Thailand Statistician
April 2022; 20(2): 357-371
http://statassoc.or.th
Contributed paper

On Interval Estimation of the Poisson Parameter in a Zero-inflated Poisson Distribution

Patchanok Srisuradetchai* and Kittanan Tonprasongrat
Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Rangsit Campus, Pathumthani, Thailand.
*Corresponding author; e-mail: patchanok@mathstat.sci.tu.ac.th

Received: 23 September 2020
Revised: 8 February 2021
Accepted: 21 February 2021

Abstract

The zero-inflated Poisson distribution is the most widely applied model for count data with excessive zeros. In this paper, confidence intervals for the Poisson parameter are derived by using score statistics. The resulting intervals are compared to the Wald confidence interval (WCI), which stems from properties of the asymptotic normal distribution. For interval constructions, a Bernoulli parameter, known as a nuisance parameter, is eliminated by the profile likelihood approach. The Wald-type intervals can be formulated explicitly, while the score intervals have no closed forms. Furthermore, the observed and expected Fisher information matrices are shown to be the same. Using a simulation study, the confidence intervals are compared in many situations where the Poisson and Bernoulli parameters and sample sizes are varied. The coverage probability (CP), average length, and coverage per unit length (CPUL) are obtained from Monte Carlo methods. The results reveal that the score confidence intervals are superior to the WCIs in an aspect of CP with small sample sizes, but all of these intervals are comparable in terms of CPUL.

Keywords: Score interval, Wald interval, profile likelihood, Monte Carlo simulations.

1. Introduction

The Poisson random variable is a discrete variable with a probability mass function (p.m.f.) that is defined as $f(x;\lambda)=e^{-\lambda}\lambda^{x}/x!$, where $x=0,1,2,...$ and $\lambda>0$. The p.m.f. expresses the probability of the number of interesting events occurring in a certain period or space, presuming that these events occur at a certain rate λ and independently of the time (Cameron and Trivedi 2013). The Poisson random variable can be found in many fields related to counting, such as the number of mutations on a strand of DNA per unit length (Balin and Cascalho 2010), the number of auto insurance claims occurring in a given period of time (David and Jemna 2015), and the total number of calls made in a given time interval (Ibrahim et al. 2016). The Poisson distribution can also be used to approximate the binomial distribution when the number of observations is large and the probability of the event of interest is small. Thus, it can be seen that this distribution is widely used for counting data that is commonly found in real-life situations.

As is well-known, the expected value and variance of a Poisson distribution are equal, which is described as "equi-dispersed". The data that is Poisson-like but with a more excessive number of zeros will lack the equidispersion property as its variance exceeds the expected value. Thus, this reduces the usefulness of the Poisson distribution. In this case, the zero-inflated Poisson (ZIP) distribution, which is a modification of the regular Poisson distribution, becomes beneficial. Ridout et al. (1998) pointed out that it is, in practice, possible to have fewer zero counts than expected, but this is a much less common occurrence. Note that not every data set with many zeros will lead to the ZIP distribution, as the regular Poisson distribution with an exceptionally low value of mean can also produce many zeros. Regarding such a case, Tlhaloganyang and Sakia (2020) showed that the standard Poisson process performed better than the ZIP distribution; so conducting the overdispersion testing before selecting the fitted distribution is suggested.

The ZIP distribution is considered as a convex combination of two subpopulations: a degenerating distribution at zero and a regular Poisson distribution. Its p.m.f. is defined as (1).

$$f(x;\lambda,\pi)=\begin{cases}\pi+(1-\pi)e^{-\lambda} & ,x=0\\ (1-\pi)e^{-\lambda}\lambda^x/x!, & x=1,2,\ldots\end{cases}=\left[\pi+(1-\pi)e^{-\lambda}\right]^{I_{\{0\}}(x)}\left[(1-\pi)e^{-\lambda}\lambda^x/x!\right]^{I_{\{1,2,\ldots\}}(x)}, \tag{1}$$

where $I_A(x)$ is an indicator function whose value equals 1 if $x\in A$ and 0 otherwise. The mean and variance of the distribution with the p.m.f. in (1) are $(1-\pi)\lambda$ and $\lambda(1-\pi)(1+\pi\lambda)$, respectively (Tlhaloganyang and Sakia 2020). Suppose that $X_1,\ldots,X_n$ are a random sample of size n from the ZIP, N_0 denotes the number of zeros, which is also a random variable, and $E(N_0)$ is equal to $n\left[\pi+(1-\pi)e^{-\lambda}\right]$. The mixing parameter π is usually unknown and is referred to as the inflation parameter at zero, giving the ZIP model more flexibility than the regular Poisson distribution (Unhapipat 2018). Many publications have employed the ZIP model for fitting data with many zeros. Xu et al. (2014) showed that the ZIP model gave a higher accuracy than the Poisson model for urinary tract infection (UTI) data. Beckett et al. (2014) modeled data regarding several natural calamities. Sarul and Sahin (2015) applied the ZIP distribution to claim frequencies for the automobile portfolios of a Turkish insurance company that occurred between 2012 and 2014, and the ZIP model was found to be superior to the standard Poisson model.

Most papers have focused on the inference in regression models, but some studies were devoted to parameter estimations of the ZIP distribution. Lambert (1992) demonstrated the log-likelihood function of the ZIP regression and described the steps of the EM algorithm to estimate regression coefficients. Lee et al. (2001) added the individual exposure variable to the ZIP distribution and derived Fisher information (FI). DeGroot and Schervish (2018) mentioned only one kind of Fisher information. To distinguish it from the other kind, $I(\lambda,\pi)$ in this paper is called observed FI while the other is called expected FI. Beckett et al. (2014) used the method of moment (MME) to find the estimators of π and λ: $\hat{\lambda}_{MM}=\bar{X}+S^2/\bar{X}-1$ and $\hat{\pi}_{MM}=\left[S^2-\bar{X}\right]/\left[\bar{X}^2+(S^2-\bar{X})\right]$, where $\bar{X}=\sum_{i=1}^{n}X_i/n$ and $S^2=\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/(n-1)$. Defining n_0 (the value of random variable N_0) as the number of x_i's taking the value 0, the maximum likelihood estimators satisfy the following system of equations

$$\hat{\pi}_{ML}+(1-\hat{\pi}_{ML})e^{-\hat{\lambda}_{ML}}=n_0/n \text{ and } (1-\hat{\pi}_{ML})\hat{\lambda}_{ML}=\bar{x} \tag{2}$$

