Comparison of estimation methods for one-inflated positive Poisson distribution

Razik Ridzuan Mohd Tajuddin, Noriszura Ismail and Kamarulzaman Ibrahim

Department of Mathematical Sciences, Universiti Kebangsaan Malaysia, Bangi, Malaysia

ABSTRACT

This paper aims to propose estimation methods for one-inflated positive Poisson (OIPP) distribution and compare their properties in terms of unbiasedness, consistency, efficiency and deficiency. All estimators considered in the study are asymptotically unbiased and consistent. The maximum likelihood estimator (MLE) for the OIPP distribution is asymptotically normal. When compared to the MLE, the ordinary least square estimator (OLSE) is the most efficient, followed by the method of moments estimator (MME) and the ratio of probability estimator (RPE). A novel one-inflation index was also proposed to assess the presence of excess ones in the dataset for the positive Poisson distribution to determine whether a one-inflated distribution is required for model fitting. A real dataset with a large number of ones, as identified by the proposed one-inflation index, was used for model fitting. It is found that the OLSE and MLE are the best estimators for an OIPP distribution.

1. Introduction

In discrete count data distribution, one method of modelling positive data is to truncate the distribution at zero, resulting in a zero-truncated distribution. Positive count data modelling can be traced back to the mid-twentieth century, when the first zero-truncated distribution, known as the zero-truncated Poisson distribution [1], was proposed. The probability mass function of the zero-truncated Poisson distribution is given as

\[ Pr(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!(\exp(-\lambda) - 1)}; x = 1, 2, 3, \ldots, \]

where \( \lambda > 0 \) and the maximum likelihood estimator (MLE) of \( \lambda \) can be obtained numerically by solving

\[ \hat{\lambda} = \frac{\hat{x}}{1 - \exp(-\hat{\lambda})}. \]  

(1)

Subsequently, other estimators for a zero-truncated Poisson distribution were developed and analysed [2–5]. An asymptotically unbiased estimator of \( \lambda \) was proposed [2] as follows:

\[ \hat{\lambda}_2 = \left( \sum_{x=0}^{r} nx \right) / \left( \sum_{x=0}^{r} nx \right), \]

where \( n_x \) is the frequency of \( x \)-valued data and \( r \) is the maximum number taken by \( x \). Similarly, an efficient estimator of \( \lambda \) was proposed [3] as follows:

\[ \hat{\lambda}_3 = \frac{1}{n} \sum_{x=2}^{\infty} nx, \]

(3)

Both (1) and (3) were used to estimate the mortality rate of the number of gall-cells per flower-head [4], and it was discovered that both estimators provide similar estimates for mortality rate estimation. In addition to the estimators in (1–3), a minimum variance unbiased estimator of \( \lambda \) was proposed [5] as follows:

\[
\hat{\lambda}_4 = \begin{cases} 
\frac{t}{n} C(n, t); & 1 \leq n \leq t, 1 \leq t \\
\frac{t}{n} \left[ 1 - \left( \frac{n - 1}{n} \right)^{t-1} \right]; & 2 \leq n \leq 15, t \\
\hat{\lambda}_1 \text{ or } \hat{\lambda}_3; & \text{otherwise,} 
\end{cases}
\]

(4)

where \( t = nx \), \( C(n, t) = 1 - S_{t-1}^n / S_t^n \) and \( S_t^n \) is the Stirling number of the second kind. The function used for this estimator is dependent on the subdomain as shown in (4).

The negative binomial distribution was modified to the zero-truncated negative binomial distribution as an
2. One-inflated positive Poisson (OIPP) distribution and its statistical properties

Let \( Y \sim \text{OIPP}(\omega, \lambda) \), then the probability mass function of \( Y \) is given in (5), where \( \lambda > 0 \) refers to the rate parameter and \( \omega \) refers to the one-modification parameter for the OIPP distribution.

\[
\Pr(Y = y | \omega, \lambda) = \begin{cases} \omega + (1 - \omega) \frac{\lambda^y}{\exp(\lambda) - 1}, & y = 1 \\ (1 - \omega) \frac{\lambda^y}{y! \exp(\lambda - 1)}, & y \geq 2 \end{cases}
\]

