both excessive zeros and ones. Zhang, Tian, and NG (2016) initially studied the likelihood-based ZOIP model.

A nonnegative integer-valued random variable Y in a ZOIP model can be represented as $Y = V(1 - B_1) + B_1(1 - B_2)$, where B_1 is a Bernoulli random variable with success probability p_0, B_2 is a Bernoulli random variable with success probability p_1, V follows a Poisson distribution with rate parameter θ, and B_1, B_2 and V are mutually independent. The relation between Y and (B_1, B_2, V) is

$$\begin{cases} (Y=0) \Leftrightarrow (V=0, B_1=0) \cup (B_1=1, B_2=1), \\ (Y=1) \Leftrightarrow (V=1, B_1=0) \cup (B_1=1, B_2=0), \\ (Y=k) \Leftrightarrow (V=k, B_1=0),\ k=2,3,\ldots \end{cases} \tag{1}$$

Then the probability mass function of Y is

$$\Pr(Y=k) = \begin{cases} p_0 p_1 + (1-p_0)\,\mathrm{e}^{-\theta}, & \text{if } k=0, \\ p_0(1-p_1) + (1-p_0)\,\theta \mathrm{e}^{-\theta}, & \text{if } k=1, \\ (1-p_0)\,\frac{\theta^k}{k!}\mathrm{e}^{-\theta}, & \text{if } k \geq 2, \end{cases} \tag{2}$$

with $0 \leq p_0 \leq 1$, $0 \leq p_1 \leq 1$, and $\theta > 0$. We denote this zero-and-one-inflated Poisson model as ZOIP (p_0, p_1, θ). An EM algorithm to get maximum likelihood estimation (MLE) from incomplete data is proposed by Dempster, Laird, and Rubin (1977). In this paper, B_1, B_2 and V are looked upon as latent variables to help get the MLE via EM algorithm and the Bayesian estimation via Gibbs sampling for ZOIP (p_0, p_1, θ).

The rest of the paper is organised as follows. In Section 2, the maximum likelihood estimates of the parameters (p_0, p_1, θ) are obtained and shown to be unique

CONTACT Yincai Tang yctang@stat.ecnu.edu.cn

under a mild condition. We focus our attention on the Bayesian estimation in Section 3. The latent variable method is used to get the MLE and Bayesian estimates. A simulation study is conducted in Section 4 to compare the performance of MLE and Bayesian estimates. Finally two real data-sets are analysed in Section 5 to illustrate the practicability of the proposed method. Our conclusions are presented in the final section.

2. Maximum likelihood estimation

2.1. Usual maximum likelihood estimate

Given a random sample $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ of size n from ZOIP (p_0, p_1, θ), the likelihood function of (p_0, p_1, θ) is

$$\begin{aligned} L\left(p_0, p_1, \theta | \boldsymbol{Y}\right) &\propto \left[p_0 p_1 + \left(1 - p_0\right) \mathrm{e}^{-\theta}\right]^{S_0} \\ &\times \left[p_0\left(1 - p_1\right) + \left(1 - p_0\right) \theta \mathrm{e}^{-\theta}\right]^{S_1} \\ &\times \left(1 - p_0\right)^{n - S_0 - S_1} \theta^S \mathrm{e}^{-(n - S_0 - S_1)\theta}, \end{aligned} \tag{3}$$

where $S_0 = S_0(\boldsymbol{Y}) = \sharp\{i : Y_i = 0\}$, $S_1 = S_1(\boldsymbol{Y}) = \sharp\{i : Y_i = 1\}$, $S = S(\boldsymbol{Y}) = \sum_{Y_i \geq 2} Y_i$. Here, $\sharp X$ is defined to be the number of elements in the set X.

Let q_0 and q_1 be the probabilities of Y being zero and one respectively, i.e.,

$$\begin{cases} q_0 = p_0 p_1 + (1 - p_0)\mathrm{e}^{-\theta}, \\ q_1 = p_0(1 - p_1) + (1 - p_0)\theta \mathrm{e}^{-\theta}. \end{cases} \tag{4}$$

Then the likelihood function of (q_0, q_1, θ) becomes

$$L\left(q_0, q_1, \theta | \boldsymbol{Y}\right) \propto q_0^{S_0} q_1^{S_1} \left(1 - q_0 - q_1\right)^{n - S_0 - S_1} \times \frac{\theta^S \mathrm{e}^{-(n - S_0 - S_1)\theta}}{\left(1 - \mathrm{e}^{-\theta} - \theta \mathrm{e}^{-\theta}\right)^{n - S_0 - S_1}}. \tag{5}$$

Then it is easy to get the MLEs of q_0 and q_1

$$\hat{q}_i = \frac{S_i}{n}, i = 0, 1, \tag{6}$$

and the MLE of θ, $\hat{\theta}$, is the solution of the following equation:

$$S\left(\mathrm{e}^\theta - \theta - 1\right) - (n - S_0 - S_1)\theta\left(\mathrm{e}^\theta - 1\right) = 0, \tag{7}$$

which can be solved numerically according to the Newton-Raphson iterative algorithm.

From (4) it is easy to check that the transformation between (p_0, p_1, θ) and (q_0, q_1, θ) is one-to-one. Thus, based on the invariance property for the maximum likelihood estimation, we get the MLEs of p_0 and p_1 as follows

$$\hat{p}_0 = \frac{\hat{q}_0 + \hat{q}_1 - (1 + \hat{\theta})\mathrm{e}^{-\hat{\theta}}}{1 - (1 + \hat{\theta})\mathrm{e}^{-\hat{\theta}}},$$

$$\hat{p}_1 = \frac{\hat{q}_0 - (1 - \hat{p}_0)\mathrm{e}^{-\hat{\theta}}}{\hat{p}_0}.$$

Furthermore, we have the following property.

Theorem 2.1: *If at least one observation is larger than one, i.e. $n - S_0 - S_1 > 0$, then the maximum likelihood estimation of ZOIP model (2) uniquely exist.*

The proof is given in the Appendix A.1.

2.2. MLE with EM algorithm

Next we provide an EM algorithm to calculate the MLE of (p_0, p_1, θ). As mentioned in Section 1, a ZOIP random variable Y can be represented in terms of three independent latent variables: two Bernoulli variables B_1 and B_2, and a Poisson random variable V. Thus, if we could observe these latent variables, then the likelihood would become three likelihood functions multiplied together. To be more specific, let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ be a sample from ZOIP model (2), with corresponding latent variables $\boldsymbol{B_1} = (B_{11}, B_{12}, \ldots, B_{1n})$, $\boldsymbol{B_2} = (B_{21}, B_{22}, \ldots, B_{2n})$ and $\boldsymbol{V} = (V_1, V_2, \ldots, V_n)$. If we could observe B_{1i}, B_{2i} and thus observe V_i, then according to $Y_i = V_i(1 - B_{1i}) + B_{1i}(1 - B_{2i})$ and the relation (1), the augmented likelihood function with the augmented data $(\boldsymbol{B_1}, \boldsymbol{B_2})$ or $(\boldsymbol{B_1}, \boldsymbol{B_2}, \boldsymbol{V})$ would be

$$\begin{aligned} &L_c(p_0, p_1, \theta | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2}) \\ &= \prod_{i=1}^n p_0^{B_{1i}}(1 - p_0)^{(1 - B_{1i})} \prod_{i=1}^n p_1^{B_{2i}}(1 - p_1)^{(1 - B_{2i})} \\ &\times \prod_{i=1}^n \left[\frac{\theta^{Y_i}}{Y_i!}\mathrm{e}^{-\theta}\right]^{(1 - B_{1i})} \\ &\triangleq L_c(p_0 | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2}) L_c(p_1 | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2}) L_c(\theta | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2}). \end{aligned} \tag{8}$$

The maximisation of Equation (8) is easy, since $L_c(p_0 | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2})$, $L_c(p_1 | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2})$, and $L_c(\theta | \boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2})$ can be maximised separately.

