Statistics and Applications {ISSN 2454-7395 (online)}
Volume 24, No. 1, 2026 (New Series), pp 127–153
http://www.ssca.org.in/journal

Zero-One-Inflated Poisson-Garima Distribution and its Applications in Biomedical Studies

Divya A.[1,2]**, Prasanth C. B.**[3] **and Muhammed Anvar P.**[4]

[1]*Department of Statistics, Govt. Victoria College, Affiliated to University of Calicut, Palakkad, Kerala-678001*
[2]*Department of Statistics, St.Thomas college, Thrissur, Kerala-680001*
[3]*Department of Statistics, Sree Kerala Varma College, Thrissur, Kerala-680011*
[4]*Department of Statistical Sciences, Kannur University, Kannur, Kerala-670567*

Received: 13 December 2024; Revised: 21 July 2025; Accepted: 23 July 2025

Abstract

Count data appears in diverse field of study with a variety of patterns in certain frequencies such as excessive occurrence of zeros and ones compared to other possible values. Here a new zero-one-inflated Poisson-Garima distribution and its applications is studied. After introducing the modified Poisson-Garima distribution, its various statistical properties and estimation of parameters are discussed. The estimation methods are then illustrated with simulated samples and the proposed model was applied to two real data sets. It is demonstrated that the new model outperformed its competitors like zero-one-inflated Poisson distribution and zero-one-inflated Poisson-Lindley distribution.

Key words: Count data; EM algorithm; Garima distribution; Inflated distributions; Poisson mixture; Zero-one-inflated data.

AMS Subject Classifications: 62E10, 62E15, 62F10

1. Introduction

In any statistical data analysis, the choice of a suitable probability distribution has a significant role. In the case of discrete valued count data, the most adopted model is the standard Poisson distribution. This is due to its greater simplicity and its property of equi-dispersion. But in reality, most of the count data shows overdispersion and long tail behaviour. So, for modeling such overdispersed count data, many researchers have relied upon distributions such a Negative Binomial, Generalized Poisson etc. There are many reasons for over/under dispersion. Existence of excess/less number of zero counts is one of them. In such cases we use zero-inflated/deflated models (zero-modified models). Zero-inflated models are a mixture of two distributions of which one generates only zeros and the other generates non-negative counts from a baseline distribution. Thus, by taking mixtures of a degenerate distribution at zero and a non-negative count distribution such as Poisson,

Corresponding Authors: Divya A., Muhammed Anvar P.
Email: divy.312@gmail.com, anvarp@kannuruniv.ac.in

Negative Binomial, Geometric, Poisson-Lindley etc., we get a zero-inflated model. Many zero-inflated models are developed to address the issue of presence of an excess number of zeros. The zero-inflated Poisson (ZIP) regression model proposed by Lambert (1992) is one among them to handle count data with excess zeros.

In addition to the excessive occurrence of zeros alone, there are cases in which count data contain excess zeros and ones simultaneously. For example, virus infections occur for at most one time as after the first infection, antibodies will be generated and so there is less chance of further infection. In their unpublished manuscript Melkersson and Olsson (1999) expanded upon the modeling of count data characterized by both excess zeros and excess ones. They introduced an extension of the ZIP distribution, creating a novel distribution known as the zero-one-inflated Poisson distribution (ZOIP). Later, Zhang *et al.* (2016) conducted a comprehensive investigation of the ZOIP distribution, initially establishing five equivalent stochastic representations for the ZOIP random variable and then deduced its various essential distributional properties. The statistical inference for ZOIP models was explored by Tang *et al.* (2017). They addressed the inference problem by employing both maximum likelihood estimation and Bayesian estimation techniques. Liu *et al.* (2018) have done the estimation of ZOIP by using expectation-maximization (EM) algorithm to get maximum likelihood estimate (MLE) and Gibbs sampling to get samples from posterior distribution based on latent variables. Also they have compared the MLEs with Bayesian estimates by Monte Carlo simulation. Recently Tajuddin *et al.* (2022) introduced the zero–one-inflated Poisson–Lindley (ZOIPL) distribution as an alternative to ZOIP model. They studied the statistical properties and characteristics of this distribution. Altun *et al.* (2023) proposed a zero-inflated Poisson generalized Lindley regression model as an alternative to the zero-inflated negative-binomial regression model. Recently, more discrete distributions based on mixing of a Poisson rate parameter with new continuous models are emerging in the literature. For example, the Poisson-Garima (PG) distribution introduced by Shanker (2017) is one among them. It has been demonstrated that this model is capable of handling the over dispersion and heavy tailed-ness in a more effective way than that of previously discussed models.This model was extended to the case of modeling count data without zeros by Shanker and Shukla (2017), resulting in the zero-truncated PG distribution.

It is observed that in some situations, count data contains large values with non-negligible frequency, in addition to the presence of excessive zeros and/or ones. For example, in the monthly number of drug offenses recorded from January 1990 to December 2001, in Pittsburgh census tract 2206 (Garay *et al.* (2022)) which we analyze in a later section, the observed maximum count is 29 and the data is right skewed. The basic Poisson model, however, cannot effectively capture this pattern as for having higher frequency for large values, the rate parameter θ should be large. But when θ becomes large, it tends to a symmetric behaviour around θ in contrast to right tailed nature of the data, as we can see in the frequency pattern given in Figure 1a (See Appendix I). In Figure 1b (See Appendix I) the tail probabilities are plotted for both Poisson and PG distributions with different values of θ. For small values of θ, the tail probabilities of PG are higher than Poisson. Although the tail probabilities of the Poisson distribution become slightly larger for higher values of θ, the distribution tends to lose its right-skewness, which may limit its suitability for modeling right skewed data. This problem persists even when we take mixtures of zero/one inflated Poisson model. To model this pattern in a better way it would be beneficial to use PG distribution with mixtures of zeros and ones. In view of this, we aim to study the properties of multiple values inflated PG model with special focus on zero inflation and simultaneous inflation of both zero and one. In this paper, a new zero-one-inflated PG distribution is introduced and studied its properties.

The paper is arranged as follows. Section 2 revisits the details of PG distribution and we introduce a modified version of PG distribution that is, zero-one-inflated PG (ZOIPG) distribution. Also its statistical properties and different methods of estimation are discussed. In the same section, a special case of ZOIPG is introduced, i.e., zero-inflated PG (ZIPG) distribution and its statistical properties and estimation details are discussed. Section 3 describes the details of simulation study. In section 4, we demonstrate the applications of the proposed models with three real data sets. Some concluding remarks are given in the last section.

2. Inflated Poisson-Garima distribution

2.1. Poisson-Garima distribution

The PG distribution was introduced by Shanker (2017) by compounding Poisson distribution with Garima distribution which was introduced by Shanker (2016). That is, when the parameter of the Poisson distribution follows a Garima distribution, we obtain PG distribution. The probability mass function (PMF) of PG distribution with parameter θ is

$$f(x) = \frac{\theta}{\theta+2}\frac{\theta x+(\theta^2+3\theta+1)}{(\theta+1)^{x+2}}; x=0,1,2,\cdots;\;\; \theta>0. \tag{1}$$

The first four moments about origin of PG distribution are given by

$$\mu_1' = \frac{\theta+3}{\theta(\theta+2)},$$

$$\mu_2' = \frac{\theta^2+5\theta+8}{\theta^2(\theta+2)},$$

$$\mu_3' = \frac{\theta^3+9\theta^2+30\theta+30}{\theta^3(\theta+2)},$$

$$\mu_4' = \frac{\theta^4+17\theta^3+92\theta^2+204\theta+144}{\theta^4(\theta+2)}.$$

Shanker (2017) examined the properties of PG distribution, including its shape, moments, skewness, and kurtosis. Additionally, PG distribution exhibits an increasing hazard rate, making it a unimodal distribution. In the next section, we study ZOIPG distribution.

2.2. Zero-one-inflated Poisson-Garima distribution

Let V be a discrete random variable with non-negative integers as support having a PMF $g(.)$. A modification of the PMF g is required when excess frequencies at certain values of V are observed relative to the base distribution g. For example, a situation may arise when the expected number of occurrences of zeros and ones are larger than the baseline distribution. A random variable X is said to follow a ZOIPG distribution with parameters p_1, p_2 and θ, if it admits the following stochastic representation.

$$X = V(1 - B_1) + B_1(1 - B_2), \tag{2}$$

where B_1 is a Bernoulli random variable with success probability p_1, $0 < p_1 < 1$, B_2 is another Bernoulli random variable with success probability p_2, $0 < p_2 < 1$, and V follows a PG distribution with rate parameter θ. Also B_1, B_2 and V are assumed to be mutually independent.
The relation between X and (B_1, B_2, V) can be understood by noticing the following expression,

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

Then the PMF of a ZOIPG random variable with parameters p_1, p_2 and θ, (ZOIPG(p_1, p_2, θ))is given by

$$P[X = k] = f(k) = \begin{cases} p_1p_2 + (1 - p_1)\frac{\theta}{\theta+2}\frac{(\theta^2+3\theta+1)}{(\theta+1)^2} & \text{for } k = 0 \\ p_1(1 - p_2) + (1 - p_1)\frac{\theta}{\theta+2}\frac{(\theta^2+4\theta+1)}{(\theta+1)^3} & \text{for } k = 1 \\ (1 - p_1)\frac{\theta}{\theta+2}\frac{(\theta^2+3\theta+1+\theta k)}{(\theta+1)^{k+2}} & \text{for } k \geq 2. \end{cases} \tag{4}$$

Remark: When $p_1 = 0$, this model reduces to PG distribution and when $p_2 = 1$, $0 < p_1 < 1$, this model becomes ZIPG model.
Next, we discuss the properties of ZOIPG model.