If \( \omega > 0 \), then the distribution is known as an OIPP distribution [9]. If \( \omega = 0 \), the OIPP distribution is reduced to a zero-truncated Poisson distribution [1]. It is possible for \( \omega \) to be negative, such as \( \omega \in (-\lambda/\exp(\lambda) - 1 - \lambda, 0) \), in which case the distribution is known as a one-deflated positive Poisson distribution. In this study, we restrict \( \omega > 0 \) to ensure the one-inflation property. The formulae for the first two moments about the origin, the dispersion index and the moment generating function are given respectively as

\[
\begin{align*}
\mu &= \omega + (1 - \omega) \frac{\lambda \exp(\lambda)}{\exp(\lambda) - 1} \\
\mu_2 &= \omega + (1 - \omega) \frac{\lambda \exp(\lambda)[\lambda + 1]}{\exp(\lambda) - 1} = \mu + (\mu - \omega)\lambda, \\
d &= \sigma^2/\mu = 1 + \lambda - \omega\lambda/\mu, \\
M_Y(t) &= \omega \exp(t) + (1 - \omega) \frac{\exp[\lambda \exp(t)] - 1}{\exp(\lambda) - 1}. \\
\end{align*}
\]

Figure 1 illustrates the dispersion index of various pairs of parameters \( \lambda \) and \( \omega \). The heatmap in Figure 1 shows that the OIPP distribution can be either underdispersed or overdispersed.

For \( y \geq 2 \), the OIPP distribution is unimodal, which can be deduced based on the decreasing function from the following ratio:

\[
\frac{\Pr(Y = y + 1 | \omega, \lambda)}{\Pr(Y = y | \omega, \lambda)} = \frac{(1 - \omega) \frac{\lambda^{y+1}}{(y+1)! \exp(\lambda - 1)}}{(1 - \omega) \frac{\lambda^y}{y! \exp(\lambda - 1)}} = \frac{\lambda}{y + 1}.
\]

3. Estimation of parameters for OIPP distribution

Several estimation methods for an OIPP distribution were developed based on the method of moments, maximum likelihood, one-proportion, ratio of probability and ordinary least squares. Fitting a model to a dataset based on its method of moments, maximum likelihood and/or ordinary least squares estimators is a common practice in statistical modelling. On the other hand, the one-proportion and ratio of probability estimators are rarely used.

3.1. Method of moments estimator (MME)

The method of moments estimator (MME) is obtained by equating sample moments with theoretical moments. By equating the first two sample moments with the first two equations in (6), the MMEs of \( \lambda \) and \( \omega \) are obtained by solving

\[
(1 - m_1)\tilde{\lambda}^2 \exp(\tilde{\lambda}) + (m_2 - m_1)\tilde{\lambda} \exp(\tilde{\lambda}) + (m_1 - m_2)\exp(\tilde{\lambda} - 1) = 0
\]

and

\[
\hat{\omega} = \frac{(m_1 - \tilde{\lambda}) \exp(\tilde{\lambda}) - m_1}{\exp(\tilde{\lambda}) - 1 - \tilde{\lambda} \exp(\tilde{\lambda})}
\]

where \( m_k = \sum_{i=1}^{n} x_i^k / n \) for \( n \) data, and \( \hat{\omega} \) and \( \tilde{\lambda} \) are the respective MMEs of \( \omega \) and \( \lambda \).
3.2. Maximum likelihood estimation (MLE)

The log-likelihood function $l$ for a random variable $Y$ that follows the OIPP distribution is given as:

$$l \propto n_1 \ln \left[ \omega + (1 - \omega) \frac{\lambda}{\exp(\lambda) - 1} \right] + (n - n_1) \left[ \ln(1 - \omega) - \ln[\exp(\lambda) - 1] \right] + \sum_{y=2}^{\infty} n_y \ln(y),$$

where $n_y$ refers to the number of $y$-valued observations and $n = \sum_{y=1}^{\infty} n_y$. By differentiating $l$ with respect to $\omega$ and $\lambda$ and setting it to zero, the respective MLEs of $\omega$ and $\lambda$ can be obtained by solving

$$\hat{\omega} = \frac{n_1[\exp(\hat{\lambda}) - 1] - n\hat{\lambda}}{n[\exp(\hat{\lambda}) - 1 - \hat{\lambda}]}$$

and

$$(\hat{\lambda} - A) \exp(\hat{\lambda}) + (A - 1)\hat{\lambda} + A = 0,$$

where $A = (nm_1 - n_1)/(n - n_1)$ and $\hat{\omega}$ and $\hat{\lambda}$ are the respective MLEs of $\omega$ and $\lambda$.

A one-proportion estimator (OPE) is the extension to the estimator for the generalized negative binomial distribution. The OPE for the generalized negative binomial distribution can be obtained by comparing the first two sample moments and the sample proportion of zeros to the population proportion [13]. Similarly, the zero-inflated Poisson distribution parameter estimator is obtained by equating the empirical probability with the theoretical probability of zero-valued observations [14]. For the OIPP distribution, the OPE of $\omega$ is obtained by equating the theoretical proportion of ones to the sample proportion of one, while the OPE of $\lambda$ is obtained by equating the population mean to the sample mean. Surprisingly, the OPEs of $\omega$ and $\lambda$ are identical to their respective MLEs.