With the EM algorithm, the MLEs of p_1, p_2 and θ can be found iteratively alternating between two steps, first the E-step by the expectation with respect to (B_{1i}, B_{2i}) under the current estimates of (p_0, p_1, θ), and then the M-step through maximising $L_c(p_0, p_1, \theta | Y, B_1, B_2)$ with the (B_{1i}, B_{2i}) fixed at their expected values from the E step. See more for example, Dempster et al. (1977). When the estimated values of (p_0, p_1, θ) converges, the iteration stops and the final iteration gives the MLEs of p_0, p_1 and θ for the full likelihood (8).

In more detail, the $(k + 1)$-th iteration of the EM algorithm can be described as follows.

E step: Under the current estimates $(p_0^{(k)}, p_1^{(k)}, \theta^{(k)})$, find

$$B_{1i}^{(k+1)} = P(B_{1i}=1|Y_i, p_0^{(k)}, p_1^{(k)}, \theta^{(k)}) = \begin{cases} \dfrac{p_0^{(k)}p_1^{(k)}}{p_0^{(k)}p_1^{(k)} + (1-p_0^{(k)})P(V=0)}, & \text{if } Y_i = 0, \\ \dfrac{p_0^{(k)}(1-p_1^{(k)})}{p_0^{(k)}(1-p_1^{(k)}) + (1-p_0^{(k)})P(V=1)}, & \text{if } Y_i = 1, \\ 0, & \text{if } Y_i = k, \\ & k = 2, 3, \ldots \end{cases} \quad (9)$$

$$B_{2i}^{(k+1)} = P(B_{2i}=1|Y_i, p_0^{(k)}, p_1^{(k)}, \theta^{(k)}) = \begin{cases} \dfrac{p_0^{(k)}p_1^{(k)} + (1-p_0^{(k)})p_1^{(k)}P(V=0)}{p_0^{(k)}p_1^{(k)} + (1-p_0^{(k)})P(V=0)}, & \text{if } Y_i = 0, \\ \dfrac{(1-p_0^{(k)})p_1^{(k)}P(V=1)}{p_0^{(k)}(1-p_1^{(k)}) + (1-p_0^{(k)})P(V=1)}, & \text{if } Y_i = 1, \\ p_1^{(k)}, & \text{if } Y_i \\ & = k, k = 2, 3, \ldots \end{cases} \quad (10)$$

M step: Let $\boldsymbol{B_1^{(k+1)}} = (B_{11}^{(k+1)}, \ldots, B_{1n}^{(k+1)})$ and $\boldsymbol{B_2^{(k+1)}} = (B_{21}^{(k+1)}, \ldots, B_{2n}^{(k+1)})$. Maximising $L_c(p_0|\boldsymbol{Y}, \boldsymbol{B_1^{(k+1)}}, \boldsymbol{B_2^{(k+1)}})$, $L_c(p_1|\boldsymbol{Y}, \boldsymbol{B_1^{(k+1)}}, \boldsymbol{B_2^{(k+1)}})$ and $L_c(\theta|\boldsymbol{Y}, \boldsymbol{B_1^{(k+1)}}, \boldsymbol{B_2^{(k+1)}})$, respectively, gives immediately

$$p_0^{(k+1)} = \frac{\sum_{i=1}^n B_{1i}^{(k+1)}}{n}, \quad p_1^{(k+1)} = \frac{\sum_{i=1}^n B_{2i}^{(k+1)}}{n},$$
$$\theta^{(k+1)} = \frac{\sum_{i=1}^n (1-B_{1i}^{(k+1)})Y_i}{\sum_{i=1}^n (1-B_{1i}^{(k+1)})}.$$

The proof of Equations (9) and (10) is given in Appendix A.2.

Theorem 2.2: *Let $\boldsymbol{\eta} = (p_0, p_1, \theta)$ and $\{\boldsymbol{\eta}^{(k)}\}$ be a sequence generated by this EM algorithm. Then according to the Theorem 3 in Wu (1983), all the limit points of instance $\{\boldsymbol{\eta}^{(k)}\}$ are local maxima of $L(\boldsymbol{\eta}|\boldsymbol{Y})$ and $L(\boldsymbol{\eta}^{(k)}|\boldsymbol{Y})$ converges monotonically to $L^\star = L(\boldsymbol{\eta}^\star|\boldsymbol{Y})$ for some local maximum $\boldsymbol{\eta}^\star$.*

The proof is given in Appendix A.3.

3. Bayesian estimation

In this section, the Bayesian analysis is studied. Laplace noninformative flat prior and conjugate prior are the usual prior distributions. However, for the two forms of the ZOIP model, the exact conjugate priors cannot be derived. The noninformative priors for objective Bayesian method is very complicated and the posterior sampling can hardly be realised, which is thoroughly discussed in another paper (Liu, Tang, & Xu, 2017). The current paper focuses on the comparison between the Bayesian approach under the naive flat prior and maximum likelihood estimation. And we also consider choosing a seemingly conjugate prior in the first data analysis to match the MLE. Here the flat prior is used, then the observed posterior distribution of (p_0, p_1, θ) is

$$\begin{aligned} \pi\left(p_0, p_1, \theta|\boldsymbol{Y}\right) &\propto L(p_0, p_1, \theta|\boldsymbol{Y}) \\ &\propto \left[p_0p_1 + \left(1-p_0\right)\mathrm{e}^{-\theta}\right]^{S_0} \\ &\quad \times \left[p_0\left(1-p_1\right) + \left(1-p_0\right)\theta\mathrm{e}^{-\theta}\right]^{S_1} \\ &\quad \times \left(1-p_0\right)^{n-S_0-S_1}\theta^S\mathrm{e}^{-(n-S_0-S_1)\theta}. \end{aligned}$$

Expanding the right part of the above equation, we have

$$\pi\left(p_0, p_1, \theta|\boldsymbol{Y}\right) \propto \sum_i p_0^{l_{1i}}p_1^{l_{2i}}\left(1-p_0\right)^{l_{3i}}\left(1-p_1\right)^{l_{4i}}\theta^S\mathrm{e}^{-(n-l_1i)\theta}. \quad (11)$$

Here $0 \le l_{ji} \le n$, $j = 1, 2, 3, 4$, and the maximum value of l_{1i} is $S_0 + S_1$. It is easy to find that the joint posterior distribution of (p_0, p_1, θ) with the flat prior is proper, when $S \ge 2$. This condition is weak since it will be met if there exist at least one observation larger than one.

The joint posterior distribution of (p_0, p_1, θ) is a nonstandard density. Though the MCMC methods such as Gibbs sampling can be used directly, instead, we make use of the data augmentation method proposed by Tanner and Wong (1987) with the help of the latent variables mentioned in the EM algorithm.

Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ be a sample from ZOIP (p_0, p_1, θ), with corresponding latent variables $\boldsymbol{B_1} = (B_{11}, B_{12}, \ldots, B_{1n})$, $\boldsymbol{B_2} = (B_{21}, B_{22}, \ldots, B_{2n})$ and $\boldsymbol{V} = (V_1, V_2, \ldots, V_n)$. Then the joint posterior distribution of $(p_0, p_1, \theta, \boldsymbol{V}, \boldsymbol{B_1}, \boldsymbol{B_2})$ given the observed data $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ is

$$\pi(p_0, p_1, \theta, \boldsymbol{B_1}, \boldsymbol{B_2}, \boldsymbol{V}|\boldsymbol{Y}) = \prod_{i=1}^n p_0^{B_{1i}}(1-p_0)^{(1-B_{1i})}p_1^{B_{2i}}(1-p_1)^{(1-B_{2i})} \times \left[\frac{\theta^{V_i}}{V_i!}\mathrm{e}^{-\theta}\right]^{(1-B_{1i})}. \quad (12)$$

The posterior samples can be obtained using Gibbs sampler in blocks of latent variables $(\boldsymbol{B_1}, \boldsymbol{B_2}, \boldsymbol{V})$ and model parameters (p_0, p_1, θ).

First, from $Y = V(1-B_1) + B_1(1-B_2)$, we can get the condition predictive distribution of the latent data (B_1, B_2, V), $\pi(B_1, B_2, V|Y=y, p_0, p_1, \theta)$, given one observation $Y = y$.