Distribution function

Let X follows a ZOIPG(p_1, p_2, θ) distribution. Then its distribution function $F(x)$ is obtained as

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ p_1p_2 + K_0 & \text{for } 0 \leq x < 1 \\ p_1 + \sum_{r=0}^{\lfloor x \rfloor} K_r & \text{for } x \geq 1 \end{cases} \tag{5}$$

where $K_r = (1 - p_1)\frac{\theta}{\theta+2}\frac{(\theta^2+3\theta+1+\theta r)}{(\theta+1)^{r+2}}$, $r = 0, 1, 2, \ldots$ and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

2.2.1. Moments and other related measures

The r^{th} factorial moment of a ZOIPG(p_1, p_2, θ) distribution is obtained as follows

$$\begin{aligned} \mu'_{(r)} &= E[X(X - 1)(X - 2) \ldots (X - r + 1)] \\ &= (1 - p_1)\frac{r!(\theta + r + 2)}{\theta^r(\theta + 2)}, \quad r \geq 2. \end{aligned}$$

For $r = 1$, we have $\mu'_{(1)}$ directly from the definition and we obtain it as

$$\mu'_{(1)} = p_1(1-p_2) + (1-p_1)\frac{\theta+3}{\theta(\theta+2)}. \tag{6}$$

For $r = 2$, we get

$$\mu'_{(2)} = (1-p_1)\frac{2(\theta+4)}{\theta^2(\theta+2)}.$$

By using the relationship between factorial moments and moments about the origin we can get the first four moments about the origin as follows

$$\mu'_1 = p_1(1-p_2) + (1-p_1)\frac{\theta+3}{\theta(\theta+2)}, \tag{7}$$

$$\mu'_2 = p_1(1-p_2) + (1-p_1)\frac{\theta^2+5\theta+8}{\theta^2(\theta+2)}, \tag{8}$$

$$\mu'_3 = 5p_1(1-p_2) + (1-p_1)\frac{5\theta^3+21\theta^2+30\theta+30}{\theta^3(\theta+2)}, \tag{9}$$

$$\mu'_4 = 25p_1(1-p_2) + (1-p_1)\frac{25\theta^4+89\theta^3+92\theta^2+204\theta+144}{\theta^4(\theta+2)}. \tag{10}$$

2.2.2. Moment generating function

The Moment generating function (MGF) of ZOIPG(p_1, p_2, θ) can be obtained as

$$\begin{aligned} M_X(t) &= E(e^{tx}) \\ &= p_1p_2 + p_1(1-p_2)e^t + (1-p_1)M_Y(t) \\ &= p_1p_2 + p_1(1-p_2)e^t + (1-p_1)\frac{\theta^3+(4-e^t)\theta^2+2(2-e^t)\theta+1-e^t}{(\theta+1)(\theta+2)(\theta+1-e^t)^2}, \end{aligned}$$

where $M_Y(t)$is the MGF of PG distribution.

2.2.3. Probability generating function

The probability generating function (PGF) of ZOIPG(p_1, p_2, θ) is obtained as

$$\begin{aligned} P_X(t) &= E(t^x) \\ &= p_1p_2 + p_1(1-p_2)t + (1-p_1)\frac{\theta^3+(4-t)\theta^2+2(2-t)\theta+1-t}{(\theta+1)(\theta+2)(\theta+1-t)^2}. \end{aligned}$$

2.2.4. Estimation of parameters

In this section, we discuss the estimation methods for the ZOIPG distribution. The parameter vector to be estimated is denoted as $\Lambda = (p_1, p_2, \theta)'$. First we give details of method of moment (MOM) estimation and then discuss the ML estimation method. The mixture structure of ZOIPG makes it easy to handle the ML estimation method using the EM algorithm.

Method of moments estimation

Let $x_1, x_2, ...x_n$ be a observed sample of size n from ZOIPG(p_1, p_2, θ). Define

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i(x_i - 1)$$

and

$$m_3 = \frac{1}{n}\sum_{i=1}^{n} x_i(x_i - 1)(x_i - 2).$$

The corresponding population moments are given by

$$E[X] = p_1(1 - p_2) + (1 - p_1)\frac{(\theta + 3)}{\theta(\theta + 2)},$$

$$E[X(X - 1)] = (1 - p_1)\frac{2(\theta + 4)}{\theta^2(\theta + 2)},$$

$$E[X(X - 1)(X - 2)] = (1 - p_1)\frac{6(\theta + 5)}{\theta^3(\theta + 2)}.$$

Equating the sample moments with the corresponding population moments and solving the equations, we get the MOMs of p_1, p_2 and θ as follows

$$\tilde{\theta} = \frac{(3m_2 - 4m_3) + \sqrt{(3m_2 + 4m_3)^2 + 12m_2m_3}}{2m_3},$$

$$\tilde{p_1} = \frac{2\tilde{\theta} + 8 - m_2\tilde{\theta}^3 - 2m_2\tilde{\theta}^2}{2(\tilde{\theta} + 4)},$$

$$\tilde{p_2} = \frac{\tilde{p_1} + (1 - \tilde{p_1})\frac{\tilde{\theta}+3}{\tilde{\theta}(\tilde{\theta}+2)} - m_1}{\tilde{p_1}}.$$

The MOM $\tilde{\theta}$ needs special attention while conducting numerical computations. We need to ensure the positivity of resulting value of the $\tilde{\theta}$ for a given sample. This is ensured by taking the positive square root of $(4m_3 + 3m_2)^2 + 12m_2m_3$. This is important because other estimators depend on this value of $\tilde{\theta}$. The details of the behaviour of $\tilde{\theta}$ is thus discussed in simulation study section. Next we describe the ML procedure.

Maximum likelihood estimation

Let $x = (x_1, x_2, ...x_n)$ be an observed sample of size n from ZOIPG(p_1, p_2, θ). Then the likelihood function is given by

$$\begin{aligned} L(p_1, p_2, \theta) = &\left[p_1 p_2 + (1-p_1)\frac{\theta(\theta^2+3\theta+1)}{(\theta+2)(\theta+1)^2}\right]^{S_0} \\ &\times\left[p_1(1-p_2) + (1-p_1\frac{\theta(\theta^2+4\theta+1)}{(\theta+2)\theta+1^3})\right]^{S_1} \\ &\times\left[\frac{(1-p_1)\theta}{(\theta+2)(\theta+1)^2}\right]^{n-S_0-S_1}\prod_{x_i\geq 2}\frac{\theta^2+3\theta+1+\theta x_i}{(\theta+1)^{x_i}} \\ = &{q_0}^{S_0}{q_1}^{S_1}\left[\frac{(1-p_0)\theta}{(\theta+2)(\theta+1)^2}\right]^{n-S_0-S_1}\prod_{x_i\geq 2}\frac{\theta^2+3\theta+1+\theta x_i}{(\theta+1)^{x_i}}, \end{aligned}$$

where S_0 and S_1 are the number of zeros and ones in the data respectively and
$q_0 = P[X = 0]$ and $q_1 = P[X = 1]$.
Then the likelihood function becomes

$$L(q_0, q_1, \theta|x) = {q_0}^{S_0}{q_1}^{S_1}\left[\frac{(1-q_0-q_1)\theta(\theta+1)}{\theta^2+5\theta+2}\right]^{n-S_0-S_1}\prod_{x_i\geq 2}\frac{\theta^2+3\theta+1+\theta x_i}{(\theta+1)^{x_i}}.$$

Now the log likelihood function can be written as

$$\begin{aligned} \log L = &(n-S_0-S_1)[\log(1-q_0-q_1) + \log\theta + \log(\theta+1) - \log(\theta^2+5\theta+2)] \\ &+ S_0\log q_0 + S_1\log q_1 + \sum_{x_i\geq 2}[\log(\theta^2+3\theta+1+\theta x_i) - x_i\log(\theta+1)]. \end{aligned}$$

Solving the likelihood equations $\frac{\partial \log L}{\partial q_0} = 0$ and $\frac{\partial \log L}{\partial q_1} = 0$ we get

$$\hat{q_0} = \frac{S_0}{n}$$

and

$$\hat{q_1} = \frac{S_1}{n}.$$

Then the MLE of θ, $\hat{\theta}$, is the solution of the non-linear equation

$$(n-S_0-S_1)\left[\frac{4\theta^2+4\theta+2}{\theta(\theta+1)(\theta^2+5\theta+2)}\right] + \sum_{x_i\geq 2}\left[\frac{2\theta+3+x_i}{\theta^2+3\theta+1+\theta x_i} - \frac{x_i}{\theta+1}\right] = 0,$$

which can be solved numerically by Newton-Raphson type iterative algorithm. Also note that the transformation between $(p_1, p_2, \theta)'$ and $(q_1, q_2, \theta)'$ is one to one. Thus the MLE of p_1 and p_2 are obtained as follows

$$\hat{p_1} = \frac{S_0+S_1}{n} - \frac{\hat{\theta}(\hat{\theta}^3+5\hat{\theta}^2+8\hat{\theta}+2)}{(\hat{\theta}+2)(\hat{\theta}+1)^3} \quad \text{and}$$

$$\hat{p_2} = \frac{\frac{S_0}{n} - (1-\hat{p_1})\frac{\hat{\theta}(\hat{\theta}^2+3\hat{\theta}+1)}{(\hat{\theta}+2)(\hat{\theta}+1)^2}}{\hat{p_1}}.$$

The finite sample properties of the above MLEs are studied using simulated data sets and are reported in Section 3. The special structure of the model in the form of mixture distribution makes it easier to do ML estimation by applying the EM algorithm. Next we describe the details of ML estimation using EM algorithm.