(Waguespack et al. 2020). Accordingly, there are no explicit forms of $\hat{\pi}_{ML}$ and $\hat{\lambda}_{ML}$. Vandenbroek (1995) derived the score test statistic to conclude whether the number of zeros is too large for a standard Poisson distribution to fit the data, i.e., $H_0 : \pi = 0.$ Xie et al. (2001) discussed six test statistics for testing whether a parameter π is equal to zero, and one of these is the score test proposed by Vandenbroek (1995). Similarly, Numna (2009) developed a Wald statistic for testing $\pi.$ Paneru et al. (2018) applied the bootstrap method to compute the confidence intervals of a zero-inflated population mean by giving an example of normal distribution. Schwartz and Giles (2016) provided the bias-adjusted maximum likelihood estimators for both parameters of ZIP in the large-sample situations where the smallest sample is n = 50. The proposed estimators were compared to the parametric bootstrap bias-adjusted estimators. Schwartz and Giles (2016) also commented that when the sample size is equal to 10, some estimates are negative. Unhapipat et al. (2018) derived the predictive distributions of the ZIP distribution based on the generalized and Jeffrey's noninformative priors. Wagh and Kamalja (2018) introduced a new estimator, referred to as a probability estimator (PE) of the inflation parameter of the ZIP distribution, i.e., parameter π based on a moment estimator (ME) of the mean parameter, and compared its performance with that of ME and the maximum likelihood estimator (MLE) through a simulation study. Sakthivel and Rajitha (2018) proposed the probability-based inflation estimator (PBIE) for parameter $\pi,$ and the performance of the proposed estimator was assessed by means of a simulation approach. Through the simulation study, Srisuradetchai and Junnumtuam (2020) studied the Wald confidence intervals for parameter π of the ZIP and zero-altered Poisson (ZAP) models and investigated the effects of the model choices between ZIP and ZAP with three different link functions: logit, probit, and complementary loglog.

The other parameter of ZIP distribution is $\lambda,$ but it has rarely been studied in the literature. It is known that the maximum likelihood is commonly used to find the point estimators for λ and $\pi.$ However, interval estimations for λ have not been identified. In this paper, the Wald confidence intervals using both observed and expected Fisher information matrices (FIMs) will be calculated. Furthermore, as π is assumed to be unknown, the profile likelihood approach is used to eliminate this nuisance parameter, and therefore, the score and Wald confidence intervals are also constructed from the profile likelihood.

2. Wald intervals

From the frequentist viewpoint, a $100(1-\alpha)\%$ confidence interval for θ will satisfy the following property

$$P\left[\theta_L\left(T_n\right) \le \theta \le \theta_U\left(T_n\right)\right] = 1-\alpha, \tag{3}$$

where $1-\alpha$ is a confidence coefficient, the statistics $\theta_L(T_n)$ and $\theta_U(T_n)$ are the limits of the confidence interval, and $\theta_L(T_n) \le \theta_U(T_n).$ The interval in (3) is a random interval, and when an observed value, $t_{\text{obs}},$ is substituted for $T_n,$ the resulting interval is no longer a random variable (Rohde 2014).

We will start with the log likelihood function in which $X = \left(X_1, X_2, \dots, X_n\right)$ is a random sample of size n from (1). This will have a form of

$$\log L\left(\lambda,\pi;x\right) = n_0 \log\left[\pi + (1-\pi)e^{-\lambda}\right] + \left(n-n_0\right)\log\left(1-\pi\right) - \lambda\left(n-n_0\right) + \sum_{i=1,x_i\neq 0}^{n} x_i \log\lambda - \log \prod_{i=1,x_i\neq 0}^{n} x_i\,!, \tag{4}$$

where $n_0 = \sum_{i=1}^{n} I_{\{0\}}(x_i)$ or the number of zeros, and $(n-n_0) = \sum_{i=1}^{n} I_{\{1,2,\ldots\}}(x_i)$ or the number of x_i having a value that is a positive integer. The plot of the relative log likelihood, $\log L(\lambda,\pi;x)/\max \log L(\lambda,\pi;x)$, is illustrated in Figure 1. It shows that the shapes of the likelihood contour can be noticeably different, depending upon the parameters. In the right figure, the contour shape largely differs from an ellipse, so it can severely affect the asymptotic normality assumption of an MLE in situations where a sample size is small. This can cause the Wald confidence interval (WCI) to have poor performance.

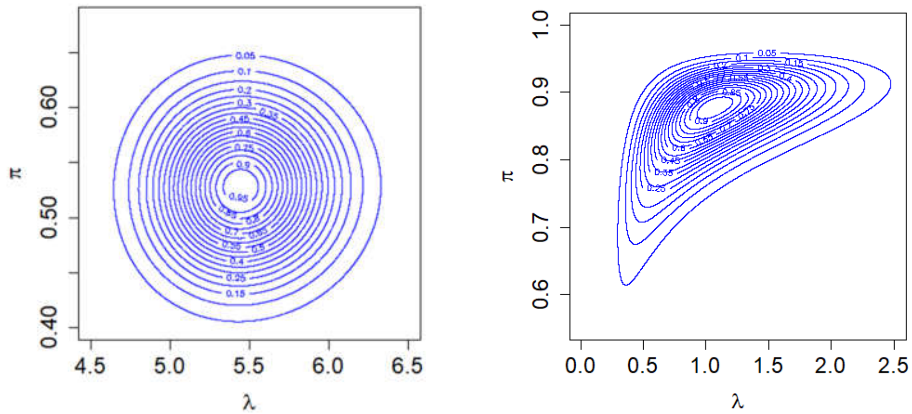

Figure 1 Contour plots of the relative likelihood functions of (λ,π) based on a sample of size 100 from (left) ZIP$(\lambda=5,\pi=0.5)$ and (right) ZIP$(\lambda=1,\ \pi=0.9)$.