Lemma 3.1: *Based on the relation $Y = V(1-B_1) + B_1(1-B_2)$, we can get*

$$P(B_1 = i, B_2 = j, V = v|Y = 0, p_0, p_1, \theta)$$

$$= \begin{cases} \dfrac{(1-p_0)(1-p_1)P(V=0)}{p_0p_1+(1-p_0)P(V=0)}, & \text{if } v=i=j=0,\\ \dfrac{(1-p_0)p_1P(V=0)}{p_0p_1+(1-p_0)P(V=0)}, & \text{if } v=i=0,\ j=1,\\ \dfrac{p_0p_1P(V=v)}{p_0p_1+(1-p_0)P(V=0)}, & \text{if } i=j=1,\ v=0,\\ & 1,\ldots,\\ 0, & \text{otherwise;} \end{cases} \quad (13)$$

$$P(B_1=i,B_2=j,V=v|Y=1,p_0,p_1,\theta)$$
$$= \begin{cases} \dfrac{(1-p_0)(1-p_1)P(V=1)}{p_0(1-p_1)+(1-p_0)P(V=1)}, & \text{if } v=1,\ i=j=0,\\ \dfrac{(1-p_0)p_1P(V=1)}{p_0(1-p_1)+(1-p_0)P(V=1)}, & \text{if } v=j=1,\ i=0,\\ \dfrac{p_0(1-p_1)P(V=v)}{p_0(1-p_1)+(1-p_0)P(V=1)}, & \text{if } i=1,\ j=0,\\ & v=0,1,\ldots,\\ 0, & \text{otherwise;} \end{cases} \quad (14)$$

and

$$P(B_1=i,B_2=j,V=v|Y=v,p_0,p_1,\theta) = \begin{cases} 1-p_1, & \text{if } i=j=0,\ v=2,3,\ldots,\\ p_1, & \text{if } i=0,\ j=1,\ v=2,3,\ldots,\\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

The proof is given in Appendix A.2.

Then from (13), (14) and (15), we can obtain the samples of $(\boldsymbol{V},\boldsymbol{B_1},\boldsymbol{B_2})$ given observations of $\boldsymbol{Y}$ as follows.

- If $Y_i=0$ is observed, then according to (13) and the relation (1), we can assume a coin is flipped with probability of $\frac{p_0p_1}{p_0p_1+(1-p_0)P(V=0)}$ getting head. If it comes up a head, set $B_{1i}=1, B_{2i}=1$ and draw V_i from the distribution of V with parameter θ. If it comes up a tail, set $B_{1i}=0$, $V_i=0$ and a coin is tossed with probability of a head being p_1. Then if a head occurs, set $B_{2i}=1$. If a tail occurs, set $B_{2i}=0$.
- If $Y_i=1$ is observed, then according to (14) and the relation (1), a coin with probability $\frac{p_0(1-p_1)}{p_0(1-p_1)+(1-p_0)P(V=1)}$ of getting head is flipped. If it comes up a head, set $B_{1i}=1$, $B_{2i}=0$ and draw V_i from the distribution of V with parameter θ. If it comes up a tail, set $B_{1i}=0$, $V=1$, and a coin is tossed with probability of a head being p_1. If a head occurs, set $B_{2i}=1$. And if tail occurs, set $B_{2i}=0$.
- If $Y_i=v$ $(v=2,3,\ldots)$ is observed, then according to (15) and the relation (1), a coin is flipped with probability p_1 of getting a head. If it comes up a head, set $B_{1i}=0$, $B_{2i}=1$ and $V_i=v$. If it comes up a tail, set $B_{1i}=0$, $B_{2i}=0$ and $V_i=v$.

Second, it is easy to get from (12) the augmented posterior distribution, $\pi(p_0,p_1,\theta|\,\boldsymbol{Y},\boldsymbol{B_1},\boldsymbol{B_2},\boldsymbol{V})$, from which to get the posterior samples of (p_0,p_1,θ) from their full conditional distributions given below.

$$p_0|p_1,\theta,\boldsymbol{B_1},\boldsymbol{B_2},\boldsymbol{V},\boldsymbol{Y} \sim \text{Beta}\left(\sum_{i=1}^n B_{1i}+1,\ n+1-\sum_{i=1}^n B_{1i}\right).$$

$$p_1|p_0,\theta,\boldsymbol{B_1},\boldsymbol{B_2},\boldsymbol{V},\boldsymbol{Y} \sim \text{Beta}\left(\sum_{i=1}^n B_{2i}+1,\ n+1-\sum_{i=1}^n B_{2i}\right),$$

and

$$\theta|p_0,p_1,\boldsymbol{B_1},\boldsymbol{B_2},\boldsymbol{V},\boldsymbol{Y} \sim \text{Gamma}\left(\sum_{i=1}^n (1-B_{1i})V_i+1,\ n-\sum_{i=1}^n B_{1i}\right).$$

Thus, the Bayesian inference of the ZOIP model (2) can be performed based on the Gibbs sampling procedure below.

(1) Set initial values for $p_0^{(0)}$, $p_1^{(0)}$ and $\theta^{(0)}$.
(2) For $t=1,2,\ldots$, perform the following iterative update
    (a) Use data augmentation steps to get $(B_{1i}^{(t)},B_{2i}^{(t)},V_i^{(t)})$, $i=1,2,\ldots,n$ with the parameters $p_0^{(t-1)}$, $p_1^{(t-1)}$ and $\theta^{(t-1)}$.
    (b) Sample $p_0^{(t)}$ from $\text{Beta}(\sum_{i=1}^n B_{1i}^{(t)}+1, n+1-\sum_{i=1}^n B_{1i}^{(t)})$.
    (c) Sample $p_1^{(t)}$ from $\text{Beta}(\sum_{i=1}^n B_{2i}^{(t)}+1, n+1-\sum_{i=1}^n B_{2i}^{(t)})$.
    (d) Sample $\theta^{(t)}$ from $\text{Gamma}(\sum_{i=1}^n (1-B_{1i}^{(t)})V_i^{(t)}+1, n-\sum_{i=1}^n B_{1i}^{(t)})$.

4. Simulation study

In this section, we will assess the performance of the maximum likelihood and Bayesian estimation methods proposed for model (2). Two simulations are conducted. The case of p_0 and p_1 around 0.5 is studied in the first simulation, in which the sample sizes are set to $n=20$, 50 and 100, the value of p_0 is set to 0.3 and 0.4, the value of p_1 is set to 0.4 and 0.5, the value of θ is set to 3, 5 and 8, the confidence level α is set to 95% and all simulations are replicated for 10,000 times. The second simulation studies the case of extreme values of p_0 and p_1, in which the sample sizes are set to 50, and 100, the value of p_0 is set to 0.1 and 0.7, the value of p_1 is set to 0.4 and 0.8, the value of θ is set to 3 and 8, the confidence level α is also set to 95% and all simulations are still replicated for 10,000 times. In both cases, the initial values of $p_0^{(0)}$ and $p_1^{(0)}$ are generated from uniform distribution and the initial value of $\theta^{(0)}$ is generated from gamma distribution. When we do simulation, according to ZOIP model (2), the sample of size n is reserved only when $n-S_0-S_1>0$, considering the existence of

Table 1. RMSE of parameter estimation for model (2) of first group simulation.