Maximum likelihood estimation with EM algorithm

For the iterative computation of MLE, we can use the EM algorithm. A ZOIPG random variable can be represented by three independent latent variables i.e., two Bernoulli variables B_1, B_2 and a PG random variable V. If we can observe these latent variables then the likelihood function of ZOIPG will be a product of three likelihood functions. Let $X = (X_1, X_2, ..., X_n)$ be a random sample from ZOIPG with corresponding latent variables $B_1 = (B_{11}, B_{12}, ..., B_{1n})$, $B_2 = (B_{21}, B_{22}, ..., B_{2n})$ and $V = (V_1, V_2, ..., V_n)$. Also we have

$$X_i = V_i(1-B_{1i}) + B_{1i}(1-B_{2i}) \tag{11}$$

So if we could observe B_{1i} and B_{2i} and hence V_i then by relation (11), the augmented likelihood function with augmented data (B_1, B_2, V) would be

$$\begin{aligned} L_c(p_1, p_2, \theta|X, B_1, B_2) &= \prod_{i=1}^{n} p_1^{B_{1i}}(1-p_1)^{1-B_{1i}} \prod_{i=1}^{n} p_2^{B_{2i}}(1-p_2)^{1-B_{2i}} \\ &\quad \prod_{i=1}^{n}\Big\{\frac{\theta(\theta X_i+\theta^2+3\theta+1)}{(\theta+2)(\theta+1)^{X_i+2}}\Big\}^{1-B_{1i}} \\ &\triangleq L_c(p_1|X, B_1, B_2)L_c(p_2|X, B_1, B_2)L_c(\theta|X, B_1, B_2) \end{aligned} \tag{12}$$

The maximisation of $L_c(p_1, p_2, \theta|X, B_1, B_2)$ is carried out by maximizing $L_c(p_1|X, B_1, B_2)$, $L_c(p_2|X, B_1, B_2)$ and $L_c(\theta|X, B_1, B_2)$ separately. Thus the MLEs of p_1, p_2 and θ with EM algorithm can be found iteratively using two steps i.e. E step and M step respectively. In the E step, the expected values of B_{1i} and B_{2i} under the current estimates of (p_1, p_2, θ) are found. In the M step, maximization of $L_c(p_1, p_2, \theta|X, B_1, B_2)$ with the expected values of B_{1i} and B_{2i} from the E step, is done. The iteration is continued until the estimated values of p_1, p_2 and θ converge. The final iteration gives the MLEs of p_1, p_2 and θ, i.e. $\hat{p_1}_{EM}, \hat{p_2}_{EM}$ and $\hat{\theta_{EM}}$. The $(k+1)^{th}$ iteration of the EM algorithm is described as follows.

E step : Let $(p_1^{(k)}, p_2^{(k)}, \theta^{(k)})$ be the estimates obtained in the k^{th} step. Then

$$B_{1i}^{(k+1)} = E[B_{1i}|X_i, p_1^{(k)}, p_2^{(k)}, \theta^{(k)}] = \begin{cases} \frac{p_1^{(k)}p_2^{(k)}}{p_1^{(k)}p_2^{(k)}+(1-p_1^{(k)})\frac{\theta^{(k)}}{\theta^{(k)}+2}\frac{(\theta^{(k)})^2+3\theta^{(k)}+1)}{(\theta^{(k)}+1)^2}} & ; X_i = 0 \\ \frac{p_1^{(k)}(1-p_2^{(k)})}{p_1^{(k)}(1-p_2^{(k)})+(1-p_1^{(k)})\frac{\theta^{(k)}}{\theta^{(k)}+2}\frac{(\theta^{(k)})^2+4\theta^{(k)}+1)}{(\theta^{(k)}+1)^3}} & ; X_i = 1 \\ 0 & ; X_i = 2, 3, ... \end{cases}$$

$$B_{2i}^{(k+1)} = E[B_{2i}|X_i, p_1^{(k)}, p_2^{(k)}, \theta^{(k)}] = \begin{cases} \frac{p_1^{(k)}p_2^{(k)}+(1-p_1^{(k)})p_2^{(k)}\frac{\theta^{(k)}}{\theta^{(k)}+2}\frac{({\theta^{(k)}}^2+3\theta^{(k)}+1)}{(\theta^{(k)}+1)^2}}{p_1^{(k)}p_2^{(k)}+(1-p_1^{(k)})\frac{\theta^{(k)}}{\theta^{(k)}+2}\frac{({\theta^{(k)}}^2+3\theta^{(k)}+1)}{(\theta^{(k)}+1)^2}} & ; X_i = 0 \\ \frac{p_2^{(k)}(1-p_1^{(k)})\frac{\theta^{(k)}}{\theta^{(k)}+2}\frac{({\theta^{(k)}}^2+4\theta^{(k)}+1)}{(\theta^{(k)}+1)^3}}{p_1^{(k)}(1-p_2^{(k)})+(1-p_1^{(k)})\frac{\theta^{(k)}}{\theta^{(k)}+2}\frac{({\theta^{(k)}}^2+4\theta^{(k)}+1)}{(\theta^{(k)}+1)^3}} & ; X_i = 1 \\ p_1^{(k)} & ; X_i = 2, 3, ... \end{cases}$$

M step:
Let

$$B_1^{(k+1)} = (B_{11}^{(k+1)}, B_{12}^{(k+1)}, ..., B_{1n}^{(k+1)})$$

and

$$B_2^{(k+1)} = (B_{21}^{(k+1)}, B_{22}^{(k+1)}, ..., B_{2n}^{(k+1)}).$$

Maximising $L_c(p_1|X, B_1^{(k+1)}, B_2^{(k+1)})$ and $L_c(p_2|X, B_1^{(k+1)}, B_2^{(k+1)})$ we get the MLE of p_1 and p_2 respectively as follows.

$$p_1^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} B_{1i}^{(k+1)}$$

and

$$p_2^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} B_{2i}^{(k+1)}.$$

$\theta^{(k+1)}$ is obtained by maximising $L_c(\theta|X, B_1^{(k+1)}, B_2^{(k+1)})$ and it is the solution of the equation

$$\sum_{i=1}^{n}(1 - B_{1i})\left\{\frac{1}{\theta} + \frac{X_i + 2\theta + 3}{\theta X_i + \theta^2 + 3\theta + 1} - \frac{1}{\theta + 2} - \frac{X_i + 2}{\theta + 1}\right\} = 0,$$

which can be solved numerically. Under the regularity conditions, it can be shown that the estimators are consistent and asymptotically normally distributed (See McLachlan and Krishnan (2008)).

2.3. Zero-inflated Poisson-Garima distribution

A special case of the ZOIPG distribution is the situation when $p_2 = 1$ and $0 < p_1 < 1$, that is, ZIPG. A random variable Y is said to have a ZIPG distribution with parameters p_1 and θ, (ZIPG(p_1, θ)) if the PMF of Y is specified as follows.

$$P[Y = k] = \begin{cases} p_1 + (1 - p_1)\frac{\theta}{\theta+2}\frac{(\theta^2+3\theta+1)}{(\theta+1)^2} & \text{for } k = 0 \\ (1 - p_1)\frac{\theta}{\theta+2}\frac{(\theta^2+3\theta+1+\theta k)}{(\theta+1)^{k+2}} & \text{for } k \geq 1. \end{cases} \quad (13)$$

2.3.1. Moments and other related measures

The r^{th} factorial moment of a random variable Y following ZIPG(p_1, θ) is obtained as follows

$$\begin{aligned}\mu'_{(r)} &= E[Y(Y-1)(Y-2)\dots(Y-r+1)]\\ &= (1-p_1)\frac{r!(\theta+r+2)}{\theta^r(\theta+2)}, \quad r \geq 1\end{aligned}$$

For $r = 1$, we get $\mu'_{(1)}$ as

$$\mu'_{(1)} = (1-p_1)\frac{\theta+3}{\theta(\theta+2)}.$$

For $r = 2$, we get

$$\mu'_{(2)} = (1-p_1)\frac{2(\theta+4)}{\theta^2(\theta+2)}.$$

By using the relationship between factorial moments and moments about the origin we can get the first four moments about the origin as

$$\mu'_1 = (1-p_1)\frac{\theta+3}{\theta(\theta+2)}, \qquad \mu'_2 = (1-p_1)\frac{\theta^2+5\theta+8}{\theta^2(\theta+2)},$$

$$\mu'_3 = (1-p_1)\frac{\theta^3+9\theta^2+30\theta+30}{\theta^3(\theta+2)}, \qquad \mu'_4 = (1-p_1)\frac{\theta^4+17\theta^3+92\theta^2+204\theta+144}{\theta^4(\theta+2)}.$$

Thus the mean and variance of ZIPG distribution can be obtained as

$$\mu_Y = (1-p_1)\frac{\theta+3}{\theta(\theta+2)}, \quad \sigma^2_Y = (1-p_1)\frac{p_1\theta^2+6p_1\theta+9p_1+\theta^3+6\theta^2+12\theta+7}{\theta^2(\theta+2)^2}.$$

2.3.2. Moment generating function

Let Y follows ZIPG(p_1, θ).Then its MGF can be obtained as

$$\begin{aligned}M_Y(t) &= E(e^{ty})\\ &= p_1 + (1-p_1)\frac{\theta(\theta^2+3\theta+1)}{(\theta+2)(\theta+1)^2} + \sum_{y=1}^{\infty} e^{ty}(1-p_1)\frac{\theta(\theta y+\theta^2+3\theta+1)}{(\theta+2)(\theta+1)^{y+2}}\\ &= p_1 + (1-p_1)M_X(t)\\ &= p_1 + (1-p_1)\frac{\theta\{\theta^3+(4-e^t)\theta^2+2(2-e^t)\theta+1-e^t\}}{(\theta+1)(\theta+2)(\theta+1-e^t)^2},\end{aligned}$$

where $M_X(t)$is the MGF of PG distribution with parameter θ.