2.1. Standard Wald confidence interval

A key component of Wald intervals is Fisher information (FI), which will be described here. Given $x=(x_1,x_2,\ldots,x_n)$, the observed FIM, $I(\lambda,\pi)$, is a matrix with the element that is the negative second derivative of the likelihood. For the ZIP, $I(\lambda,\pi)$ is already known from Numna (2009) that

$$I(\lambda,\pi)=\begin{bmatrix} I_{11}=\dfrac{\sum_{i=1}^{n} x_i}{\lambda^2}-\dfrac{n_0\pi(1-\pi)e^{-\lambda}}{\left[\pi+(1-\pi)e^{-\lambda}\right]^2} & I_{12}=\dfrac{-n_0e^{-\lambda}}{\left[\pi+(1-\pi)e^{-\lambda}\right]^2} \\ I_{12}=\dfrac{-n_0e^{-\lambda}}{\left[\pi+(1-\pi)e^{-\lambda}\right]^2} & I_{22}=\dfrac{n-n_0}{(1-\pi)^2}+\dfrac{n_0\left(1-e^{-\lambda}\right)^2}{\left[\pi+(1-\pi)e^{-\lambda}\right]^2} \end{bmatrix}, \tag{5}$$

where I_{11}, I_{12} and I_{22} are calculated by taking the second derivative of the log likelihood function $-\frac{\partial^2}{\partial\lambda^2}\log L(\lambda,\pi;x)$, $-\frac{\partial^2}{\partial\lambda\partial\pi}\log L(\lambda,\pi;x)$, and $-\frac{\partial^2}{\partial\pi^2}\log L(\lambda,\pi;x)$, respectively. In our paper, the expected FIM is derived by taking the expectation to matrix $I(\lambda,\pi)$, (5), resulting in

$$J(\lambda,\pi)=\begin{bmatrix} n\left[\dfrac{1-\pi}{\lambda}-\dfrac{\pi(1-\pi)e^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}}\right] & \dfrac{-ne^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}} \\ \dfrac{-ne^{-\lambda}}{\pi+(1-\pi)e^{-\lambda}} & \dfrac{n\left(1-e^{-\lambda}\right)}{\left[\pi+(1-\pi)e^{-\lambda}\right](1-\pi)} \end{bmatrix}=\begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix}. \tag{6}$$

It is not difficult to show that $I(\hat{\lambda}_{ML},\hat{\pi}_{ML})$ and $J(\hat{\lambda}_{ML},\hat{\pi}_{ML})$ have the same value. Firstly, consider the differences between I_{ij} and J_{ij}, $i,j=1,2$ in the following

$$I_{11}-J_{11}=\frac{\sum_{i=1}^{n}x_i-n\hat{\lambda}_{ML}\left(1-\hat{\pi}_{ML}\right)}{\left(\hat{\lambda}_{ML}\right)^2}+\frac{\hat{\pi}_{ML}\left(1-\hat{\pi}_{ML}\right)e^{-\hat{\lambda}_{ML}}}{P^2}\left(n-\frac{n_0}{P}\right),\ I_{12}-J_{12}=\frac{ne^{-\hat{\lambda}_{ML}}}{P}-\frac{n_0e^{-\hat{\lambda}_{ML}}}{P^2},\text{ and}$$

$$I_{22}-J_{22}=\frac{n-n_0}{\left(1-\hat{\pi}_{ML}\right)^2}+\frac{n_0\left(1-e^{-\hat{\lambda}_{ML}}\right)^2}{P^2}-\frac{n\left(1-e^{-\hat{\lambda}_{ML}}\right)}{P\left(1-\hat{\pi}_{ML}\right)},$$

where $P=\left[\hat{\pi}_{ML}+\left(1-\hat{\pi}_{ML}\right)e^{-\hat{\lambda}_{ML}}\right]$. From (2), it arrives at $1-e^{-\hat{\lambda}_{ML}}=\hat{\lambda}_{ML}\left(n-n_0\right)/n\overline{x}$ and $P=n_0/n$. Substitute them into $I_{ij}-J_{ij}$ and they will all be zeros. Note that $I(\lambda,\pi)$ and $J(\lambda,\pi)$ are different, but the estimated FIMs are the same, i.e. $I(\hat{\lambda}_{ML},\hat{\pi}_{ML})=J(\hat{\lambda}_{ML},\hat{\pi}_{ML})$.

For testing the hypothesis $H_0:\lambda=\lambda_0$ versus $H_1:\lambda\neq\lambda_0$, the Wald statistic is the squared difference of $\hat{\lambda}_{ML}-\lambda_0$, which is weighed by the curvature of the log-likelihood function. It has the form of

$$W=\frac{\left(\hat{\lambda}_{ML}-\lambda_0\right)^2}{\operatorname{var}(\hat{\lambda}_{ML})}\sim\chi_1^2. \tag{7}$$

The statistic $\sqrt{W}$ follows an asymptotic standard normal distribution. The value of var($\hat{\lambda}_{ML}$) will be estimated by $J^{11}(\hat{\lambda}_{ML},\hat{\pi}_{ML})$, which is the element in the first row and the first column of the inverse matrix of $J(\hat{\lambda}_{ML},\hat{\pi}_{ML})$. Because $J^{11}(\lambda,\pi)=J_{22}/\left(J_{11}J_{22}-J_{12}^2\right)$, consider

$$\begin{aligned}J_{11}J_{22}&=n\left[\frac{1-\pi}{\lambda}-\frac{\pi\left(1-\pi\right)e^{-\lambda}}{\pi+\left(1-\pi\right)e^{-\lambda}}\right]\frac{n\left(1-e^{-\lambda}\right)}{\left[\pi+\left(1-\pi\right)e^{-\lambda}\right]\left(1-\pi\right)}\\&=\frac{n^2\pi+n^2e^{-\lambda}-2n^2\pi e^{-\lambda}-n^2e^{-2\lambda}+n^2\pi e^{-2\lambda}-n^2\pi\lambda e^{-\lambda}+n^2\pi\lambda e^{-2\lambda}}{\lambda\left[\pi+\left(1-\pi\right)e^{-\lambda}\right]^2}.\end{aligned} \tag{8}$$

Subtracting J_{12}^2 from (8), the following is obtained.

$$\begin{aligned}J_{11}J_{22}-J_{12}^2&=\frac{n^2\pi+n^2e^{-\lambda}-2n^2\pi e^{-\lambda}-n^2e^{-2\lambda}+n^2\pi e^{-2\lambda}-n^2\pi\lambda e^{-\lambda}+n^2\pi\lambda e^{-2\lambda}-n^2\lambda e^{-2\lambda}}{\lambda\left[\pi+\left(1-\pi\right)e^{-\lambda}\right]^2}\\&=\frac{n^2\pi\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)+\left(n^2-n^2\pi\right)e^{-\lambda}\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)}{\lambda\left[\pi+\left(1-\pi\right)e^{-\lambda}\right]^2}=\frac{\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)n^2}{\lambda\left[\pi+\left(1-\pi\right)e^{-\lambda}\right]}.\end{aligned}$$