θ	p_0	p_1	n	θ_M	θ_B	p_{0M}	p_{0B}	p_{1M}	p_{1B}
3	0.3	0.4	20	0.5990	0.6569	0.1535	0.1113	0.2754	0.1482
			50	0.3772	0.4108	0.1052	0.0926	0.1725	0.1424
			100	0.2073	0.2815	0.0791	0.0749	0.1203	0.1151
		0.5	20	0.6005	0.6642	0.1604	0.1102	0.2755	0.1424
			50	0.4468	0.4032	0.1093	0.0929	0.1947	0.1355
			100	0.2910	0.3178	0.0756	0.0729	0.1344	0.1216
	0.4	0.4	20	0.6503	0.6463	0.1624	0.1257	0.2597	0.1487
			50	0.4134	0.4455	0.1061	0.1070	0.1329	0.1359
			100	0.2844	0.3113	0.0770	0.0829	0.0897	0.0997
		0.5	20	0.6681	0.6431	0.1717	0.1309	0.2424	0.1461
			50	0.3967	0.4281	0.1123	0.1056	0.1629	0.1252
			100	0.2531	0.3069	0.0784	0.0778	0.1046	0.0952
5	0.3	0.4	20	0.6494	0.6707	0.1149	0.0981	0.2285	0.1632
			50	0.3646	0.4296	0.0707	0.0673	0.1371	0.1227
			100	0.2813	0.2773	0.0490	0.0506	0.0962	0.0903
		0.5	20	0.7142	0.7024	0.1114	0.0995	0.2285	0.1564
			50	0.4596	0.4232	0.0703	0.0660	0.1466	0.1235
			100	0.3289	0.2721	0.0517	0.0492	0.1007	0.0936
	0.4	0.4	20	0.6551	0.6768	0.1173	0.1050	0.1937	0.1499
			50	0.4350	0.4072	0.0717	0.0716	0.1156	0.1112
			100	0.3548	0.2618	0.0536	0.0511	0.0819	0.0752
		0.5	20	0.6312	0.6424	0.1186	0.1062	0.1998	0.1501
			50	0.4237	0.3839	0.0832	0.0704	0.1221	0.1075
			100	0.2877	0.3509	0.0654	0.0501	0.0861	0.0777
8	0.3	0.4	20	0.7349	0.6891	0.1010	0.0922	0.2077	0.1541
			50	0.4947	0.4625	0.0634	0.0634	0.1305	0.1150
			100	0.3122	0.3386	0.0468	0.0443	0.0911	0.0835
		0.5	20	0.7401	0.7231	0.1041	0.0991	0.2172	0.1501
			50	0.4025	0.4726	0.0657	0.0622	0.1357	0.1164
			100	0.3879	0.3098	0.0465	0.0442	0.0910	0.0877
	0.4	0.4	20	0.6491	0.7369	0.1085	0.0990	0.1805	0.1461
			50	0.3996	0.3671	0.0689	0.0691	0.1105	0.1020
			100	0.2739	0.3319	0.0498	0.0485	0.0780	0.0766
		0.5	20	0.7049	0.6702	0.1090	0.0969	0.1765	0.1454
			50	0.4349	0.4147	0.0772	0.0657	0.1100	0.1015
			100	0.3273	0.3411	0.0533	0.0495	0.0797	0.0777

MLE of θ. The method proposed in Section 2.1 is used to calculate the MLE in our simulation. And the asymptotic normality is used to derive the confidence interval. Here, the two-sided equal tail confidence intervals are used. The comparison results for the root mean squared errors are listed in Tables 1 and 2 and the comparison results for coverage probabilities are listed in Tables 3 and 4. In these tables, the subscripts M and B represent the maximum likelihood estimation and Bayesian estimation, respectively.

For the point estimate of θ, the MLE performs slightly better than the Bayeisan estimate when θ is small but they are similar when θ is large. For the point estimates of p_0 and p_1, the MLE performs slightly worse than the Bayesian estimates when n is small and the MLE and Bayesian estimates perform similarly when n is large. For the interval estimates of θ, p_0 and p_1, the coverage probabilities based on the Bayesian estimates are generally more accurate than that based on the MLE. As the sample size increases, the accuracy

Table 2. RMSE of parameter estimation for model (2) of second group simulation.

θ	p_0	p_1	n	θ_M	θ_B	p_{0M}	p_{0B}	p_{1M}	p_{1B}
3	0.1	0.4	50	0.3027	0.3602	0.1018	0.0845	0.1607	0.1382
			100	0.2311	0.2663	0.0884	0.0686	0.1354	0.1286
		0.8	50	0.3252	0.3938	0.0936	0.1108	0.1867	0.1431
			100	0.2278	0.2475	0.0872	0.0709	0.1483	0.1304
	0.7	0.4	50	0.6208	0.6129	0.0914	0.1065	0.1101	0.1319
			100	0.3601	0.4035	0.0633	0.0576	0.0611	0.0723
		0.8	50	0.6032	0.6485	0.0867	0.1309	0.2424	0.2298
			100	0.3215	0.3927	0.0708	0.0550	0.1072	0.1131
8	0.1	0.4	50	0.4516	0.3821	0.0906	0.0852	0.1301	0.1445
			100	0.3209	0.2973	0.0501	0.0555	0.0640	0.0619
		0.8	50	0.4239	0.4059	0.0952	0.1022	0.1302	0.1490
			100	0.2727	0.2998	0.0484	0.0612	0.1026	0.0944
	0.7	0.4	50	0.8610	0.8911	0.0991	0.1107	0.0956	0.1113
			100	0.4601	0.4335	0.0585	0.0613	0.0605	0.0682
		0.8	50	0.8225	0.8408	0.0906	0.1144	0.1904	0.1887
			100	0.4220	0.4298	0.0504	0.0599	0.0935	0.1056

Table 3. Coverage probabilities of confidence intervals for model (2) of first group simulation.

θ	p_0	p_1	n	θ_M	θ_B	p_{0M}	p_{0B}	p_{1M}	p_{1B}
3	0.3	0.4	20	0.9315	0.9701	0.9030	0.9751	0.9067	0.9871
			50	0.9282	0.9433	0.9285	0.9456	0.9325	0.9706
			100	0.9422	0.9516	0.9347	0.9443	0.9190	0.9602
		0.5	20	0.9290	0.9632	0.8992	0.9782	0.9179	0.9892
			50	0.9466	0.9443	0.9195	0.9491	0.9254	0.9655
			100	0.9488	0.9404	0.9393	0.9410	0.9215	0.9476
	0.4	0.4	20	0.9172	0.9544	0.9132	0.9656	0.9287	0.9684
			50	0.9323	0.9362	0.9359	0.9524	0.9280	0.9472
			100	0.9430	0.9393	0.9300	0.9365	0.9322	0.9487
		0.5	20	0.9211	0.9587	0.9140	0.9577	0.9247	0.9725
			50	0.9510	0.9344	0.9295	0.9452	0.9312	0.9643
			100	0.9373	0.9455	0.9493	0.9441	0.9474	0.9614
5	0.3	0.4	20	0.9375	0.9391	0.9196	0.9572	0.9008	0.9720
			50	0.9535	0.9404	0.9430	0.9581	0.9235	0.9550
			100	0.9453	0.9411	0.9415	0.9464	0.9492	0.9570
		0.5	20	0.9535	0.9474	0.9260	0.9666	0.9165	0.9650
			50	0.9553	0.9473	0.9360	0.9626	0.9190	0.9513
			100	0.9492	0.9571	0.9412	0.9442	0.9372	0.9472
	0.4	0.4	20	0.9485	0.9383	0.9244	0.9570	0.9133	0.9622
			50	0.9473	0.9416	0.9410	0.9512	0.9285	0.9486
			100	0.9435	0.9527	0.9390	0.9528	0.9401	0.9552
		0.5	20	0.9444	0.9499	0.9273	0.9530	0.9123	0.9714
			50	0.9427	0.9412	0.9335	0.9590	0.9225	0.9606
			100	0.9493	0.9491	0.9421	0.9540	0.9346	0.9551
8	0.3	0.4	20	0.9455	0.9502	0.9183	0.9725	0.9533	0.9655
			50	0.9543	0.9670	0.9493	0.9521	0.9173	0.9466
			100	0.9450	0.9601	0.9305	0.9602	0.9240	0.9577
		0.5	20	0.9371	0.9523	0.9282	0.9483	0.9105	0.9661
			50	0.9482	0.9366	0.9429	0.9586	0.9044	0.9391
			100	0.9503	0.9611	0.9306	0.9553	0.9325	0.9497
	0.4	0.4	20	0.9511	0.9502	0.9422	0.9713	0.9190	0.9605
			50	0.9465	0.9565	0.9470	0.9475	0.9345	0.9531
			100	0.9470	0.9411	0.9353	0.9456	0.9481	0.9493
		0.5	20	0.9425	0.9560	0.9402	0.9754	0.9086	0.9583
			50	0.9521	0.9373	0.9425	0.9532	0.9360	0.9626
			100	0.9452	0.9522	0.9422	0.9431	0.9445	0.9410

Table 4. Coverage probabilities of confidence intervals for model (2) of second group simulation.