2.3.3. Probability generating function

The PGF of a random variable Y following ZIPG(p_1, θ) is obtained as

$$\begin{aligned}P_Y(t) &= E(t^y)\\ &= p_1 + (1-p_1)\frac{\theta\{\theta^3+(4-t)\theta^2+2(2-t)\theta+1-t\}}{(\theta+1)(\theta+2)(\theta+1-t)^2}.\end{aligned}$$

2.3.4. Estimation of parameters

Like in the case of ZOIPG, we describe the different methods of estimation used for ZIPG as follows. In this case the parameter vector to be estimated is given by $\Lambda = (p_1, \theta)'$.

Method of moments estimation

Let $y_1, y_2, \ldots, y_n$ be the observed sample of size n from ZIPG distribution with probability mass as given in equation (13)
Let

$$m_1 = \bar{y}, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} y_i(y_i - 1),$$

where $\bar{y}$ is the sample mean. The corresponding population moments are given by

$$E[Y] = (1-p_1)\frac{(\theta+3)}{\theta(\theta+2)}, \quad E[Y(Y-1)] = (1-p_1)\frac{2(\theta+4)}{\theta^2(\theta+2)}.$$

Equating the sample moments with the corresponding population moments and solving the resulting equations, we get the MOMs of p_1 and θ as follows

$$\tilde{\theta} = \frac{(2m_1 - 3m_2) + \sqrt{(2m_1+3m_2)^2 + 8m_1m_2}}{2m_2}, \quad \tilde{p_1} = \frac{\tilde{\theta} + 3 - m_1\tilde{\theta}(\tilde{\theta}+2)}{(\tilde{\theta}+3)}.$$

As in the case of ZOIPG, the value of the MOM $\tilde{\theta}$ for a given sample should satisfy the positivity condition. To ensure the positivity of $\tilde{\theta}$, we take positive square root of $(2m_1 + 3m_2)^2 + 8m_1m_2$.

Maximum likelihood estimation

Let $y = (y_1, y_2, \ldots, y_n)$ be the observed sample from ZIPG(p_1, θ). Then the likelihood function is given by

$$\begin{aligned} L(p_1, \theta | y) &= \left[p_1 + (1-p_1)\frac{\theta(\theta^2+3\theta+1)}{(\theta+2)(\theta+1)^2}\right]^{S_0} \\ &\quad \times \left[\frac{(1-p_1)\theta}{(\theta+2)(\theta+1)^2}\right]^{n-S_0} \prod_{y_i \geq 1} \frac{\theta^2+3\theta+1+\theta y_i}{(\theta+1)^{y_i}} \\ &= {q_0}^{S_0}\left[\frac{(1-p_0)\theta}{(\theta+2)(\theta+1)^2}\right]^{n-S_0} \prod_{y_i \geq 2} \frac{\theta^2+3\theta+1+\theta y_i}{(\theta+1)^{y_i}}, \end{aligned}$$

where S_0 is the number of zeros in the data and $q_0 = P[Y = 0]$.
Then the likelihood function becomes

$$L(q_0, \theta | y) = {q_0}^{S_0}\left[\frac{(1-q_0)\theta}{7\theta^2+6\theta+2}\right]^{n-S_0} \prod_{y_i \geq 1} \frac{\theta^2+3\theta+1+\theta y_i}{(\theta+1)^{y_i}}.$$

The log-likelihood is given by

$$\log L = S_0 \log q_0 + (n - S_0)\Big[\log(1-q_0) + \log\theta - \log(7\theta^2+6\theta+2)\Big] + \sum_{y_i \geq 1}\Big[\log(\theta^2+3\theta+1+\theta y_i) - y_i\log(\theta+1)\Big].$$

Solving the likelihood equation $\frac{\partial \log L}{\partial q_0} = 0$, we get $\hat{q_0} = \frac{S_0}{n}$.
The MLE of θ, $\hat{\theta}$, is the solution of the non-linear equation

$$(n-S_0)\left[\frac{2-7\theta^2}{\theta(\theta^2+6\theta+2)}\right] + \sum_{y_i \geq 1}\left[\frac{2\theta+3+y_i}{\theta^2+3\theta+1+\theta y_i} - \frac{y_i}{\theta+1}\right] = 0,$$

which can be solved numerically by Newton-Raphson iterative algorithm. Note that the transformation between (p_1, θ) and (q_1, θ) is one to one, thus the MLE of p_1, $\hat{p_1}$, is obtained as follows

$$\hat{p_1} = 1 - \frac{(1-\frac{S_0}{n})(\hat{\theta}+1)^2(\hat{\theta}+2)}{(7\hat{\theta}^2+6\hat{\theta}+2)}.$$

Remark: The MLE can again be computed using the help of EM algorithm, we skip the details for the sake of space considerations.

3. Simulation study

To assess the finite sample performance of the above mentioned estimators, we conducted a simulation study. First, we consider the zero-one-inflated situation. Since the ZOIPG distribution can be expressed as mixture, the simulation from the model is easy. We consider different parameter combinations and simulate observations of varying sizes like $n = 100, 200, 1000$ from that particular ZOIPG. Then we apply the three estimation methods to this artificial data set. The resulting point estimates are then saved and this procedure is repeated for a large number, say, $N = 1000$. After that, we take the arithmetic mean of these 1000 saved estimates as the final estimate in each of the estimation methods. Root mean squared errors (RMSE) are then evaluated to assess the variability of the estimates. Let $\tilde{\Lambda} = (\tilde{p_1}, \tilde{p_2}, \tilde{\theta})'$, $\hat{\Lambda} = (\hat{p_1}, \hat{p_2}, \hat{\theta})'$ and $\hat{\Lambda}_{EM} = (\hat{p}_{1EM}, \hat{p}_{2EM}, \hat{\theta}_{EM})'$ denote respectively the MOM, MLE and EM estimators of Λ. Since the moment estimators have closed form expressions, first we computed the moment estimates. These estimates are then used as starting values to solve the non-linear system of likelihood equations in both the case of MLE and MLE using EM algorithm. The non-linear system of equations derived from equation (12) is solved numerically using iterative Newton-Raphson method implemented in **R** function **optim**. For EM method, we stop the iterative algorithm when difference of consecutive values of log-likelihood is less than a threshold, i.e., $|\log L(\hat{\theta}_k) - \log L(\hat{\theta}_{k-1})| < 0.001$. We also examined the speed of convergence of estimates obtained from EM algorithm for all parameter combinations and for different sample sizes. As an illustration, we give details of convergence for the parameter combination $p_1 = 0.8,\ p_2 = 0.7,\ \theta = 0.4$, exhibited in Figure (2)(See Appendix I). From the figure, it is clear that for large sample size, the convergence speed is high compared to the combinations with small sample size. In both cases, estimator of p_2 converges rapidly. When n is small, estimates of p_1 and θ converges after 10 iterations where as it takes just 4 iterations when n increases to 500.

The simulation results are given in Tables 1,2,3 and 4 (see Appendix I). In Tables 1 and 2, the mean estimates of p_1, p_2 and θ are given when $\theta = 0.4$ and $\theta = 0.8$ respectively. We have chosen different combinations of values of (p_1, p_2), which may represent some practical situations. For instance, we cover the values of parameters including those estimated from fitting real data sets (discussed in the next section). For large values of θ, say, $\theta \geq 4$, we get only lesser number of distinct possible values from the range of a ZOIPG distribution $\{0, 1, 2, 3, \ldots\}$. Hence, we decided to focus on lower values of θ. Other combinations were also studied, but the results are not reported here due to the space restrictions. The complete simulation results are available up on request from the corresponding author.The simulation study uses sample sizes 100, 200 and 1000 which reflects typical data volumes covering small to moderately large samples encountered in practice. These sizes also match with those of the datasets analyzed in the data analysis section, ensuring relevant and realistic evaluation. As seen from the Tables 1 and 2, the bias of the estimates in all three methods decreases when sample size n increases. Tables 3 and 4 exhibit the RMSE of the three estimators when $\theta = 0.4$ and $\theta = 0.8$ respectively. Again from these Tables, we can see that the RMSE is decreasing when the sample size is increasing. For instance, when $\theta = 0.4, p_1 = p_2 = 0.7$, substantial reduction in RMSE is observed when there is an increase of sample size from $n = 100$ to $n = 1000$. This pattern is seen in all other parameter combinations and across the estimation methods. Considering the performance measures of ML and EM algorithm based estimation, almost similar results are obtained. But, as expected, the RMSE of moment estimates are not comparable with that of MLE and EM estimates.

Next we describe the simulation study of ZIPG model. Here, the parameter vector is $\Lambda = (p_1, \theta)'$. Similar to the case of ZOIPG, we repeated the simulation experiment $N = 1000$ times and stored the resulting estimates. The results are exhibited in Table 5 (See Appendix I). In Table 5, we report the point estimates along with their RMSE, when $\theta = 0.2$ and $\theta = 0.8$ respectively. Here also, the performance of MLEs in terms of RMSE is satisfactory. All the computations in the simulation are done through R Core Team (2022) software. The R code of the computations are available in Appendix II. As a final remark on the simulation study, we calculated computational times for different estimation methods. The average computational times are calculated separately for each estimation method across all parameter combinations over the entire $N = 1000$ repetitions. The results are given in Table 6 (See Appendix I). The MOM and MLEs take a little time to computations compared to EM. This is expected since explicit expressions for MOM are available. EM is slow, but better RMSE is obtained. Next, we illustrate the applicability of our proposed models with three real data sets.