Theorem 1: The MLE $\hat{\lambda}$ of $\lambda$ is consistent and asymptotically normal, such that:

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, I^{-1}(\lambda)),$$

where

$$I(\lambda) = \left[ \omega + (1 - \omega) \frac{\lambda}{u} \right] \left[ (1 - \omega)uv(\lambda e^v) + e^v(u - \lambda e^v) \right] \times \frac{(\omega u - v + 1) + \omega u + 1}{u^2v^2}$$
Theorem 2: The MLE \( \hat{\lambda} \) is consistent and asymptotically normal, such that
\[
\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, I^{-1}(\lambda)),
\]
where
\[
I(\lambda) = \lambda - \frac{u}{\lambda} \left\{ (u - \lambda)(\omega(u - \lambda) + \lambda) + \frac{1}{1 - \omega} \right\}
\]
is the Fisher information of \( \lambda \) with \( u = \exp(\lambda) - 1 \) and \( v = \omega(u - \lambda) + \lambda \).

Proof: The regularity conditions under which the MLE \( \hat{\lambda} \) is consistent and asymptotically normal is satisfied by the OIPP distribution (see [1, Chapter 6]), therefore
\[
l(\lambda) = -E \left( \frac{\partial^2 \ln f(y)}{\partial \lambda^2} \right)
= -E \left[ \frac{\partial^2}{\partial \lambda^2} \left( l_{y=1} \ln \left( \omega + (1 - \omega) \frac{\lambda}{\exp(\lambda) - 1} \right)
+ l_{y>1} \ln(1 - \omega) + y \ln \lambda
- \ln y! - \ln(\exp(\lambda) - 1) \right) \right]
= -E[l_{y=1}^2 + l_{y>1}Q] = - \left[ Pf(1) + \sum_{y=2}^{\infty} Qf(y) \right],
\]
where \( f(y) = Pr(Y = y) \), and \( P \) and \( Q \) are respectively given as
\[
P = \frac{(1 - \omega)uv[-\lambda \exp(\lambda)]}{u^2 v^2},
\]
and
\[
Q = -\frac{y}{\lambda^2} + \frac{\exp(\lambda)}{u^2}.
\]
Solving the summation in \( l(y) \) yields
\[
l(\lambda) = \left[ \omega + (1 - \omega) \frac{\lambda}{u} \right]
\times \frac{[(1 - \omega)uv(\lambda e^\lambda) + e^\lambda (u - \lambda e^\lambda) (ouv - v + 1) + ouv + 1]}{u^2 v^2}
+ \frac{1 - \omega}{\lambda} \frac{e^\lambda (u - \lambda)(1 - \omega)}{u^3}.
\]
From Theorem 1, the asymptotic 100(1 - \( \alpha \))% confidence interval of \( \lambda \) is given as,
\[
\hat{\lambda} \pm z_{\alpha/2} I^{-1/2}(\lambda) \sqrt{n}.
\]

3.3. Ratio of probability estimator (RPE)

The ratio of probability estimator (RPE) is the extension to the estimator for the generalized negative binomial distribution that can be obtained either by equating the ratio of one-valued observations to two-valued observations or by equating the first two theoretical moments with the sample moments [13]. The same method has been employed to obtain the parameter estimators for the zero-inflated Poisson–Lindley distribution [16]. Our study is premised on [13,16], in which the respective estimators of \( \lambda \) and \( \omega \) were obtained by considering the ratio of probability for 3-valued observations to 2-valued observations and by equating the population mean with the sample mean. Parameter \( \omega \) can be eliminated by considering the ratio of probability for 3-valued observations to 2-valued observations, allowing parameter \( \lambda \) to be easily obtained as follows:
\[
\frac{Pr(Y = 3|\omega, \lambda)}{Pr(Y = 2|\omega, \lambda)} = \frac{(1 - \omega) \lambda^2}{2^{\lambda(1 - \omega)}} = \frac{\lambda}{3}.
\]
By further equating \( \lambda/3 \) to the empirical ratio of \( n_3/n_2 \), the resulting RPE of \( \lambda \) is given as
\[
\tilde{\lambda} = \frac{3n_3}{n_2},
\]
where \( \tilde{\lambda} \) is the RPE of \( \lambda \) and \( \hat{\lambda} \) can only be obtained when both \( n_2 \) and \( n_3 \) are greater than zero. Note that
the formula of $\hat{\lambda}$ is very simple, hence can be solved manually. The calculation of RPE of $\omega$ is similar to the calculation of MME of $\omega$. Notice that the RPE of $\lambda$ is a special case of the general form of the Zelterman estimator studied by Böhning [17] given as $(i + 1)\eta_{i+1}/\eta_{i}$, where $i = 2$. Since the OIPP distribution is an extended zero-truncated Poisson distribution, it is only appropriate to use the first approach given by Böhning [17]. This entails truncating the Poisson model at all counts except 2 and 3, resulting in the following log-likelihood function.