θ	p_0	p_1	n	θ_M	θ_B	p_{0M}	p_{0B}	p_{1M}	p_{1B}
3	0.1	0.4	50	0.9390	0.9360	0.9005	0.9173	0.9126	0.9280
			100	0.9343	0.9662	0.9327	0.9389	0.9401	0.9552
		0.8	50	0.9353	0.9406	0.9119	0.9252	0.9264	0.9128
			100	0.9569	0.9601	0.9445	0.9331	0.9292	0.9400
	0.7	0.4	50	0.9218	0.9240	0.9334	0.9460	0.9244	0.9308
			100	0.9409	0.9357	0.9306	0.9390	0.9528	0.9612
		0.8	50	0.9309	0.9194	0.9405	0.9270	0.9086	0.9277
			100	0.9388	0.9620	0.9504	0.9445	0.9321	0.9646
8	0.1	0.4	50	0.9339	0.9448	0.9206	0.9342	0.9330	0.9247
			100	0.9445	0.9551	0.9384	0.9569	0.9396	0.9428
		0.8	50	0.9418	0.9362	0.9276	0.9230	0.9387	0.9451
			100	0.9350	0.9493	0.9628	0.9439	0.9360	0.9501
	0.7	0.4	50	0.9343	0.9225	0.9391	0.9328	0.9286	0.9308
			100	0.9569	0.9409	0.9372	0.9395	0.9310	0.9361
		0.8	50	0.9446	0.9301	0.9324	0.9289	0.9305	0.9242
			100	0.9400	0.9446	0.9371	0.9418	0.9497	0.9319

Table 5. Fitted frequencies and parameter estimation for legionellosis data, ZIP model.

	Frequency estimation Count of legionellosis cases					Estimation of	
	0	1	2	3	4		
Observed frequency	36	23	3	0	1	θ	p
MLE(Poisson)	34	14	3	0	0	0.423 (0.246,0.600)	
MLE(ZIP)	36	11	4	1	0	0.675 (0.160,1.190)	0.692 (0.567,0.818)
Bayes_J(ZIP)	35	11	4	1	0	0.725 (0.291,1.328)	0.682 (0.553,0.799)
Bayes_R(ZIP)	36	11	4	1	0	0.725 (0.291,1.328)	0.689 (0.559,0.805)

Table 6. Fitted frequencies and parameter estimation for legionellosis data, ZOIP model.

	Frequency estimation Count of legionellosis cases					Estimation of		
	0	1	2	3	4			
Observed frequency	36	23	3	0	1	θ	p_0	p_1
MLE(ZOIP)	36	23	3	1	0	1.229 (0.294,2.165)	0.817 (0.644,0.990)	0.633 (0.488,0.778)
Bayes(ZOIP)	32	24	5	2	0	0.895 (0.388,2.488)	0.503 (0.046,0.910)	0.603 (0.264,0.859)

of all the estimates increases. The result of the second simulation shows that the larger p_0 leads to worse estimation of θ, in that the data has a smaller probability coming from the Poisson distribution.

Table 7. Model comparison for legionellosis data: AIC, DIC, WAIC1, WAIC2.

	MLE	Bayes		
Model	AIC	DIC	WAIC1	WAIC2
ZIP	66.18	123.27	124.12	124.34
ZOIP	40.53	121.10	121.65	122.01

5. Real data analysis

5.1. Singapore legionnaires disease data

In this subsection, one example about Legionnaires disease in Singapore from the healthcare industry is presented to illustrate our method and this data-set was analysed by Xu et al. (2014). Legionellosis (Legionnaires disease and Pontiac fever) is an acute respiratory infection caused by gram negative, rod-shaped bacteria of the genus Legionella (Lam, Ang, Tan, James, & Goh, 2011). In Singapore, legionnaires disease has been recognised as a potential public health threat. In order to make relevant control polices, it is useful to know the distribution of the counts of legionellosis cases. For illustration, here we apply our model in the study of the weekly legionellosis count data in the year 2005. The data were reported by the Ministry of Health of Singapore. Xu et al. (2014) derived the Jeffreys prior and reference prior for the ZIP model, and then presented the Bayesian fitted frequencies and compared with likelihood method for both the ZIP and pure Poisson models. See Table 5 for their detailed results. The estimation results of ZOIP model (2) is presented in Table 6. The 95% confidence or credible intervals are put in the parentheses. Furthermore, the observed and fitted frequency distributions for different models based on MLE and Bayesian estimation are shown in Figure 1.

As is noted by Xu et al. (2014), the difference between the estimation accuracy of the ZIP model and the Poisson model is not clear according to the fitted frequency distributions. According to the fitted frequency distribution shown in Figure 1, the frequency of one is underestimated overall in Xu et al. (2014) (the estimated value is nearly half of the observed frequency). In our result, both the MLEs and Bayes estimates for all frequencies are closer to the observed values. Besides, the estimation of parameter θ using the ZOIP model is nearly twice of the estimation by the ZIP model. And the results of AIC (Akaike information criterion) values with MLEs for the ZIP and ZOIP models are presented in Table 7. The AIC value of the ZOIP model is smaller than the value of the ZIP model. The results of DIC (deviance information criterion) and WAIC (Watanabe-Akaike information criterion) given by Gelman et al. (2014) are also presented in Table 7 for Bayesian model comparison. For the ZIP model, we only list those values under Jeffreys prior, which is quite the same as those under the reference prior. For the ZOIP model, we use the flat priors. Both of them show that the ZOIP model is more appropriate than the ZIP model under all the criteria. The fitted frequencies also show that the Bayes ZOIP underperforms MLE in this example. This may due to the influence of the prior. To judge the effect of the prior on the final posterior inference, we tried some seemingly conjugate priors, the beta prior for p_0 and p_1 and gamma prior for θ. We found that beta(5,2) and beta(2,2) for p_0 and p_1 and gamma(2.5,2.5) for θ give the similar result as the ZOIP MLE. To reduce the sensitivity of the prior, objective Bayesian method can be considered. Please refer to the detailed derivation and examples given by Liu et al. (2017). They do give consistent result when the sample size is moderate.

5.2. US Detroit accidental death data

In this section, one accidental data-set from Detroit, Michigan, is used to demonstrate the zero-and-one-inflated Poisson model introduced in the previous section. Here, we apply our model in the study of the daily accidental deaths data in the year 1994 available from the NMMAPS of air pollution and health in the United States. The fitted frequency distributions based on the MLE of a Poisson model and zero-inflated Poisson model (Xu et al. 2014) are presented in Table 8 and the ML and Bayesian estimation results of ZOIP model (2) are presented in Table 9. The 95% confidence or credible intervals are put in the parentheses. Furthermore, the observed and fitted frequency distributions are shown in Figure 2.

Table 8. Fitted frequencies and parameter estimation for accidental death data, ZIP model.

	Frequency estimation Count of accidental deaths								Estimation of	
	0	1	2	3	4	5	6	7		
Observed frequency	181	122	28	25	5	2	1	1	θ	p
MLE(Poisson)	162	132	53	14	3	0	0	0	0.8110 (0,2.5760)	
MLE(ZIP)	181	105	54	19	5	1	0	0	1.0402 (1.0236,1.0568)	0.4959 (0.4945,0.4972)
$Bayes_J$(ZIP)	181	104	54	19	5	1	0	0	1.0416 (1.0094,1.0718)	0.4962 (0.4943,0.4974)
$Bayes_R$(ZIP)	181	105	54	19	5	1	0	0	1.0397 (1.0098,1.0704)	0.4960 (0.4943,0.4976)

Table 9. Fitted frequencies and parameter estimation for accidental death data, ZOIP model.