4. Data analysis

In order to demonstrate the performance of ZOIPG and ZIPG, these distributions are fitted to three real data sets. Comparison is done with other distributions such as ZIP, ZIPL, ZOIP and ZOIPL. Parameters are computed by using ML estimation technique. For each of these distributions, values of χ^2 statistic, p-value, AIC, BIC, DI(Dispersion Index) values are calculated to compare how well the distribution fits the data.

Data set I

Zeger (1988) analyzed a data set which consists of monthly information about Polio cases reported by U.S. Centre for Disease Control over the years 1970 to 1983. This data was reanalyzed by many authors, among them, see Qi *et al.* (2019). In this data set, there is a total of 168 observations of which 37.87 % are zeroes and 32.74 % are ones. The data have a mean of 1.33 and a variance of 3.35, which indicates overdispersion. Also it is right skewed. Figure 3a (See Appendix I) gives the barplot of the data. Since the data consists of excess number of zeroes and ones, we fitted ZOIP, ZOIPL and ZOIPG models. The results obtained are summarized in Table 7 (See Appendix I). The AIC and BIC values of ZOIPG are comparable with that of ZOIPL, with a slight improvement. Note that the value of χ^2 is minimum for ZOIPG. It is also observed that the ZOIPG model-implied DI is close to the empirical DI. The fitted probabilities for all models against empirical proportions are plotted in Figure 4 (See Appendix I). It is clear that on an average, the ZOIPG models fits well.

Data set II

A data set consisting of weekly number of syphilis cases in the United States from 2007 to 2010 is obtained from **R** package **ZIM** by Yang *et al.* (2018). The data consists of 209 observations of which 77.03 % are zeroes. Figure 3b(See Appendix I) gives the barplot of the data. Since this data consists of excessive zeros, we fitted ZIP, ZIPG and ZIPL for the data. This data have mean 0.9760 and variance 6.6773, implying a clear presence of overdispersion. The results obtained are summarized in Table 8 (See Appendix I). The values of AIC, BIC and χ^2 values are minimum for ZIPG model. Comparing to other models, the DI implied by ZIPG model is closer to empirical DI, indicating the suitability of ZIPG model to this data set. The fitted probabilities for all models against empirical probabilities are plotted in Figure 5(See Appendix I). It is obvious that on an average, the ZIPG models fits well the data.

Data set III

The monthly number of drug offenses recorded from January 1990 to December 2001, in Pittsburgh census tract 2206 (Garay *et al.* (2022)) is analyzed. The data consists of 144 observations and it includes 43.06 % zeroes and 14.58 % ones which can be seen from Figure 3c (See Appendix I). The data have mean 2.11 and variance 12.9106, leading to confirming presence of overdispersion. So we fitted ZOIP, ZOIPG and ZOIPL models for the data. The results are obtained in Table 9 (See Appendix I). A substantial reduction in AIC, BIC and χ^2 values of ZOIPG model are observed in comparison to ZOIP and ZOIPL models. Although none of the implied dispersion indices of all three models is close to empirical DI, the one implied by ZOIPG is still better among them. The fitted probabilities for all models against empirical probabilities are plotted in Figure 6(See Appendix I).

It is observed that, for all the data sets, the ZOIPG/ZIPG models consistently outperform the competing models. Specifically, these models yield the lowest values for the AIC, BIC, and χ^2 criteria. Additionally, the p-values from the goodness-of-fit tests are consistently higher for the proposed models, indicating better fit. Furthermore, the DI of the ZOIPG/ZIPG models is closer to the empirical DI across all data sets.

5. Conclusion

When count data exhibit presence of excessive number of zeros and ones, a mixture of degenerate random variables at 0 and 1 and a Poisson distribution is usually employed. However, in some situations, in addition to the presence of extra zeros and ones, large values are also observed. In this case the zero-one-inflated Poisson model may not be well fitted. Several other Poisson mixture distributions are introduced in literature. In this paper, we have introduced a new inflated Poisson-Garima distribution and its properties. In particular, ZOIPG and ZIPG distributions are considered. The estimation of model parameters using MOM, ML method and ML using EM algorithm are then discussed. A simulation study is conducted to evaluate the finite sample performance of the above estimators. Three real data sets were analyzed to show the applicability of the proposed models. The introduced model outperformed the other competing models in terms of AIC, BIC and DI.

This class of distributions falls under the modified Poisson mixture models where the rate parameter of Poisson distribution follows a one parameter continuous random variable namely Garima distribution. One can generalize this model as a family of Poisson mixture distributions where the intensity parameter follows some non-negative distributions other than Garima distribution. Also, we can construct regression models for a response variable with abundance of occurrence of zero and one in presence of covariates. Moreover, one can develop an auto regressive count time series models based on the ZOIPG distribution. We leave these aspects for our future research study.

Acknowledgements

The authors are grateful to the editor and reviewer for their guidance and valuable comments and suggestions which led to substantial improvement of the earlier draft. Divya A. and Prasanth C. B. acknowledges the valuable support of the Department of Statistics, St.Thomas college, Thrissur. Muhammed Anvar P. wishes to acknowledge the project grant received under PM USHA Scheme G.O number G.O.(Rt)No.239/2025/HEDN dated 22.02.2025

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

Altun, E., Alqifari, H., and Eliwa, M. S. (2023). A novel approach for zero-inflated count regression model: Zero-inflated Poisson generalized-Lindley linear model with applications. *AIMS Mathematics*, **8**, 23272–23290.

Garay, A. M., Medina, F. L., Jales CS, I., and Bertail, P. (2022). First-order integer valued ar processes with zero-inflated innovations. In *Nonstationary Systems: Theory and Applications: Contributions to the 13th Workshop on Nonstationary Systems and Their Applications, February 3-5, 2020, Grodek nad Dunajcem, Poland 13*, pages 19–40. Springer.

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

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

McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. John Wiley & Sons.

Melkersson, M. and Olsson, C. (1999). *Is Visiting the Dentist a Good Habit?: Analyzing Count Data with Excess Zeros and Excess Ones*. University of Umeå.

Qi, X., Li, Q., and Zhu, F. (2019). Modeling time series of count with excess zeros and ones based on INAR (1) model with zero-and-one inflated Poisson innovations. *Journal of Computational and Applied Mathematics*, **346**, 572–590.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shanker, R. (2016). Garima distribution and its application to model behavioral science data. *Biometrics and Biostatistics International Journal*, **4**, 1–9.

Shanker, R. (2017). The discrete Poisson-Garima distribution. *Biometrics and Biostatistics International Journal*, **5**, 1–7.

Shanker, R. and Shukla, K. K. (2017). Zero-truncated Poisson-Garima distribution and its applications. *Biostat Biometrics Open Access Journal*, **3**, 555605.

Tajuddin, R. R. M., Ismail, N., Ibrahim, K., and Bakar, S. A. A. (2022). A new zero-one-inflated Poisson-Lindley distribution for modelling overdispersed count data. *Bulletin of the Malaysian Mathematical Sciences Society*, **45**, 21–35.

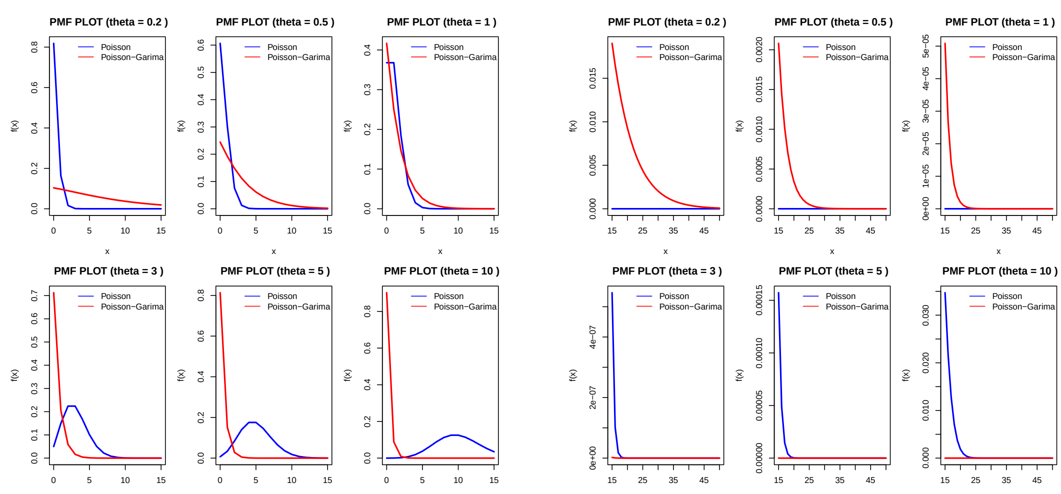
Tang, Y., Liu, W., and Xu, A. (2017). Statistical inference for zero-and-one-inflated Poisson models. *Statistical Theory and Related Fields*, **1**, 216–226.

Yang, M., Zamba, G., Cavanaugh, J., and Yang, M. M. (2018). Package 'zim'. .

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621–629.