$$
\ell = \ln L = \left[ \frac{1}{1 + \frac{\omega}{3}} \right]^{n_2} \left[ \frac{1}{1 + \frac{\lambda}{3}} \right]^{n_3},
$$

with maximum likelihood estimate $\hat{\lambda} = 3n_3/n_2 = \hat{\lambda}$ (refer to [17] for a detailed explanation). Based on the findings in [17], the variance of RPE of $\lambda$, which is a special case of Zelterman estimator, can be written as

$$
\text{Var}(\hat{\lambda}) = \lambda^2 \left( \frac{1}{n_2} + \frac{1}{n_3} \right).
$$

The variance can be further estimated by substituting $\lambda$ with $\hat{\lambda}$, resulting in

$$
\text{Var}(\hat{\lambda}) = 9 \left[ \frac{n_2}{n^2 / (n_2 + n_3)} \right].
$$

Identical results can be obtained using the second approach given by Böhning [17], which considers the nonparametric multinomial approach. Based on the variance and the estimated variance formulae, it can be concluded that the RPE of $\lambda$ is consistent since as $n$ increases, so does $n_i$. Also, the one-inflation and unimodality properties indicate that $n_2 > n_3$, resulting in zero $\text{Var}(\hat{\lambda})$ and $\text{Var}(\hat{\lambda})$.

### 3.4. Ordinary least squares estimator (OLSE)

The ordinary least squares estimator (OLSE) is an estimator that minimizes the function of the squared difference between theoretical and empirical cumulative functions. Suppose $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ is the order statistics of the data that follows the OIPP distribution. It is known that $E[F(Y_{(i)})] = i/(n + 1)$ for $i = 1, 2, \ldots, n$. However, for count data, the expected value of the order statistics is best expressed with respect to the data frequency, such that $E[F(Y_{(i)})] = \left( \sum_{j=1}^{i} \eta_{j} \right) / (n + 1)$. The OLSE of $\omega$ and $\lambda$ can be obtained by minimizing the following function:

$$
S_{Y}(\omega, \lambda) = \sum_{y=1}^{\infty} \left[ F(y|\omega, \lambda) - \sum_{j=1}^{\infty} \frac{\eta_{j}}{n + 1} \right]^{2}
$$

$$
= \sum_{y=1}^{\infty} \left[ \sum_{j=1}^{y} \left( \omega + (1 - \omega) \frac{\lambda}{\exp(\lambda) - 1} I_{j=1} \right) \right]^{2},
$$

where $I_{j=1}$ is an indicator function and $\eta_{j}$ is the frequency of $y$-valued data. The estimators $\omega^*$ and $\lambda^*$ that give the smallest $S_{Y}(\omega^*, \lambda^*)$ are the OLSEs of $\omega$ and $\lambda$, respectively. The resulting estimators have invariance property since OLSE is a special case of the minimum-distance estimation [18].

### 4. Simulation study

Two simulation studies were conducted to evaluate the performance of the proposed estimators in terms of unbiasedness, consistency and efficiency. Both simulation studies used $\lambda > 1$ to ensure that $n_2 > 0$ and $n_3 > 0$, so that the RPEs of $\omega$ and $\lambda$ can be obtained.

#### 4.1. Unbiasedness and consistency properties of the estimators

The first simulation study was conducted to evaluate the unbiasedness and consistency of the proposed estimators. The setting for the first simulation study is as follows:

**Simulation setting:**

**Step 1.** Generate $n = 100, 200, \ldots, 1000$ random data samples that follow a positive Poisson distribution with parameters $\omega$ and $\lambda$, where $\omega = 0.1, 0.3$ and $\lambda = 1.2, 2.0$.

**Step 2.** Change the value of $\omega n$ data to 1 at random to ensure the one-inflation property.

**Step 3.** Estimate the parameters using MME, MLE, RPE and OLSE.

**Step 4.** Repeat Steps 1–3 for a total of 2000 times, and calculate bias $\text{Bias}(\delta')$ and mean squared error $\text{MSE}(\delta')$ for parameter $\delta$ using the following respective formulae

$$
\text{Bias}(\delta') = \frac{1}{2000} \sum_{j=1}^{2000} |\delta - \delta'|/2000
$$

and

$$
\text{MSE}(\delta') = \frac{1}{2000} \sum_{j=1}^{2000} (\delta - \delta')^2/2000,
$$

where $\delta'$ is the estimated $\delta$ and $\delta = \omega, \lambda$. The absolute value of the deviation about the true parameter was considered when computing the bias to provide a smooth trend of bias as $n$ increases and avoid bias value fluctuations.