Thus,

$$J^{11}(\lambda,\pi)=\frac{J_{22}}{J_{11}J_{22}-J_{12}^2}=\frac{\lambda\left(1-e^{-\lambda}\right)}{n\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)\left(1-\pi\right)}. \tag{9}$$

Consequently, the $(1-\alpha)100\%$ Wald confidence interval for λ can be simplified as follows

$$\hat{\lambda}_{ML} \pm z_{1-\alpha/2}\sqrt{J^{11}(\hat{\lambda}_{ML},\hat{\pi}_{ML})} = \hat{\lambda}_{ML} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}_{ML}\left(1-e^{-\hat{\lambda}_{ML}}\right)}{n\left(1-e^{-\hat{\lambda}_{ML}}-\hat{\lambda}_{ML}e^{-\hat{\lambda}_{ML}}\right)\left(1-\hat{\pi}_{ML}\right)}}, \tag{10}$$

where $\hat{\lambda}_{ML}$ and $\hat{\pi}_{ML}$ are the maximum likelihood estimators.

2.2. Wald confidence intervals using profile likelihood

The profile likelihood functions behave like "ordinary" likelihoods in that they can be expanded in quadratic terms, but the profile likelihoods have only the parameters of interest because the nuisance parameters have been profiled out (Murphy and Van Der Vaart 2000). To find the profile likelihood function of λ, let parameter λ be a fixed value and first find the π that maximizes $\log L(\lambda,\pi;x)$, (4). The resulting π is $\tilde{\pi}_P = \frac{n_0 - ne^{-\lambda}}{n(1-e^{-\lambda})}$, and substitute π in (4) with $\tilde{\pi}_P$ to find the log profile likelihood as the following:

$$\log L_P\left(\lambda,\tilde{\pi}_p;x\right)$$
$$= n_0\log\left[\tilde{\pi}_P + (1-\tilde{\pi}_P)e^{-\lambda}\right] + (n-n_0)\log(1-\tilde{\pi}_P) - \lambda(n-n_0) + \sum_{i=1}^{n} x_i \log\lambda - \log\prod_{i=1,x_i\neq 0}^{n} x_i!$$
$$= n_0\log\left[\left(\frac{n_0-ne^{-\lambda}}{n-ne^{-\lambda}}\right) + \left(\frac{n-n_0}{n-ne^{-\lambda}}\right)e^{-\lambda}\right] + (n-n_0)\log\left[\frac{n-n_0}{n-ne^{-\lambda}}\right] - \lambda(n-n_0) + \sum_{i=1}^{n} x_i\log\lambda + c$$
$$= n_0\log\left(\frac{n_0}{n}\right) + (n-n_0)\log\left(\frac{n-n_0}{n-ne^{-\lambda}}\right) - \lambda(n-n_0) + \sum_{i=1}^{n} x_i\log\lambda + c, \tag{11}$$

where $c = -\log\prod_{i=1,x_i\neq 0}^{n} x_i!$. With the same data that produced Figure 1, the corresponding relative profile likelihood functions are illustrated in Figure 2. If the shape of the contour likelihood does not appear to be a circle or an ellipse, the profile likelihood will present an asymmetrical curve (right panel in Figure 2).

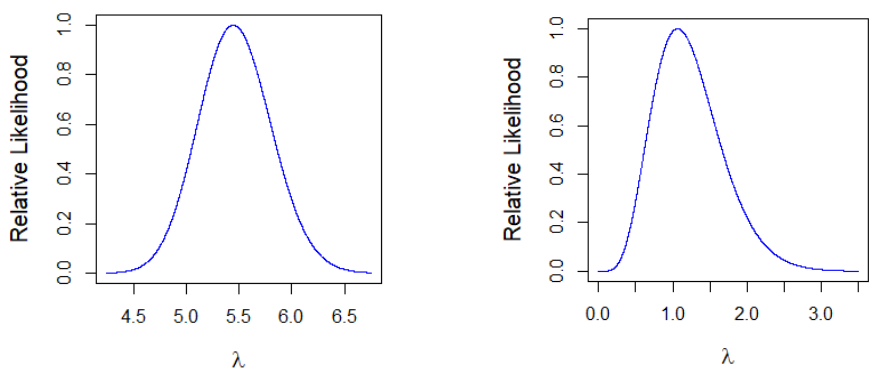

Figure 2 Relative profile likelihood functions of λ given that $x=(x_1,\ldots,x_{100})$ is sampled from (left) ZIP($\lambda=5,\pi=0.5$) and (right) ZIP($\lambda=1,\pi=0.9$).

The λ that maximizes (11) can be acquired by solving the equation $\frac{\partial}{\partial\lambda}\log L_P\left(\lambda,\tilde{\pi}_P;x\right)=0$ as the following

$$\frac{-\left(n-n_0\right)\left(ne^{-\lambda}\right)}{n-ne^{-\lambda}}-\left(n-n_0\right)+\frac{\sum_{i=1,x_i\neq 0}^{n} x_i}{\lambda}=0$$

$$\lambda\left(n-n_0\right)e^{-\lambda}+\left(1-e^{-\lambda}\right)\left(\lambda(n-n_0)-\sum_{i=1,x_i\neq 0}^{n} x_i\right)=0. \tag{12}$$

From the above equations, it can be further simplified and leads to a profile maximum likelihood estimator (PMLE) of λ or $\hat{\lambda}_P$, which can be obtained from

$$\left(n-n_0\right)\hat{\lambda}_P-\sum_{i=1}^{n} x_i\left(1-e^{-\hat{\lambda}_P}\right)=0. \tag{13}$$

The root of this nonlinear equation, $\hat{\lambda}_P$, can simply be solved by R program, using the uniroot function, based on the algorithm proposed by Brent (2013). In addition, it is observed that Equations (13) and (2) are equivalent, and thus $\hat{\lambda}_P=\hat{\lambda}_{ML}$. This leads to the fact that $L(\hat{\lambda}_{ML},\hat{\pi}_{ML};x)=L(\hat{\lambda}_P,\tilde{\pi}_P;x)$ as well as the corresponding FIs. Prior to arriving at a significant conclusion, it is necessary to identify both observed and expected FIs for future use.