	Frequency estimation Count of accidental deaths								Estimation of			
	0	1	2	3	4	5	6	7				
Observed frequency	181	122	28	25	5	2	1	1	θ	p_0	p_1	
MLE(ZOIP)	181	122	31	19	8	3	1	0	1.8168 (1.4225,2.2111)	0.6866 (0.6346,0.7372)	0.6480 (0.5959,0.6926)	
Bayes(ZOIP)	181	122	31	19	8	3	1	0		1.8128 (1.3728,2.2659)	0.6481 (0.6097,0.6974)	0.6484 (0.6065,0.7028)

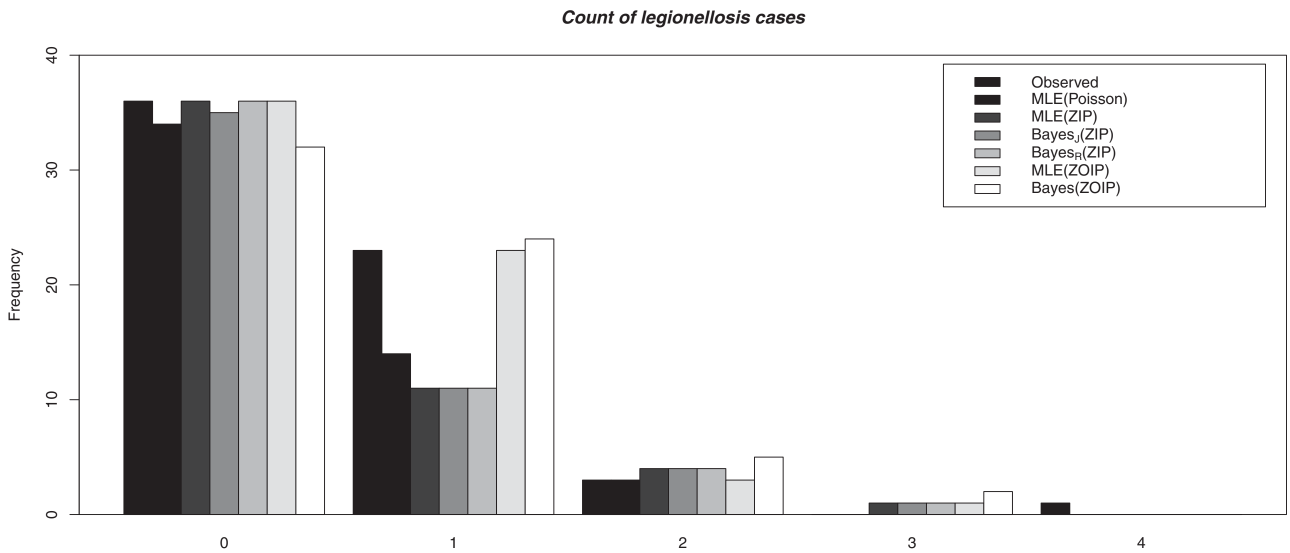

Figure 1. Frequency distributions for legionellosis cases data.

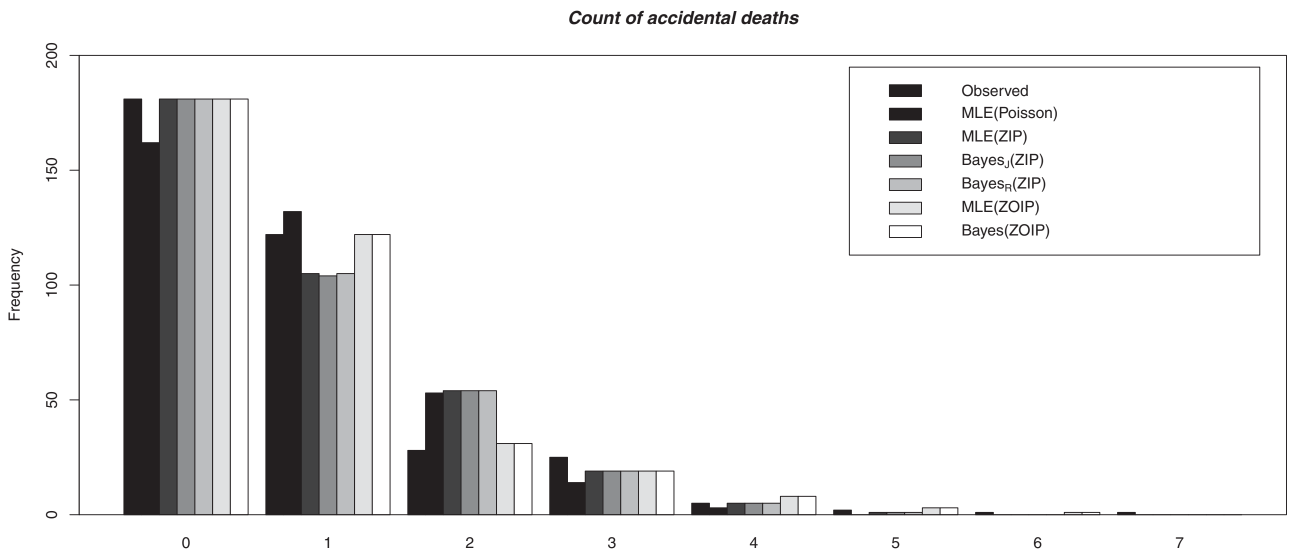

Figure 2. Frequency distributions for accidental death data.

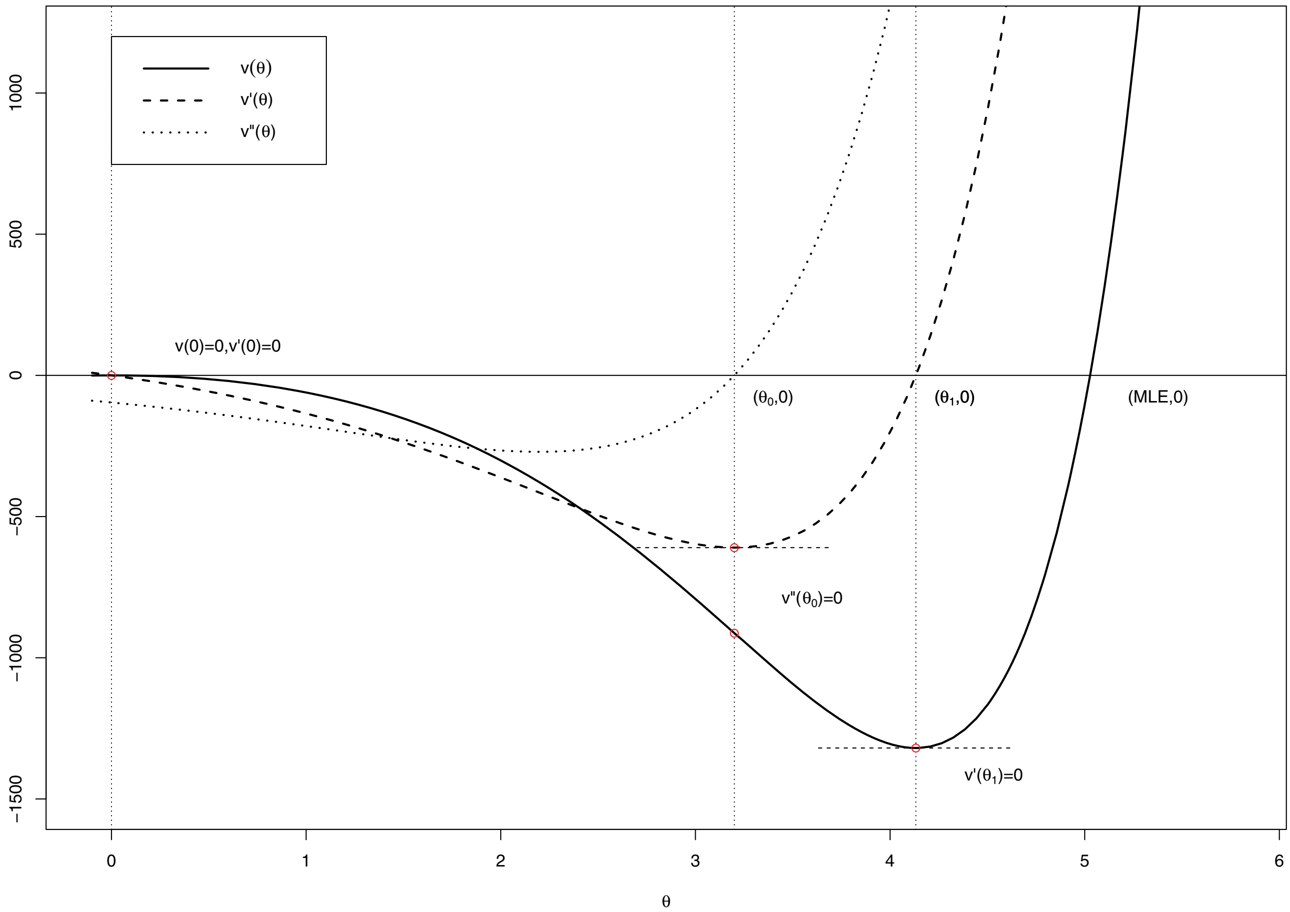

Figure 3. Illustration of one and only one solution for Equation (7).