Figures 2 and 3 respectively illustrate the bias of the proposed estimators of $\lambda$ and $\omega$. Based on Figures 2 and 3, the following conclusions can be drawn:

(i) The estimators are asymptotically unbiased. As $n$ increases, the bias of the estimators of both $\lambda$ and $\omega$ decrease and approach zero.
Figure 2. The absolute relative bias of the estimator of $\lambda$ when estimated using the MLE, MME, OLSE and RPE for $\lambda = 1.2, 2.0$ and $\omega = 0.1, 0.3$.

(ii) The biases of MMEs, MLEs and OLSEs are similar and significantly smaller than those of RPEs for the same sample size and parameters.

(iii) The estimators with the lowest to highest bias for any given sample size are MLE, OLSE, MME and RPE.

Figures 4 and 5 respectively illustrate the mean squared error of the estimators of $\lambda$ and $\omega$ using the proposed estimation methods. Based on Figures 4 and 5, the following conclusions can be drawn:

(i) The estimators are consistent for all parameters. As $n$ increases, the mean squared errors of the estimators decrease and approach zero.

(ii) The mean squared error of MMEs, MLEs and OLSEs are similar and significantly smaller than those of RPEs for the same sample size and parameters.

(iii) The estimators with the lowest to highest mean squared error for any given sample size are MLE, OLSE, MME and RPE.

In short, the proposed estimators are asymptotically unbiased and consistent for all values of parameters. RPE has a significantly larger bias and mean squared error than other estimators, but it still may provide a good model fitting for very large samples ($n \gg 1000$). The best to worst estimators in terms of unbiasedness and consistency are MLE, OLSE, MME and RPE.

4.2. Efficiency property of the estimators

The second simulation study was conducted to evaluate the efficiency of the proposed estimators compared...
Figure 3. The absolute relative bias of the estimator of \( \omega \) when estimated using the MLE, MME, OLSE and RPE for \( \lambda = 1.2, 2.0 \) and \( \omega = 0.1, 0.3 \).

Simulation setting:

Step 1. Generate \( n = 1000 \) random data samples that follow the positive Poisson distribution with parameters \( \omega \) and \( \lambda \).

Step 2. Change the value of \( \omega \) on data to 1 at random to ensure the one-inflation property.

Step 3. Estimate the parameters using MME, MLE, RPE and OLSE.

Step 4. Repeat Steps 1–3 for a total of 2000 times, and calculate the efficiency of all estimators compared to the efficiency of the MLE using

\[
eff(\delta', \hat{\delta}) = \frac{\text{Var}(\hat{\delta})}{\text{Var}(\delta')},
\]

where \( \delta' \) is the MME, RPE and OLSE of \( \delta \), and \( \hat{\delta} \) is the MLE of \( \delta \). Since the asymptotically unbiased property has been established and variance can be represented as the sum of mean squared error and squared bias, the efficiency of the estimators can be calculated using the following formula:

\[
eff(\delta', \hat{\delta}) = \frac{\text{MSE}(\hat{\delta})}{\text{MSE}(\delta')}.
\]

Figure 6 and Figure 7 respectively illustrate the efficiency of the proposed estimators of \( \lambda \) and \( \omega \) in percentage compared to the efficiency of the MLE. Based on Figure 6 and Figure 7, the following conclusions can be drawn:

(i) The MME of \( \lambda \) is approximately 70–80% as efficient as the MLE of \( \lambda \).
(ii) The MME of \( \omega \) is approximately 40–70% as efficient as the MLE of \( \omega \).
### Figure 4. The mean squared error of the estimator of $\lambda$ when estimated using the MLE, MME, OLSE and RPE for $\lambda = 1.2, 2.0$ and $\omega = 0.1, 0.3$.  

(iii) The RPE of $\lambda$ is approximately 5–30% as efficient as the MLE of $\lambda$.  
(iv) The RPE of $\omega$ is approximately 2–25% as efficient as the MLE of $\omega$.  
(v) The OLSEs of $\lambda$ and $\omega$ is approximately 90–99% as efficient as the MLE of $\lambda$ and $\omega$.  

It can be concluded that the OLSE is almost as efficient as the MLE but more efficient than the MME and RPE. Also, the MME is significantly more efficient than the RPE.  