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

Appendix I



(a) **(b)**

Figure 1: Probability plots for Poisson and Poisson-Garima

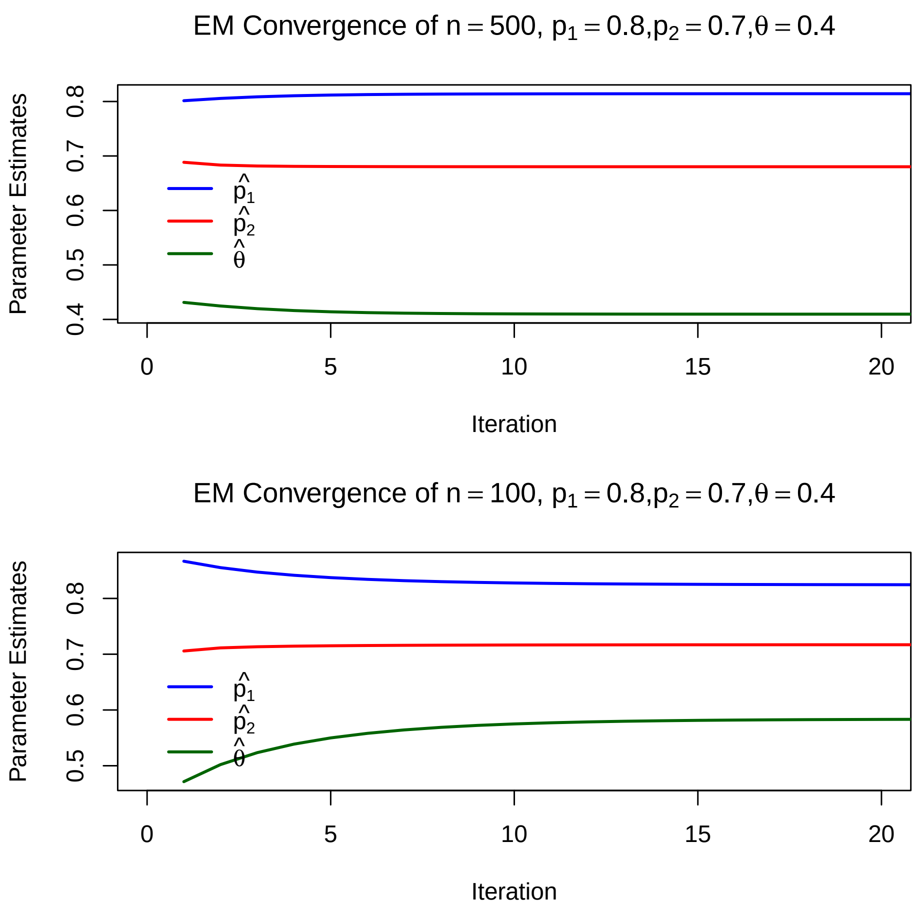
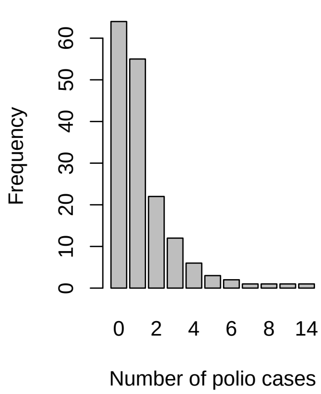
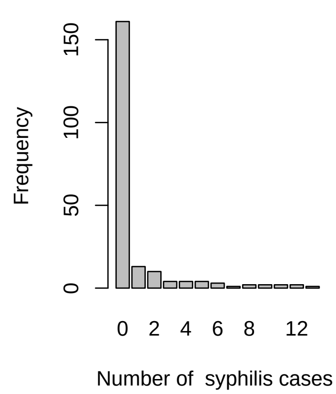


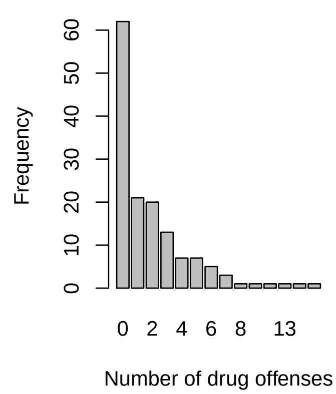
Figure 2: Convergence of EM estimates



(a) Barplot of monthly polio data



(b) Barplot of weekly number of syphilis cases



(c) Barplot of monthly number of drug offenses

Figure 3: Barplots of datasets

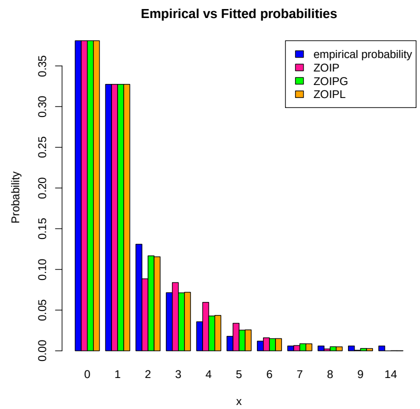


Figure 4: Bar Plot of fitted probabilities of polio data

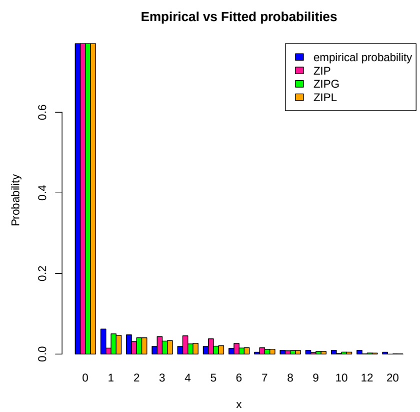


Figure 5: Bar Plot of fitted probabilities of syphilis cases data

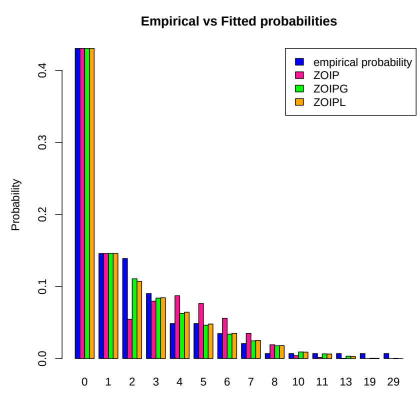


Figure 6: Bar Plot of fitted probabilities of drug data

Table 1: Parameter estimates of ZOIPG distribution with $\theta = 0.4$.

			MOM			MLE			EM		
n	p_1	p_2	$\tilde{\theta}$	$\tilde{p}_1$	$\tilde{p}_2$	$\hat{\theta}$	$\hat{p}_1$	$\hat{p}_2$	$\hat{\theta}_{EM}$	$\hat{p}_{1EM}$	$\hat{p}_{2EM}$
100			0.4457	0.3427	0.2063	0.4073	0.1225	0.4537	0.4070	0.1925	0.3537
200	0.2	0.4	0.4248	0.2977	0.2162	0.4038	0.1245	0.4471	0.4041	0.1954	0.3775
1000			0.7975	0.2825	0.3895	0.7927	0.0845	0.4767	0.7923	0.2045	0.4044
100			0.4457	0.3427	0.2069	0.4073	0.1391	0.4943	0.4100	0.186	0.4263
200	0.2	0.5	0.425	0.2977	0.2257	0.4038	0.1266	0.6027	0.4049	0.1937	0.4703
1000			0.4066	0.1814	0.3536	0.3985	0.1291	0.6746	0.3988	0.2036	0.5004
100			0.4678	0.3923	0.4080	0.4128	0.3086	0.6507	0.4141	0.4875	0.4990
200	0.5	0.5	0.4370	0.4226	0.4826	0.4053	0.3140	0.6726	0.4054	0.4970	0.4987
1000			0.4099	0.4732	0.5521	0.4002	0.3150	0.6712	0.4001	0.4983	0.4981
100			0.5170	0.5649	0.5781	0.4279	0.4284	0.6638	0.4283	0.6894	0.4977
200	0.7	0.5	0.4771	0.6165	0.5721	0.4268	0.4253	0.6625	0.4247	0.6886	0.4912
1000			0.4110	0.6823	0.5228	0.3999	0.4433	0.6707	0.3999	0.7006	0.4980
100			0.5151	0.5686	0.5660	0.4292	0.4277	0.5187	0.4537	0.6669	0.6810
200	0.7	0.7	0.4771	0.6165	0.5705	0.4268	0.4253	0.4861	0.4247	0.6886	0.7123
1000			0.4111	0.6809	0.7332	0.3977	0.4449	0.6347	0.3977	0.7017	0.6979

Table 2: Parameter estimates of ZOIPG distribution with θ=0.8

			MOM			MLE			EM		
n	p_1	p_2	$\tilde{\theta}$	$\tilde{p}_1$	$\tilde{p}_2$	$\hat{\theta}$	$\hat{p}_1$	$\hat{p}_2$	$\hat{\theta}_{EM}$	$\hat{p}_{1EM}$	$\hat{p}_{2EM}$
100			0.9528	0.4040	0.4159	0.8406	0.1817	0.3501	0.8080	0.2028	0.3188
200	0.2	0.4	0.8422	0.3602	0.3650	0.7884	0.1172	0.4025	0.7801	0.2105	0.3578
1000			0.7975	0.2825	0.3895	0.7928	0.0845	0.4767	0.7923	0.2045	0.4044
100			0.9529	0.4042	0.4149	0.8405	0.1817	0.3672	0.8203	0.1878	0.3968
200	0.2	0.5	0.8422	0.3603	0.3741	0.7884	0.1172	0.4287	0.7887	0.1988	0.4459
1000			0.7975	0.2825	0.4325	0.7928	0.0845	0.7097	0.7929	0.2037	0.5057
100			1.0290	0.4148	0.4902	0.8621	0.2029	0.5395	0.8569	0.4576	0.4507
200	0.5	0.5	0.9146	0.4174	0.4908	0.8275	0.1976	0.6362	0.8302	0.4775	0.4794
1000			0.8031	0.4823	0.5177	0.7942	0.2063	0.7169	0.7941	0.5013	0.5039
100			1.1868	0.5223	0.4732	0.9192	0.2768	0.5827	0.9167	0.6554	0.4663
200	0.7	0.5	0.9735	0.5881	0.4889	0.8449	0.2785	0.6467	0.8477	0.6844	0.4897
1000			0.8342	0.6729	0.5038	0.8048	0.2856	0.6997	0.8048	0.6992	0.4985
100			1.1870	0.5220	0.6474	0.9192	0.2768	0.2203	0.9945	0.6030	0.6447
200	0.7	0.7	0.9735	0.5881	0.7109	0.8449	0.2785	0.1494	0.8716	0.6682	0.6854
1000			0.8343	0.6729	0.7178	0.8048	0.2856	0.1000	0.8048	0.6992	0.7013