Consider the observed FI calculated from the profile likelihood

$$I_P\left(\lambda\right)=-\frac{\partial^2}{\partial\lambda^2}\log L_P\left(\lambda,\tilde{\pi}_P;x\right)=\frac{\partial}{\partial\lambda}\left[\frac{\left(n-n_0\right)\left(ne^{-\lambda}\right)}{n-ne^{-\lambda}}+\left(n-n_0\right)-\sum_{i=1,x_i\neq 0}^{n} x_i\Big/\lambda\right]$$

$$=\frac{n_0e^{-\lambda}-ne^{-\lambda}}{1-2e^{-\lambda}+e^{-2\lambda}}+\frac{\sum_{i=1}^{n}x_i}{\lambda^2}=\frac{\left(1-e^{-\lambda}\right)^2\sum_{i=1}^{n}x_i-\left(n-n_0\right)\lambda^2e^{-\lambda}}{\left[\lambda\left(1-e^{-\lambda}\right)\right]^2}. \tag{14}$$

Accordingly, a $(1-\alpha)100\%$ Wald confidence interval of λ using the profile likelihood and observed FI will be

$$\hat{\lambda}_P\pm z_{1-\alpha/2}\sqrt{I_P^{-1}(\hat{\lambda}_P)}=\hat{\lambda}_P\pm z_{1-\alpha/2}\frac{\hat{\lambda}_P\left(1-e^{-\hat{\lambda}_P}\right)}{\sqrt{\left(1-e^{-\hat{\lambda}_P}\right)^2\sum_{i=1}^{n}x_i-\left(n-n_0\right)\hat{\lambda}_P^2e^{-\hat{\lambda}_P}}}. \tag{15}$$

The expectation of $I_P\left(\lambda\right)$ in (13) will give the expected FI:

$$J_P(\lambda)=E\left[\frac{\left(1-e^{-\lambda}\right)^2\sum_{i=1}^{n}X_i-\left(n-N_0\right)\lambda^2e^{-\lambda}}{\lambda^2\left(1-e^{-\lambda}\right)^2}\right]=E\left[\frac{\sum_{i=1}^{n}X_i}{\lambda^2}-\frac{\left(n-N_0\right)e^{-\lambda}}{\left(1-e^{-\lambda}\right)^2}\right]$$

$$=\frac{ne^{-\lambda}\left(\pi+\left(1-\pi\right)e^{-\lambda}\right)-ne^{-\lambda}}{\left(1-e^{-\lambda}\right)^2}+\frac{n\left(1-\pi\right)\lambda}{\lambda^2}=\frac{n\left(1-\pi\right)\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)}{\lambda\left(1-e^{-\lambda}\right)}. \tag{16}$$

Substitute π in (15) with $\tilde{\pi}_P=\frac{n_0-ne^{-\lambda}}{n-ne^{-\lambda}}$, and $J_P(\lambda)$ will be simplified to

$$J_P(\lambda)=\frac{(n-n_0)\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)}{\lambda\left(1-e^{-\lambda}\right)^2}. \tag{17}$$

Hence, a $(1-\alpha)100\%$ Wald confidence interval using the profile likelihood and expected FI will be

$$\hat{\lambda}_P \pm z_{1-\alpha/2}\sqrt{J^{-1}(\hat{\lambda}_P)}=\hat{\lambda}_P \pm z_{1-\alpha/2}\left(1-e^{-\hat{\lambda}_P}\right)\sqrt{\frac{\hat{\lambda}_P}{(n-n_0)\left(1-e^{-\hat{\lambda}_P}-\hat{\lambda}_P e^{-\hat{\lambda}_P}\right)}}. \tag{18}$$

Furthermore, it can be shown that the estimates of observed and expected FI are the same. Consider the following:

$$\begin{aligned} I_P(\hat{\lambda}_P)-J_P(\hat{\lambda}_P)&=\frac{\left(1-e^{-\hat{\lambda}_P}\right)^2\sum_{i=1}^n x_i-(n-n_0)\hat{\lambda}_P^2 e^{-\lambda}-\hat{\lambda}_P(n-n_0)\left(1-e^{-\hat{\lambda}_P}-\hat{\lambda}_P e^{-\hat{\lambda}_P}\right)}{\left[\hat{\lambda}_P\left(1-e^{-\hat{\lambda}_P}\right)\right]^2} \\ &=\frac{\left(1-e^{-\hat{\lambda}_P}\right)^2\sum_{i=1}^n x_i-(n-n_0)\lambda\left(1-e^{-\hat{\lambda}_P}\right)}{\left[\hat{\lambda}_P\left(1-e^{-\hat{\lambda}_P}\right)\right]^2}=\frac{\left(1-e^{-\hat{\lambda}_P}\right)\sum_{i=1}^n x_i-(n-n_0)\hat{\lambda}_P}{\hat{\lambda}_P^{\,2}\left(1-e^{-\hat{\lambda}_P}\right)}. \end{aligned} \tag{19}$$

Because the nominator $\left(1-e^{-\hat{\lambda}_P}\right)\sum_{i=1}^n x_i-(n-n_0)\hat{\lambda}_P$ in (19) equals (13), the observed and expected FI values which are evaluated at the MLE are the same, i.e. $I_P(\hat{\lambda}_P)-J_P(\hat{\lambda}_P)=0.$ This implies that the Wald confidence intervals using $I_P(\lambda)$ and $J_P(\lambda)$ have the same upper and lower limits. It should be emphasized that $I_P(\lambda)$ and $J_P(\lambda)$ are not the same function, but $I_P(\hat{\lambda}_P)$ equals $J_P(\hat{\lambda}_P)$.

In general, the estimated values of $J^{-1}(\hat{\lambda}_P)$ in (18) and $J^{11}(\hat{\lambda}_{ML},\hat{\pi}_{ML})$ in (10) are not necessarily equal; however, in this work, the estimated values are found to be the same. This leads to the fact that the following four values are equal: $J^{-1}(\hat{\lambda}_P), I^{-1}(\hat{\lambda}_P)$, $J^{11}(\hat{\lambda}_{ML},\hat{\pi}_{ML})$, $I^{11}(\hat{\lambda}_{ML},\hat{\pi}_{ML})$. Consequently, WCIs using either the joint likelihood or profile likelihood are equivalent. However, the score intervals to be discussed in the next section do not employ the estimate of $J(\lambda)$ or $I(\lambda)$ like the WCIs. As a result, Wald and score intervals are different in both mathematical and simulation results.