Table 10. Model comparison for accidental death data: AIC, DIC, WAIC1, WAIC2.

	MLE	Bayes		
Model	AIC	DIC	WAIC1	WAIC2
ZIP	918.52	918.52	919.06	919.09
ZOIP	913.65	898.62	898.75	898.91

According to the fitted frequencies shown in Table 8 and Figure 2, the frequency of one is underestimated overall while the frequency of two is overestimated by the ZIP model (the estimated value of two is nearly double of the true frequency). By our model, both the MLE and Bayesian estimation of all the frequencies are closer to the observed ones. And the results of AIC for MLE model fitting, and DIC and WAIC for Bayesian model fitting under Jeffreys prior for the ZIP model and flat prior for the ZOIP model are presented in Table 10 . The results also show that in this case study the ZOIP model is more appropriate than the ZIP model.

6. Conclusions

In this paper, we proposed a zero-and-one-inflated Poisson model. The EM algorithm to get the MLE and Gibbs sampling to get the samples from the posterior distribution based on latent variables are proposed. The Bayesian method is compared with MLE via Monte Carlo simulation. Simulation results show that the Bayesian estimates perform slightly better when the sample size is small or moderate. Two real data-sets are analysed through the new method and compared with the results given by Xu et al. (2014) in terms of AIC, DIC and WAIC criteria. Both MLE and Bayesian estimates have better performance than those given by Xu et al. (2014). This study also shows that the prior will influence the posterior inference when the sample size in not large enough. Better noninformative priors than the flat prior can be considered in the future study.

Acknowledgments

The research is supported by the Natural Science Foundation of China (Nos. 11271136, 81530086, 11671303, 11201345), the 111 Project of China (No. B14019), the Natural Science Foundation of Zhejiang Province (No. LY15G010006) and the China Postdoctoral Science Foundation (No. 2015M572598).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

National Natural Science Foundation of China (CN) [grant number 11671303], [grant number 11201345]: Ministry of Education of the People's Republic of China (CN) [grant number B14019]; China Postdoctoral Science Foundation (CN) [grant number 2015M572598]; National Natural

Science Foundation of China (CN) [grant number 11271136], [grant number 81530086]; Natural Science Foundation of Zhejiang Province (CN) [grant number LY15G010006].

Notes on contributors

Yincai Tang is professor of statistics in the College of Statistics, East China Normal University, Shanghai, China. He received his PhD degree from East China Normal University. His professional publications and research interests have focused on lifetime data analysis, degradation data analysis, big data analysis and Bayesian inference.

Wenchen Liu is a PhD candidate in the College of Statistics, East China Normal University, Shanghai, China. His research interests include Bayesian statistics, reliability statistics and big data analytics.

Ancha Xu is an associate professor in the Department of Statistics, Wenzhou University, Zhejiang, China. He received his PhD degree from East China Normal University, Shanghai, China, in 2011. His research interests include Bayesian statistics, degradation modeling, and accelerated life testing.

References

Cheung, Y. B. (2002). Zero-inflated models for regression analysis of count data: A study of growth and development. *Statistics in Medicine*, *21*, 1462–1469.

Crepon, B., & Duguet, E. (1997). Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. *Journal of Econometrics*, *79*, 355–378.

Dagne, G. A. (2010). Bayesian semiparametric zero-inflated Poisson model for longitudinal count data. *Mathematical Biosciences*, *224*, 126–130.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Faddy, M. (1998). Stochastic models for analysis of species abundance data. In D.J. Fletcher, L. Kavalieris, & B. F. J. Manly (Eds.), *Statistics in Ecology and Environmental Monitoring 2: Decision Making and Risk Assessment in Biology* (pp. 33–40). University of Otago Press.

Feller, W. (1943). On a general class of contagious distributions. *Annals of Mathematical Statistics*, *16*, 319–329.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, B. D., Vehtari, A., & Rubin D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Ghosh, S. K., Mukhopadhyay, P., & Lu, J. C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, *136*, 1360–1375.

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

Lam, M. C., Ang, L. W., Tan, A. L., James, L., & Goh, K. T. (2011). Epidemiology and control of legionellosis, Singapore. *Emerging Infectious Diseases*, *17*, 1209–1215.

Liu, W. C., Tang, Y. C., & Xu, A. C. (2017). A zero-and-one inflated Poisson model and its application. *Statistics and Its Interface*, Accepted.

Melkersson, M., & Rooth, D. (2000). Modeling female fertility using inflated count data models. *Journal of Population Economics*, *13*, 189–203.

Musio, M., Sauleau, E. A., & Buemi A. (2010). Bayesian semiparametric ZIP models with space-time interactions: An application to cancer registry data. *Mathematical Medicine and Biology*, *27*, 181–194.

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

Tanner, M., & Wong, W. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, *82*, 528–550.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, *11*, 95–103.

Welsh, A., Cunningham, R., Donnelly, C., & Lindenmayer, D. (1996). Modeling the abundance of rare species: Statistical models for counts with extra zeros. *Ecological Modelling*, *88*, 297–308.

Xu, H. Y., Xie, M., & Goh, T. N. (2014). Objective bayes analysis of zero-inflated poisson distribution with application to healthcare data. *IIE Transactions*, *46*, 843–852.

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

Appendix

A.1

Proof of Theorem 2.1: Let $\upsilon(\theta) = (n - S_0 - S_1)\theta e^{\theta} - Se^{\theta} - (n - S - S_0 - S_1)\theta + S$. Then,

$$\begin{aligned}\upsilon'(\theta) &= (n - S_0 - S_1)\,\theta e^{\theta} + (n - S - S_0 - S_1)\,e^{\theta} \\ &\quad - (n - S - S_0 - S_1)\,, \\ \upsilon''(\theta) &= (n - S_0 - S_1)\,\theta e^{\theta} \\ &\quad + (2n - S - 2S_1 - 2S_0)\,e^{\theta}.\end{aligned}$$

According to the definition of S and the condition that $n - S_0 - S_1 > 0$, we have $(2n - S - 2S_1 - 2S_0) < 0$. Let

$$\theta_0 = \frac{S + 2S_1 + 2S_0 - 2n}{n - S_0 - S_1}.$$

It immediately follows that $\upsilon''(\theta) > 0$ when $\theta > \theta_0$ and $\upsilon''(\theta) < 0$ when $\theta < \theta_0$. That is, θ_0 is an inflection point of $\upsilon(\theta)$. So $\upsilon'(\theta)$ is decreasing on $(0, \theta_0)$ and increasing on $(\theta_0, +\infty)$. Additionally, it can be easily verified that $\upsilon'(0) = 0$ and $\upsilon'(+\infty) > 0$. Therefore, there exists $\theta_1 > 0$ such that $\upsilon'(\theta) < 0$ when $\theta \in (0, \theta_1)$, and $\upsilon'(\theta) > 0$, when $\theta \in (\theta_1, +\infty)$. So $\upsilon(\theta)$ is decreasing on $(0, \theta_1)$ and increasing on $(\theta_1, +\infty)$, and it can be further shown that $\upsilon(0) = 0$ and $\upsilon(+\infty) > 0$. Thus, $\upsilon(\theta_1) < 0$. With $\upsilon(\theta)$ being continuous and increasing on $(\theta_1, +\infty)$, so there is one and only one solution for θ to Equation (7). The illustration of the proof is shown in Figure 3.