Another way to evaluate estimator performance is by calculating the joint efficiencies of the estimators for the two parameters using the deficiency criterion [19], which is defined as:

$$\text{Def}(\lambda', \omega') = \text{MSE}(\lambda') + \text{MSE}(\omega'),$$

where $\lambda'$ and $\omega'$ are the estimators of $\lambda$ and $\omega$, respectively. Figure 8 shows that the joint efficiency decreases as the sample size grows for given $\lambda$ and $\omega$. The MLE, MME and OLSE have similar joint efficiencies, whereas RPE has a substantially larger joint efficiency, even when the sample size is as large as 1000. According to the deficiency criterion, the MLE is the most efficient estimator. The estimators with the best to worst efficiency are MLE, OLSE, MME and RPE. These findings corroborate the findings in Figure 6 and Figure 7.  

#### 5. One-inflation index under positive Poisson distribution

A novel index called the one-inflation index was developed to determine whether a one-inflated distribution is required to model a dataset. By definition, a $k$-inflation
index is an index that assesses the presence of $k$-valued data with respect to a certain distribution. For instance, the zero-inflation index, denoted as $z_{ip}$, assesses the presence of excess zeros in a dataset for a Poisson distribution [20,21]. Another zero-inflation index, denoted as $z_{inb}$, was introduced to assess the presence of excess zeros in the dataset for a negative binomial distribution [22]. The formulae of $z_{ip}$ and $z_{inb}$ are respectively given as:

$$z_{ip} = 1 + \ln(p_0)/\mu$$

and

$$z_{inb} = 1 + \frac{(\sigma^2 - \mu) \ln(p_0)}{\mu^2 \ln(\sigma^2/\mu)},$$

where $p_0$ is the proportion of zeros, $\mu$ is the mean and $\sigma^2$ is the variance. When the sample zero-inflation index is greater than zero, this indicates that there are excess zeros in the dataset. Following the works in [20–22], a new one-inflation index is proposed to assess the presence of excess ones in the dataset for a positive Poisson distribution. The one-inflation index is denoted as

$$o_{ipp} = 1 + \ln(p_1) - \ln(\exp(d + \mu - 1) - 1) - \ln(d + \mu - 1),$$

where $p_1$ is the proportion of ones and $d$ is the dispersion index. If a random variable follows the positive Poisson distribution, then $o_{ipp} = 0$, while $o_{ipp} > 0$ indicates that the dataset contains a large number of ones for the positive Poisson distribution. The presence of excess ones in the sample data can be determined by computing the sample one-inflation index. As an example, a dataset of size 5000 that follows the OIPP distribution was simulated, and the one-inflation index $\hat{o}_{ipp}$ for a positive Poisson distribution are shown in Table 1.
According to Table 1, the index correctly assesses the presence of excess ones in the sample data. When \( \omega = 0 \), this implies that the index is very close to zero, which indicates that there is no excess of ones in the data for the OIPP distribution since this distribution reduces to a positive Poisson distribution. Trivially, a smaller \( \lambda \) yields a higher proportion of ones in the positive Poisson distribution and a lower proportion of ones contributed by \( \omega \) in the OIPP distribution. Therefore, the resulting index will be comparatively low (refer to Table 1 for \( \lambda = 0.5 \) and all values of \( \omega \)). However, when \( \lambda \) is large, the proportion of ones in the positive Poisson distribution will be small, while the proportion of ones contributed by \( \omega \) in the OIPP distribution will be large. This results in a high index value (refer to Table 1 for \( \lambda = 2.5 \) and positive \( \omega \)). It can be concluded that the one-inflation index is useful to assess the presence of excess ones in a dataset for a positive Poisson distribution.

This one-inflation index can be a viable alternative to the score test proposed by Godwin and Böhning [9], in which the null hypothesis of the test states that there is no inflation in the data. The score test [9] is given as:

\[
S = \frac{(nT - n)\hat{\lambda}(T + 1)}{nT(T - \hat{\lambda})}
\]

where \( T = \exp(\hat{\lambda}) - 1 \) and \( S \sim \chi^2_{(1)} \). The information about whether a dataset contains excess ones or not needs to be considered in statistical modelling. If the dataset contains excess ones, then one-inflated positive count data distributions should be considered.

6. Applications
A dataset on the frequency of a person being arrested for drunk driving [23] was used to demonstrate and
Figure 7. The efficiency of the estimators of $\omega$ when estimated using the MME, RPE and OLSE with respect to the MLE.