3. Score Confidence Intervals Using Profile Likelihood

The score function of λ, $S_P(\lambda)$, is the first derivative of the log profile likelihood function or (11). Suppose the object of interest is the null hypothesis $H_0:\lambda=\lambda_0$ with the two-sided alternative $H_a:\lambda\neq\lambda_0$, and if the null hypothesis is true, in large samples the score test will follow the reference χ_1^2 distribution. Let $X=(X_1,X_2,...,X_n)$ be a random sample from ZIP(λ,π), if the profile likelihood is employed, the score statistic will be

$$\frac{S_P(\lambda)}{\sqrt{I_P(\lambda)}}\overset{a}{\sim}N(0,1) \text{ and } \frac{S_P(\lambda)}{\sqrt{J_P(\lambda)}}\overset{a}{\sim}N(0,1). \tag{20}$$

From (12), it can be further simplified as

$$S_P(\lambda)=\frac{-(n-n_0)e^{-\lambda}}{1-e^{-\lambda}}-(n-n_0)+\frac{\sum x_i}{\lambda}=\frac{\sum_{i=1}^{n}x_i\left(1-e^{-\lambda}\right)-(n-n_0)\lambda}{\left(1-e^{-\lambda}\right)\lambda}.$$

From $I_P(\lambda)$, (14), the score statistic defined in (20) will become

$$\frac{S_P(\lambda)}{\sqrt{I_P(\lambda)}}=\frac{\sum_{i=1}^{n}x_i\left(1-e^{-\lambda}\right)-(n-n_0)\lambda}{\sqrt{\left(1-e^{-\lambda}\right)^2\sum_{i=1}^{n}x_i-(n-n_0)\lambda^2e^{-\lambda}}}$$

$$=\frac{\sum_{i=1}^{n}x_i\left(1-e^{-\lambda}\right)-(n-n_0)\lambda}{\sqrt{-(n-n_0)\lambda^2e^{-\lambda}+\sum_{i=1}^{n}x_i-2e^{-\lambda}\sum_{i=1}^{n}x_i+e^{-2\lambda}\sum_{i=1}^{n}x_i}}. \tag{21}$$

To construct the score confidence intervals using the profile and observed FI, from (21), inequality $\left|S_P(\lambda)/\sqrt{I_P(\lambda)}\right|<z_{1-\alpha/2}$ can be further simplified into the following

$$C_1\lambda^2+C_2\lambda\left(e^{-\lambda}-1\right)+C_3\lambda^2e^{-\lambda}+C_4\left(1-e^{-\lambda}\right)^2<0, \tag{22}$$

where $C_1=(n-n_0)^2, C_2=2(n-n_0)\sum_{i=1}^{n}x_i, C_3=(n-n_0)z_{1-\alpha/2}^2, C_4=\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}x_i-z_{1-\alpha/2}^2\right)$ and these terms are free of parameter λ. In this paper, this interval is named "SCI". Likewise, if using $J_P(\lambda)$ instead of $I_P(\lambda)$, the score statistic, or $S_P(\lambda)/\sqrt{J_P(\lambda)}$, will be

$$\frac{S_P(\lambda)}{\sqrt{J_P(\lambda)}}=\frac{\sum_{i=1}^{n}x_i\left(1-e^{-\lambda}\right)-(n-n_0)\lambda}{\sqrt{\lambda(n-n_0)\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)}}=\frac{-(n-n_0)\lambda+\sum_{i=1}^{n}x_i-e^{-\lambda}\sum_{i=1}^{n}x_i}{\sqrt{-\lambda^2e^{-\lambda}(n-n_0)+(n-n_0)\lambda\left(1-e^{-\lambda}\right)}}. \tag{23}$$

Additionally, the score confidence interval using the profile and expected FI is a set of λ that achieves $\left|S_P(\lambda)/\sqrt{J_P(\lambda)}\right|<z_{1-\alpha/2}$, or is the root of the following equation

$$\lambda^2D_1+\lambda\left(e^{-\lambda}-1\right)D_2+\lambda^2e^{-\lambda}D_3+\left(1-e^{-\lambda}\right)D_4<0, \tag{24}$$

where $D_1=(n-n_0)^2, D_2=2(n-n_0)\sum_{i=1}^{n}x_i+(n-n_0)z_{1-\alpha/2}^2, D_3=(n-n_0)z_{1-\alpha/2}^2, D_4=\left(\sum_{i=1}^{n}x_i\right)^2$. The interval corresponding to (24) is referred to as "SCJ".

4. Simulation Study

For the simulations, count data are generated from the R function "rzipois", which is in the VGAM package. The characteristics of the data include the probability of zero (π = 0.1, 0.3, 0.5, 0.7 and 0.9), the Poisson parameters (λ =1, 3, 5, 7 and 9), and the sample sizes (n = 10, 30, 60, 100, 200 and 500). For fixed λ and n, the proportion of zeros in data depends only on parameter π. As π increases, the proportion of zeros, $p_0=E(N_0/n)=\left[\pi+(1-\pi)e^{-\lambda}\right]$, increases. It is also observed that $\partial p_0/\partial\lambda=-(1-\pi)e^{-\lambda}$ is always less than 0; therefore, as the value of λ increases, the proportion of zeros will mathematically decrease.

The coverage probability (CP), average length (AL), and coverage per unit length (CPUL) of the confidence intervals (CIs) are estimated by Monte Carlo simulations with 10,000 repetitions. Figures 3-5 illustrate CPs, ALs, and CPULs, respectively. The results corresponding to λ = 7 and 9 are not presented because they are nearly identical to those with λ = 5. Examination of the results reveals that the characteristics of the data greatly impact all criteria: CP, AL, and CPUL.

In Figure 3, the performance is examined through the CP and it is found that when λ = 3 and 5, the WCI performs as well as the SCI and SCJ for most sample sizes (n = 30, 60, 100, 200 and 500); the CPs are all close to 0.95. When λ has a low value, such as 1, and the sample size is small, the WCI is inferior to the other two intervals. Also, for a small number of n, when π increases, the CP of WCIs tends to decrease whereas it slightly increases for SCJs.

The ALs are investigated in Figure 4. From all three intervals, when the sample size increases or/and λ decreases, the range of ALs will decrease. For instance, at λ = 3 and n = 10, the AL increases from 2.5 to 6.8; at λ = 3 and n = 200, AL rises from 0.55 to 1.67. In addition, if n and π are fixed, the larger λ is, the higher AL values are in any intervals. For example, at n = 30 and π = 0.5, the ALs of SCIs for λ = 1, 3 and 5 are approximately 3, 3.5, and 4.2, respectively. Generally, SCJs tend to have the highest AL followed by SCIs and WCIs for a combination of (λ, π, n).