Table 3: RMSE of ZOIPG distribution with θ =0.4

			MOM			MLE			EM		
n	p_1	p_2	$\tilde{\theta}$	$\tilde{p_1}$	$\tilde{p_2}$	$\hat{\theta}$	$\hat{p_1}$	$\hat{p_2}$	$\hat{\theta}_{EM}$	$\hat{p_1}_{EM}$	$\hat{p_2}_{EM}$
100			0.1122	0.2298	0.2931	0.0592	0.1000	1.5571	0.0592	0.0900	0.2117
200	0.2	0.4	0.0755	0.2020	0.2891	0.0400	0.0889	0.7426	0.0400	0.0686	0.1609
1000			0.0985	0.1703	0.1575	0.0520	0.1183	0.2322	0.0520	0.0510	0.0922
100			0.1122	0.2298	0.3672	0.0600	0.1063	0.2924	0.0616	0.0975	0.2263
200	0.2	0.5	0.0755	0.2020	0.3610	0.0400	0.0894	0.2579	0.0412	0.0714	0.1587
1000			0.0316	0.1030	0.3226	0.0176	0.0728	0.1910	0.0173	0.0245	0.0490
100			0.1459	0.2102	0.3081	0.0800	0.2035	0.2193	0.0819	0.0889	0.0954
200	0.5	0.5	0.0985	0.1749	0.2769	0.0520	0.1926	0.1985	0.0529	0.0608	0.0608
1000			0.0424	0.0831	0.1497	0.0224	0.1860	0.1758	0.0224	0.0224	0.0245
100			0.2124	0.2175	0.2289	0.1122	0.2835	0.2035	0.1131	0.0787	0.0700
200	0.7	0.5	0.1884	0.1741	0.1887	0.0849	0.2804	0.1780	0.0837	0.0490	0.0436
1000			0.0583	0.0678	0.0819	0.0332	0.2583	0.1738	0.0332	0.0224	0.0200
100			0.2161	0.2161	0.3592	0.1145	0.2839	0.4468	0.1936	0.1517	0.1030
200	0.7	0.7	0.1884	0.1741	0.3406	0.0849	0.2804	0.4734	0.0831	0.0490	0.0458
1000			0.0557	0.0671	0.0949	0.0300	0.2563	0.4241	0.0300	0.0224	0.0200

Table 4: RMSE of estimates of ZOIPG distribution with θ=0.8

			MOM			MLE			EM		
n	p_1	p_2	$\tilde{\theta}$	$\tilde{p_1}$	$\tilde{p_2}$	$\hat{\theta}$	$\hat{p_1}$	$\hat{p_2}$	$\hat{\theta}_{EM}$	$\hat{p_1}_{EM}$	$\hat{p_2}_{EM}$
100			0.3503	0.2621	0.2490	0.2059	0.1694	0.3126	0.1487	0.1323	0.2681
200	0.2	0.4	0.2015	0.2466	0.2128	0.1109	0.1367	0.2897	0.1057	0.1054	0.2163
1000			0.0985	0.1703	0.1575	0.0520	0.1179	0.2322	0.0520	0.0510	0.0922
100			0.3503	0.2621	0.2655	0.2059	0.1694	0.3561	0.1572	0.1400	0.2933
200	0.2	0.5	0.2015	0.2466	0.2559	0.1109	0.1367	0.3342	0.1109	0.1145	0.2332
1000			0.0985	0.1703	0.2093	0.0520	0.1179	0.2903	0.0520	0.0520	0.0768
100			0.4631	0.2095	0.2441	0.2709	0.3138	0.2978	0.2445	0.1664	0.1718
200	0.5	0.5	0.2800	0.2052	0.1931	0.1616	0.3094	0.2615	0.1649	0.1136	0.1005
1000			0.1091	0.1170	0.0624	0.0616	0.2950	0.2276	0.0616	0.0424	0.0283
100			0.8141	0.2750	0.2020	0.5019	0.4382	0.2657	0.4004	0.1584	0.0269
200	0.7	0.5	0.3780	0.2102	0.1200	0.2149	0.4287	0.2347	0.2218	0.0922	0.0616
1000			0.1476	0.0911	0.0300	0.0900	0.4163	0.3783	0.0900	0.0387	0.0200
100			0.4004	0.1584	0.1269	0.5019	0.4382	0.5590	0.5085	0.2581	0.1741
200	0.7	0.7	0.2218	0.0922	0.0616	0.2149	0.4287	0.5851	0.2812	0.1449	0.0970
1000			0.0900	0.0387	0.0200	0.0900	0.4163	0.6000	0.0900	0.0387	0.0200

Table 5: Parameter estimates and their RMSE of ZIPG model

			MOM		RMSE		MLE		RMSE	
n	θ	p_1	$\tilde{\theta}$	$\tilde{p}_1$	$\tilde{\theta}$	$\tilde{p}_1$	$\hat{\theta}$	$\hat{p}_1$	$\hat{\theta}$	$\hat{p}_1$
100			0.2369	0.0243	0.0374	0.1758	0.2241	0.2309	0.0241	0.0308
200	0.2	0.2	0.1721	0.2775	0.0279	0.0775	0.2139	0.1566	0.0138	0.0436
1000			0.2004	0.2532	0.0004	0.0532	0.2106	0.1807	0.0105	0.0192
100			0.2096	0.4198	0.0096	0.0802	0.2249	0.4116	0.0249	0.0883
200	0.2	0.5	0.1707	0.5174	0.0293	0.0173	0.2289	0.4669	0.0289	0.0332
1000			0.1974	0.5344	0.0026	0.0346	0.2131	0.5010	0.0130	0.0010
100			0.2049	0.6575	0.0049	0.0425	0.2569	0.6772	0.0566	0.0228
200	0.2	0.7	0.1767	0.6290	0.0232	0.0707	0.2144	0.7189	0.0141	0.0190
1000			0.2044	0.7155	0.0044	0.0155	0.2171	0.6982	0.0170	0.0018
100			0.4431	0.6777	0.1212	0.0768	0.4185	0.6963	0.0883	0.0548
200	0.4	0.7	0.4198	0.6889	0.0837	0.0548	0.4085	0.6978	0.0648	0.0387
1000			0.4046	0.6979	0.0346	0.0245	0.4020	0.7000	0.0265	0.0173
100			1.0900	0.0711	0.2972	0.1288	0.8915	0.2519	0.0917	0.0520
200	0.8	0.2	0.7042	0.3856	0.0959	0.1857	0.8126	0.1244	0.0126	0.0755
1000			0.8217	0.1877	0.0217	0.0122	0.8147	0.2361	0.0148	0.0361
100			0.9431	0.4228	0.1432	0.0775	0.9374	0.5176	0.1374	0.0173
200	0.8	0.5	0.7586	0.5963	0.0412	0.0964	0.8700	0.3805	0.0700	0.1196
1000			0.8926	0.4621	0.0927	0.0374	0.7803	0.5094	0.0197	0.0094
100			1.0030	0.6309	0.2040	0.0693	1.0750	0.7159	0.2750	0.0173
200	0.8	0.7	0.6375	0.7758	0.1625	0.0755	0.8778	0.5961	0.0781	0.1039
1000			0.9161	0.6739	0.1162	0.0261	0.8206	0.7160	0.0207	0.0158

Table 6: Computation time comparison for ZOIPG and ZIPG models across different estimation methods

		ZOIPG		ZIPG
Sample Size (n)	Method	Time(in seconds)	Method	Time(in seconds)
	MOM	0.2664	MOM	0.1637
100	MLE	0.7506	MLE	1.7238
	EM	115.5400		
	MOM	0.3100	MOM	0.2340
200	MLE	0.8207	MLE	2.9412
	EM	210.2300		
	MOM	0.6338	MOM	0.4839
1000	MLE	1.2178	MLE	14.1458
	EM	619.9900		

Table 7: Fitted distributions on Dataset-I, along with their expected frequencies, estimates of parameters and model selection statistics.

Distribution of monthly number of Polio cases				
		Fitted Distributions with expected frequencies		
x	f	ZOIP	ZOIPG	ZOIPL
0	64	64	64	64
1	55	55	55	55
2	22	15	20	19
3	12	14	12	12
4	6	10	7	7
5	3	6	4	4
6	2	3	3	3
7	1	1	1	2
8	1	0	1	1
9	1	0	1	1
14	1	0	0	0
Total	168	168	168	168
Estimates	p_1	0.6241	0.2288	0.2742
	p_2	0.5753	0.3878	0.4548
	θ	2.8421	0.8712	0.9316
AIC		554.6511	**536.0119**	536.387
BIC		564.023	**545.4384**	545.7589
χ^2		5.2524	**0.4429**	0.9802
d.f		2	2	2
p-value		0.0724	**0.8014**	0.6126
DI	2.6287	1.9437	**2.2516**	2.2240

Table 8: Fitted distributions on Dataset-II, along with their expected frequencies, estimates of parameters and information statistics.