Table 1. The sample proportion of one-valued data, sample mean, sample variance, sample dispersion index and sample one-inflation index for the one-inflated positive Poisson distribution for various pairs of parameters $\lambda$ and $\omega$, based on a dataset of size 5000.

| $\lambda$ | $\omega$ | Sample proportion, $\hat{p}_1$ | Sample mean, $\bar{y}$ | Sample variance, $s^2$ | Sample dispersion index, $\hat{d}$ | Sample index, $\hat{oi}_{PP}$ |
|---|---|---|---|---|---|---|
| 0.5 | 0.0 | 0.768 | 1.276 | 0.296 | 0.232 | 0.006 |
| 0.2 | 0.815 | 1.218 | 0.246 | 0.202 | 0.061 |
| 0.4 | 0.866 | 1.163 | 0.198 | 0.171 | 0.158 |
| 0.6 | 0.909 | 1.108 | 0.135 | 0.122 | 0.185 |
| 1.0 | 0.0 | 0.584 | 1.581 | 0.664 | 0.420 | 0.007 |
| 0.2 | 0.664 | 1.464 | 0.574 | 0.392 | 0.108 |
| 0.4 | 0.756 | 1.336 | 0.453 | 0.339 | 0.214 |
| 0.6 | 0.837 | 1.234 | 0.359 | 0.291 | 0.352 |
| 1.5 | 0.0 | 0.432 | 1.936 | 1.135 | 0.586 | 0.019 |
| 0.2 | 0.539 | 1.748 | 1.015 | 0.580 | 0.161 |
| 0.4 | 0.658 | 1.553 | 0.852 | 0.549 | 0.304 |
| 0.6 | 0.767 | 1.383 | 0.666 | 0.482 | 0.428 |
| 2.0 | 0.0 | 0.317 | 2.302 | 1.582 | 0.687 | 0.004 |
| 0.2 | 0.439 | 2.081 | 1.637 | 0.786 | 0.234 |
| 0.4 | 0.589 | 1.788 | 1.369 | 0.766 | 0.395 |
| 0.6 | 0.726 | 1.512 | 1.027 | 0.679 | 0.511 |
| 2.5 | 0.0 | 0.223 | 2.743 | 2.173 | 0.792 | 0.014 |
| 0.2 | 0.384 | 2.369 | 2.154 | 0.910 | 0.289 |
| 0.4 | 0.529 | 2.062 | 2.054 | 0.996 | 0.469 |
| 0.6 | 0.692 | 1.684 | 1.533 | 0.911 | 0.592 |
Figure 8. The joint efficiency of the estimators of $\lambda$ and $\omega$ when estimated using the MLE, MME, OLSE and RPE for $\lambda = 1.2, 2.0$ and $\omega = 0.1, 0.3$.

| $x$ | $n_x$ | MME | MLE | RPE | OLSE |
|-----|-------|-----|-----|-----|------|
| 1   | 8877  | 8880.44 | 8876.98 | 8868.79 | 8876.96 |
| 2   | 481   | 471.05  | 477.33  | 492.38  | 480.36  |
| 3   | 52    | 61.08   | 58.82   | 53.23   | 56.35   |
| 4   | 8     | 5.94    | 5.44    | 4.32    | 4.96    |
| 5+  | 1     | 0.49    | 0.44    | 0.28    | 0.37    |

Estimated parameters

- $\hat{\lambda} = 0.3890$
- $\hat{\omega} = 0.6857$
- RMSE = 6.29
- $\chi^2 = 2.5885$
- $p$-value = 0.1076

The results of model fitting of the frequency of a person being arrested for drunk driving based on the MME, MLE, RPE and OLSE for the OIPP distribution.

Table 2. The results of model fitting of the frequency of a person being arrested for drunk driving based on the MME, MLE, RPE and OLSE for the OIPP distribution.

* the best model is written in bold.

analyse the performance of the proposed estimators in fitting real data. The sample one-inflation index of 0.1545 implies that the dataset contains a substantial number of ones in the data that cannot be explained by a positive Poisson distribution. The model fitting to the data using a positive Poisson distribution yields $\hat{\lambda} = 0.1275$. The estimated value of $\lambda$ generates a score $s = 30.6686$, hence rejecting the null hypothesis of no
inflation. Therefore, the OIPP distribution is appropriate for model fitting for this dataset.

The chi-square goodness-of-fit test and the root mean squared error (RMSE) of the fitted data were used to evaluate the proposed estimators, in which

\[
RMSE = \sqrt{\frac{1}{h} \sum_{x=1}^{h} (n_x - \hat{n}_x)^2 / h},
\]

where \( h \) is the number of data groups and \( \hat{n}_x \) is the estimated frequency for the respective \( n_x \). In general, the best estimator provides adequate goodness of fit and has the lowest RMSE.

The results of the model fitting for the frequency of a person being arrested for drunk driving based on the proposed estimators are shown in Table 2. Based on the goodness-of-fit test and \( p \)-values, all estimators are found to be adequate except for RPE. The RMSE of the OLSE is found to be significantly smaller compared to other estimators. Therefore, the OLSE provide the best fit for describing the frequency of a person being arrested for drunk driving.