Unhapipat et al. (2016) suggested adopting a single criterion incorporating the AL and CP. This is called a coverage per unit length (CPUL) which is defined as CP/AL. It is useful for comparisons of CIs with different CPs and ALs but the same sample size. The CPULs of three confidence intervals are shown in Figure 5. Provided that values of λ, π and n are fixed, all intervals are noticeably indifferent. Nevertheless, the parameters of the ZIP still affect the CPUL such that as λ increases and/or π increases, the CPUL will decrease.

5. Real data Analysis

To demonstrate the calculation of three confidence intervals, the yearly numbers of meteorite falls in the United States are presented from 1995 to 2020, and the dataset is shown in Figure 6 (left). A meteorite fall is defined as a collected meteorite after it was observed by people or automated devices. The other type of meteorites is called "find" and it is not included to our analysis. The data used in this paper is obtained from The Meteoritical Bulletin Database (MBD), available online at http://www.lpi.usra.edu/meteor/.

The maximum likelihood estimates of λ and π are 1.48 and 0.35, respectively, and estimated and empirical probabilities are shown in Figure 6 (right). The p-values for the goodness-of-fit test using $\mathrm{ZIP}\left(\hat{\lambda}_{ML} = 1.48, \hat{\pi}_{ML} = 0.35\right)$ and $\mathrm{Poisson}\left(\hat{\lambda}_{ML} = 0.96\right)$ are 0.5704 and 0.1395, respectively. Thus, the ZIP model is far better than the Poisson model for this dataset. (18), (22), and (24) yield intervals of (0.7072, 2.2556) for WCI, (0.6987, 2.2726) for SCI, and (0.8913, 2.4845) for SCJ, respectively. The corresponding interval lengths are 1.548, 1.574, and 1.593; this agrees with our simulation study that SCJ has the highest AL followed by SCI and WCI for small values of n and λ. The observed sample size is 26, $\hat{\pi}_{ML}$ is 0.35, and $\hat{\lambda}_{ML}$ is 1.48. Our closest setting has a sample size of 30, π of 0.3, and λ of 1 or λ of 3. In Figure 3, simulations suggest that CPs for all three methods are nearly equal at a value of 0.95.

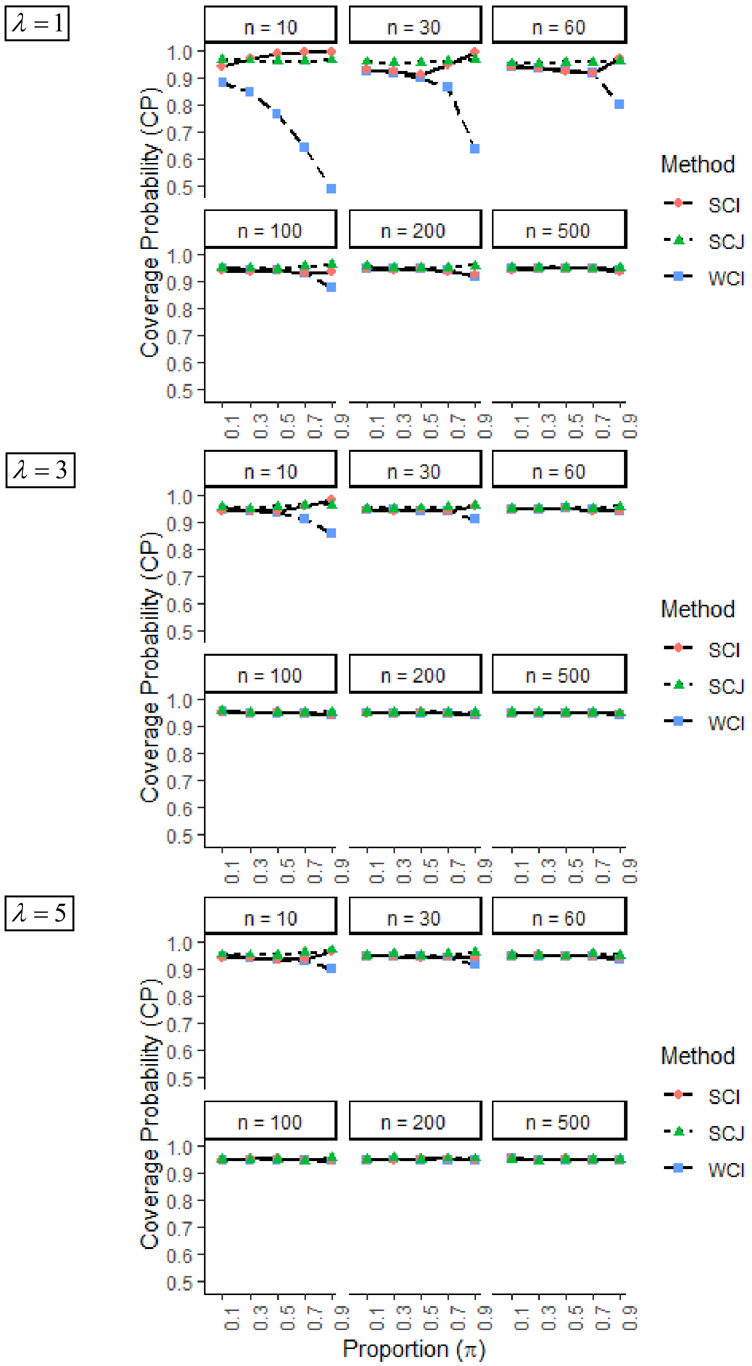

Figure 3 Coverage probabilities of three types of intervals obtained from the samples under $\text{ZIP}(\lambda, \pi)$, $\pi = 0.1, 0.3, 0.5, 0.7, 0.9$, and $\lambda = 1, 3, 5$.