Distribution of weekly number of syphilis cases				
		Fitted Distributions with expected frequencies		
x	f	ZIP	ZIPG	ZIPL
0	161	161	161	161
1	13	3	11	10
2	10	7	9	9
3	4	9	7	7
4	4	9	5	6
5	4	8	4	5
6	3	6	3	3
7	1	3	2	3
8	2	2	2	2
9	2	1	2	1
10	2	0	2	1
12	2	0	1	1
20	1	0	0	0
Total	209	209	209	209
Estimates	p_1	0.7667	0.7095	0.7211
	θ	4.19	0.4194	0.4802
AIC		512.5916	**452.5824**	453.8479
BIC		519.2763	**459.2671**	460.5325
χ^2		46.34127	**2.0716**	7.0302
d.f		5	4	5
p-value		$< .001$	**0.7226**	0.2184
DI	6.8410	4.209	**6.1837**	5.8709

Table 9: Fitted distributions on Dataset-III, along with their expected frequencies, estimates of parameters and model selection statistics.

Distribution of monthly number of drug offenses				
		Fitted Distributions with expected frequencies.		
x	f	ZOIP	ZOIPG	ZOIPL
0	62	62	62	62
1	21	21	21	21
2	20	8	16	16
3	13	12	13	12
4	7	13	9	10
5	7	11	7	7
6	5	8	5	5
7	3	5	4	4
8	1	3	3	3
10	1	1	2	2
11	1	0	1	1
13	1	0	1	1
19	1	0	0	0
29	1	0	0	0
Total	144	144	144	144
Estimates	p_1	0.5458	0.2508	0.3026
	p_2	0.7784	0 .9913	0.9463
	θ	4.3820	0.4962	0.5491
AIC		641.4777	**559.5282**	561.2334
BIC		650.3872	**568.4377**	570.1428
χ^2		23.4321	**1.8081**	2.3470
d.f		4	4	6
p-value		0.0001	**0.7710**	0.6722
DI	6.1156	3.0198	**4.0614**	3.9262

Appendix II

R code for ZOIPG Simulation

```
library(rootSolve)
sim = 1000
n = 100
p0 = 0.3
p1 = 0.9
theta = 0.5
p = (theta + 1) / (theta + 2)

MM1 = MM2 = MM3 = c()
ML1 = ML2 = ML3 = c()
EM1 = EM2 = EM3 = c()
```

```
for (i in 1:sim) {
 x1=rexp(n,theta)
x2=rgamma(n,rate=theta,shape=2)
mix=runif(n)
data=ifelse(mix<p,x1,x2)
V=rpois(n,data)

  B1 = rbinom(n, 1, p0)
  B2 = rbinom(n, 1, p1)
  y = B1 * (1 - B2) + (1 - B1) * V

  # Method of Moments
  fm1 = mean(y)
  fm2 = mean(y * (y - 1))
  fm3 = mean(y * (y - 1) * (y - 2))
  s1 = fm3 / fm2
  MM3[i] = ((3 - 4 * s1) + sqrt((4 * s1 + 3)^2 + 12 * s1)) / (2 *
     s1)
  MM1[i] = 1 - (fm2 * (MM3[i] + 2) * MM3[i]^2) / (2 * (MM3[i] + 4))
  muhat = (MM3[i] + 3) / (MM3[i] * (MM3[i] + 2))
  p21 = MM1[i] - fm1 + (1 - MM1[i]) * muhat
  MM2[i] = p21 / MM1[i]
  if (MM1[i] < 0 || MM1[i] > 1) MM1[i] = 0.5
  if (MM2[i] < 0 || MM2[i] > 1) MM2[i] = 0.1

  # MLE Estimation
  s0 = sum(y == 0)
  s1 = sum(y == 1)
  y1 = y[y >= 2]

  loglt = function(x) {
    a1 = n - s0 - s1
    a2 = (4 * x^2 + 4 * x + 2) / (x * (x + 1) * (x^2 + 5 * x + 2))
    a3 = sum((2 * x + 3 + y1) / (x^2 + 3 * x + 1 + x * y1))
    a4 = sum(y1 / (x + 1))
    F = a1 * a2 + a3 - a4
    return(F)
  }
  ML3[i] = uniroot(loglt, c(0, 100))$root

  loglp = function(x) {
    p00 = x[1]; p10 = x[2]
    c1 = (ML3[i] * (ML3[i]^2 + 3 * ML3[i] + 1)) / ((ML3[i] + 2) *
       (ML3[i] + 1)^2)
    c2 = (ML3[i] * (ML3[i]^2 + 4 * ML3[i] + 1)) / ((ML3[i] + 2) *
       (ML3[i] + 1)^3)
    F1 = c1 * (1 - p00) + p00 * p10 - s0 / n
    F2 = c2 * (1 - p00) + p00 * (1 - p10) - s1 / n
    return(c(F1, F2))
  }
```

```
ML1[i] = multiroot(loglp, c(0.5, 0.5))$root[1]
ML2[i] = multiroot(loglp, c(0.5, 0.5))$root[2]

if (ML1[i] < 0 || ML1[i] > 1) ML1[i] = 0.5
if (ML2[i] < 0 || ML2[i] > 1) ML2[i] = 0.1

# EM Algorithm
EMfit = function(par) {
  EM = function(para) {
    EMp0 = para[1]; EMp1 = para[2]; EMtheta = para[3]
    a1 = (EMp0 * EMp1) / (EMp0 * EMp1 + (1 - EMp0) * (EMtheta *
       (EMtheta^2 + 3 * EMtheta + 1)) / ((EMtheta + 2) * (EMtheta
       + 1)^2))
    a2 = (EMp0 * (1 - EMp1)) / (EMp0 * (1 - EMp1) + (1 - EMp0) *
       (EMtheta * (EMtheta^2 + 4 * EMtheta + 1)) / ((EMtheta + 2)
       * (EMtheta + 1)^3))
    B1em = a1 * (y == 0) + a2 * (y == 1)
    b1 = (EMp0 * EMp1 + (1 - EMp0) * EMp1 * EMtheta * (EMtheta^2
       + 3 * EMtheta + 1) / ((EMtheta + 2) * (EMtheta + 1)^2)) /
         (EMp0 * EMp1 + (1 - EMp0) * EMtheta * (EMtheta^2 + 3 *
            EMtheta + 1) / ((EMtheta + 2) * (EMtheta + 1)^2))
    b2 = ((1 - EMp0) * EMp1 * EMtheta * (EMtheta^2 + 4 * EMtheta
       + 1) / ((EMtheta + 2) * (EMtheta + 1)^3)) /
         (EMp0 * (1 - EMp1) + (1 - EMp0) * EMtheta * (EMtheta^2 +
            4 * EMtheta + 1) / ((EMtheta + 2) * (EMtheta + 1)^3))
    b3 = EMp1
    B2em = b1 * (y == 0) + b2 * (y == 1) + b3 * (y >= 2)

    logtheta = function(th) {
      -sum((1 - B1em) * log(th * (th * y + th^2 + 3 * th + 1) /
         ((th + 2) * (th + 1)^(y + 2))))
    }
    thetaEM = optimize(logtheta, c(0, 100))$minimum
    para[1] = mean(B1em)
    para[2] = mean(B2em)
    para[3] = thetaEM
    loglik = function(theta) {
      p01 = theta[1]; p11 = theta[2]; th = theta[3]
      if (p01 < 0 || p01 > 1) p01 = 0.5
      if (p11 < 0 || p11 > 1) p11 = 0.5
      logL = sum(B1em * log(p01) + (1 - B1em) * (log(1 - p01) + 2
         * log(th) + log(y + th + 2) - (y + 3) * log(th + 1)) +
                 B2em * log(p11) + (1 - B2em) * log(1 - p11))
      return(logL)
    }
    list(pa = para, ll = loglik(para))
  }
  iter = 0; para.old = par; like.old = EM(par)$ll
  repeat {
    iter = iter + 1
```

```
      para.new = EM(para.old)$pa
      loglik.new = EM(para.new)$ll
      if (abs(loglik.new - like.old) < 0.001) break
      para.old = para.new; like.old = loglik.new
    }
    list(para = para.new, loglik = loglik.new)
  }
  mm = c(MM1[i], MM2[i], MM3[i])
  EM1[i] = EMfit(mm)$para[1]
  EM2[i] = EMfit(mm)$para[2]
  EM3[i] = EMfit(mm)$para[3]
}

True = c(theta, p0, p1)
MM = c(mean(MM3), mean(MM1), mean(MM2))
MLE = c(mean(ML3), mean(ML1), mean(ML2))
EM = c(mean(EM3), mean(EM1), mean(EM2))

MSEmm = c(mean((MM3 - theta)^2), mean((MM1 - p0)^2), mean((MM2 -
   p1)^2))
MSEml = c(mean((ML3 - theta)^2), mean((ML1 - p0)^2), mean((ML2 -
   p1)^2))
MSEem = c(mean((EM3 - theta)^2), mean((EM1 - p0)^2), mean((EM2 -
   p1)^2))

cbind(True, MM, MSEmm, MLE, MSEml, EM, MSEem)
```

More R codes are available at https://github.com/Divya-1203/zoipg/blob/main/zoipggithub.R