7. Conclusions

Several estimators for the OIPP distribution were proposed in this study, namely MME, MLE, RPE, OPE and OLSE, in which the MLE and OPE both yield identical results. Two comprehensive simulation studies were conducted to evaluate the unbiasedness, consistency and efficiency of the proposed estimators. According to the results, the proposed estimators are asymptotically unbiased and consistent. The best to worst estimators based on the bias, mean squared error, efficiency and deficiency values are MLE, OLSE, MME and RPE. A one-inflation index to assess the presence of excess ones in a dataset for a positive Poisson distribution was also proposed. Based on the chi-square goodness-of-fit test and RMSE, the OLSE provides the best fit to the dataset adopted in this study. Therefore, both MLE and OLSE are the best estimators for the OIPP distribution.

Acknowledgements

The authors gratefully acknowledge the financial support received in the form of research grants from the Ministry of Education, Malaysia [FRGS/1/2019/STG06/UKM/01/5]; and Universiti Kebangsaan Malaysia [GUP-2019-031].

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Ministry of Education, Malaysia: [Grant Number FRGS/1/2019/STG06/UKM/01/5]; Universiti Kebangsaan Malaysia: [Grant Number GUP-2019-031].

ORCID

Razik Ridzuan Mohd Tajuddin  
http://orcid.org/0000-0001-6534-3678

References

[1] David FN, Johnson NL. The truncated Poisson. Biometrics. 1952;8:275–285.
[2] Moore PG. The estimation of the Poisson parameter from a truncated distribution. Biometrika. 1952;39:247–251.
[3] Plackett RL. The truncated Poisson distribution. Biometrics. 1953;9:485–488.
[4] Finney DJ, Varley GC. The truncated Poisson distribution. Biometrics. 1955;11:387–394.
[5] Tate RF, Goen RL. Minimum variance unbiased estimation for the truncated Poisson distribution. Ann Math Stat. 1958;29:755–765.
[6] Sampford MR. The truncated negative binomial distribution. Biometrika. 1955;42:58–69.
[7] Brass W. Simplified methods of fitting the truncated negative binomial distribution. Biometrika. 1958;45:59–68.
[8] Ghitany ME, Al-Mutairi DK, Nadarajah S. Zero-truncated Poisson-Lindley distribution and its application. Math Comput Simul. 2008;79:279–287.
[9] Godwin RT, Böhning D. Estimation of the population size by using the one-inflated positive Poisson model. J R Stat Soc: Ser C. 2017;66:425–448.
[10] Godwin RT. One-inflation and unobserved heterogeneity in population size estimation. Biometrical J. 2017;59:79–93.
[11] Godwin R. The one-inflated positive Poisson mixture model for use in population size estimation. Biometrical J. 2019;61:1541–1556.
[12] Tajuddin, R. R. M., Ismail, N., Kamarulzaman, I., 2021. Estimating population size of criminals: a new Horvitz-Thompson estimator under one-inflated positive Poisson-Lindley model. Crime & Delinquency. doi:10.1177%2F0112872111014158
[13] Famoye F. Parameter estimation for generalized negative binomial distribution. Commun Stat Simul Comput. 1997;26:269–279.
[14] Wagh YS, Kamalja KK. Zero-inflated models and estimation in zero-inflated Poisson distribution. Commun Stat – Simu Comput. 2018;47:2248–2265.
[15] Hogg RV, McKean JW, Craig AT. Maximum likelihood estimation, in Introduction to mathematical statistics, 6th ed. Pearson Prentice Hall, New Jersey, pp. 313, 2005.
[16] Böhning D. A simple variance formula for population size estimators by conditioning. Stat Methodol. 2006;3:410–423.
[17] Drossos CA, Philippou AN. A note on minimum distance estimates. Ann I Stat Math. 1980;32:121–123.
[18] Akgül FG, Şenoğlu B, Arslan T. An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. Energy Convers Manage. 2016;114:234–240.
[19] Puig P. Characterizing additively closed discrete models by a property of their maximum likelihood estimators, with an application to generalized Hermite distributions. J Am Stat Assoc. 2003;98:687–692.
[20] Puig P, Valero J. Count data distributions. J Am Stat Assoc. 2006;101:332–340.
[21] Blasco-Moreno A, Pérez-Casany M, Puig P, et al. What does a zero mean? Understanding false, random and structural zeros in ecology. Methods Ecol Evol. 2019;10:949–959.
[22] Van der Heijden PGM, Cruyff M, Van Houwelingen HC. Estimating the size of a criminal population from police records using the truncated Poisson regression model. Statistica Neerlandica. 2003;57:289–304.