Using parameter transformation, it is easy to obtain p_0 and p_1 as follows:

$$\begin{cases} p_0 = \dfrac{q_0 + q_1 - (1+\theta)e^{-\theta}}{1 - (1+\theta)e^{-\theta}} \\ p_1 = \dfrac{q_0 - (1 - p_0)e^{-\theta}}{p_0}. \end{cases} \tag{A1}$$

Based on the invariance property for the maximum likelihood estimation and the one-to-one transformation (4,A1), the maximum likelihood estimation of

ZOIP model (2) is sole, when at least one observation is larger than one. □

A.2

Proof of Equations (9), (10), (13), (14) and (15): Using the representation $Y = V(1 - B_1) + B_1(1 - B_2)$, we have

$$\begin{aligned} P(Y=0) &= P(Y=0, B_1=0) + P(Y=0, B_1=1) \\ &= P(V=0, B_1=0) + P(B_2=1, B_1=1) \\ &= p_0 p_1 + (1-p_0)P(V=0). \end{aligned}$$

Then we can calculate the conditional probabilities as follows:

$$\begin{aligned} &P(V=0, B_1=0, B_2=0|Y=0, p_0, p_1, \theta) \\ &= \frac{P(V=0, B_1=0, B_2=0, Y=0)}{P(Y=0)} \\ &= \frac{(1-p_0)(1-p_1)P(V=0)}{p_0p_1 + (1-p_0)P(V=0)}, \end{aligned}$$

$$\begin{aligned} &P(V=0, B_1=0, B_2=1|Y=0, p_0, p_1, \theta) \\ &\quad = \frac{P(V=0, B_1=0, B_2=1, Y=0)}{P(Y=0)} \\ &\quad = \frac{(1-p_0)p_1P(V=0)}{p_0p_1 + (1-p_0)P(V=0)}, \end{aligned}$$

$$\begin{aligned} &P(V=v, B_1=1, B_2=1|Y=0, p_0, p_1, \theta) \\ &\quad = \frac{P(V=v, B_1=1, B_2=1, Y=0)}{P(Y=0)} \\ &\quad = \frac{p_0p_1P(V=v)}{p_0p_1 + (1-p_0)P(V=0)}. \end{aligned}$$

It is easy to see that the sum of these conditional probabilities equals 1. So Equation (13) is proved.

Equations (14) and (15) can be proved similarly. Using Equations (13), (14) and (15), we have

$$\begin{aligned} &P(B_{1i}=1|Y_i=0, p_0^{(k)}, p_1^{(k)}, \theta^{(k)}) \\ &\quad = \sum_{v=0}^{+\infty} P(V=v, B_{1i}=1, B_{2i}=1|Y=0) \\ &\quad = \sum_{v=0}^{+\infty} \frac{p_0^{(k)}p_1^{(k)}P(V=v)}{p_0^{(k)}p_1^{(k)} + (1-p_0^{(k)})P(V=0)} \\ &\quad = \frac{p_0^{(k)}p_1^{(k)}}{p_0^{(k)}p_1^{(k)} + (1-p_0^{(k)})P(V=0)}, \\ &P(B_{1i}=1|Y_i=1, p_0^{(k)}, p_1^{(k)}, \theta^{(k)}) \\ &\quad = \sum_{v=0}^{+\infty} P(V=v, B_{1i}=1, B_{2i}=0|Y=1) \\ &\quad = \sum_{v=0}^{+\infty} \frac{p_0^{(k)}(1-p_1^{(k)})P(V=v)}{p_0^{(k)}(1-p_1^{(k)}) + (1-p_0^{(k)})P(V=1)} \\ &\quad = \frac{p_0^{(k)}(1-p_1^{(k)})}{p_0^{(k)}(1-p_1^{(k)}) + (1-p_0^{(k)})P(V=1)}, \end{aligned}$$

and when $Y_i \geq 2$,

$$P(B_{1i}=1|Y_i=y, p_0^{(k)}, p_1^{(k)}, \theta^{(k)}) = 0, y = 2, 3, \ldots$$

Equation (10) can be proved similarly. □

A.3

Proof of Theorem 2.2: Let $\ell_c(\boldsymbol{\eta}|\boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2}) = \log(L_c(\boldsymbol{\eta}|\boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2}))$. The proof depends on the smoothness of $Q(\boldsymbol{\eta}^*, \boldsymbol{\eta}) = E[\ell_c(\boldsymbol{\eta}^*|\boldsymbol{Y}, \boldsymbol{B_1}, \boldsymbol{B_2})|\boldsymbol{Y}, \boldsymbol{\eta}]$. We rewrite

$$Q(\boldsymbol{\eta}^*, \boldsymbol{\eta}) = Q_1(\boldsymbol{\eta}^*, \boldsymbol{\eta}) + Q_2(\boldsymbol{\eta}^*, \boldsymbol{\eta}) + Q_3(\boldsymbol{\eta}^*, \boldsymbol{\eta}),$$

where

$$\begin{aligned} &Q_1(\boldsymbol{\eta}^*, \boldsymbol{\eta}) \\ &\quad = \sum_{i=1}^{n} E[B_{1i}\ln p_0^* + (1-B_{1i})\ln(1-p_0^*)|\boldsymbol{Y}, \boldsymbol{\eta}] \\ &\qquad - \sum_{i=1}^{n} E[(1-B_{1i})\ln(Y_i!)|\boldsymbol{Y}, \boldsymbol{\eta}], \\ &Q_2(\boldsymbol{\eta}^*, \boldsymbol{\eta}) \\ &\quad = \sum_{i=1}^{n} E[B_{2i}\ln p_1^* + (1-B_{2i})\ln(1-p_1^*)|Y, \boldsymbol{\eta}], \end{aligned}$$

and

$$Q_3(\boldsymbol{\eta}^*, \boldsymbol{\eta}) = \sum_{i=1}^{n} E[(1-B_{1i})(Y_i\ln\theta^* - \theta^*)|\boldsymbol{Y}, \boldsymbol{\eta}].$$

Simple calculation yields

$$\begin{aligned} &Q_1(\boldsymbol{\eta}^*, \boldsymbol{\eta}) \\ &= \sum_{i=1}^{n} I\{Y_i=0\}\left[\frac{p_0p_1\ln p_0^*}{p_0p_1 + (1-p_0)P(V=0)} \right. \\ &\quad \left. + \frac{(1-p_0)P(V=0)\ln(1-p_0^*)}{p_0p_1 + (1-p_0)P(V=0)}\right] \\ &\quad + \sum_{i=1}^{n} I\{Y_i=1\}\left[\frac{p_0(1-p_1)\ln p_0^*}{p_0(1-p_1) + (1-p_0)P(V=1)} \right. \\ &\quad \left. + \frac{(1-p_0)P(V=1)\ln(1-p_0^*)}{p_0(1-p_1) + (1-p_0)P(V=1)}\right] \\ &\quad + \sum_{i=1}^{n} I\{Y_i \geq 2\}\ln(1-p_0^*) \\ &\quad - \sum_{i=1}^{n} I\{Y_i \geq 2\}\ln(Y_i!), \end{aligned}$$

which is continuous in $\boldsymbol{\eta}^*$ and $\boldsymbol{\eta}$. A similar conclusion can be applied to $Q_2(\boldsymbol{\eta}^*, \boldsymbol{\eta})$ and $Q_3(\boldsymbol{\eta}^*, \boldsymbol{\eta})$. It is easy to find that Equation (11) in Wu (1983) for $Q(\boldsymbol{\eta}^{(k)}, \boldsymbol{\eta})$ is correct. So according to Theorem 3 in Wu (1983), the conclusion is obvious. □