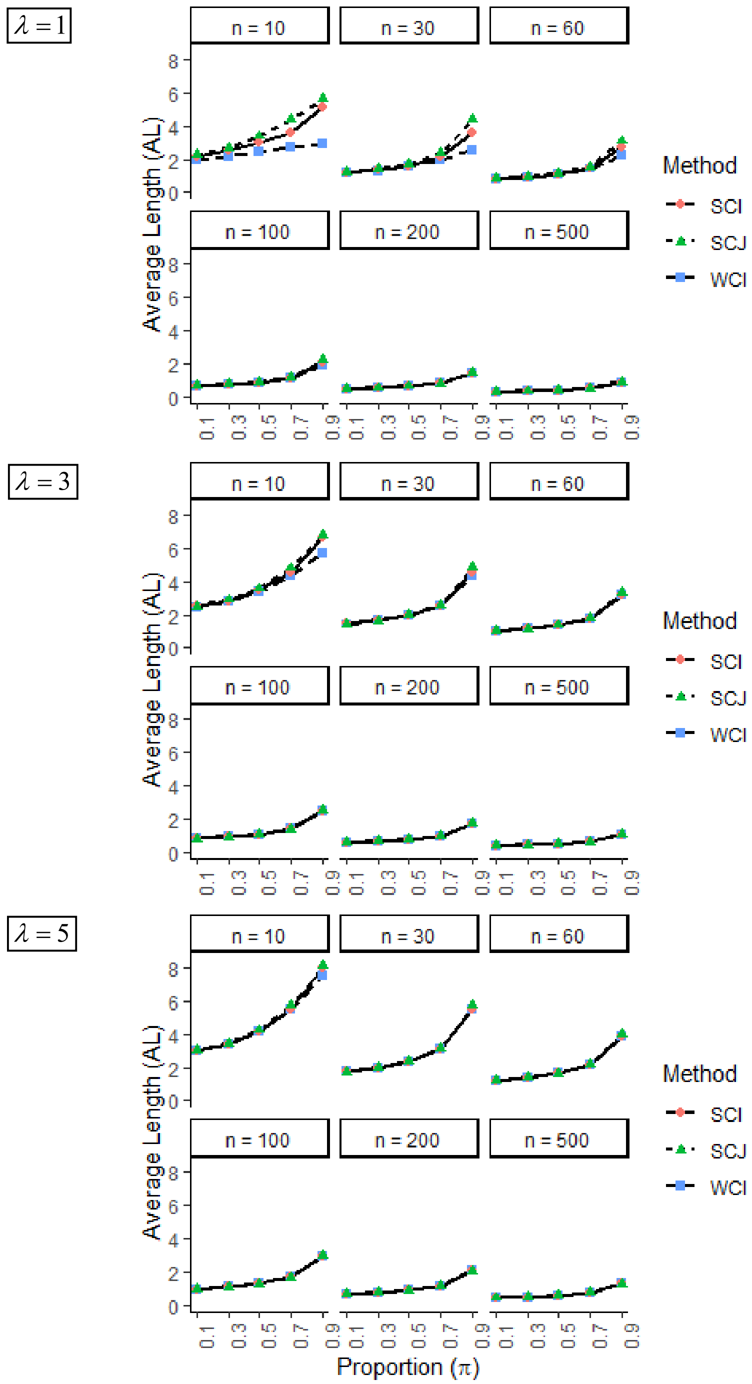

Figure 4 Average length of three types of intervals obtained from the samples under $\text{ZIP}(\lambda,\pi)$, $\pi = 0.1, 0.3, 0.5, 0.7, 0.9$, and $\lambda = 1, 3, 5$.

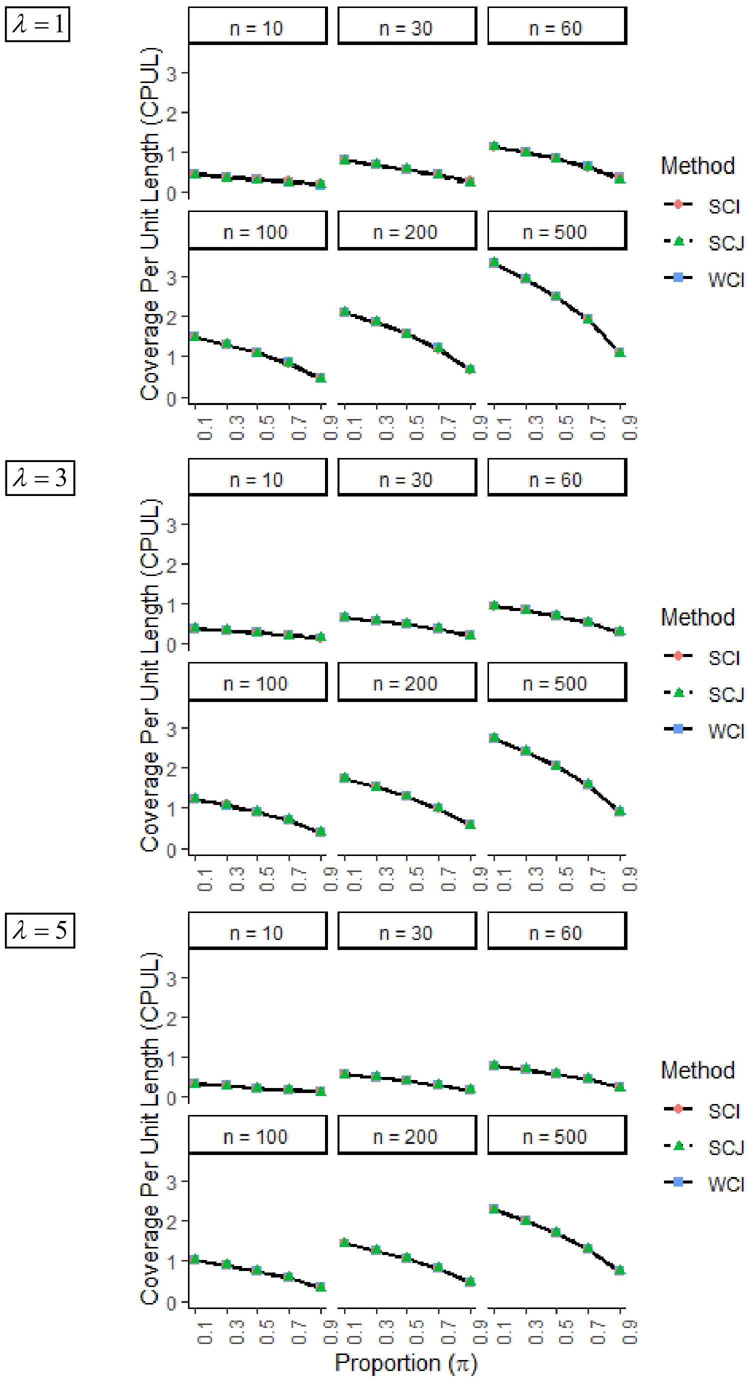

Figure 5 Coverage per unit length of three types of intervals obtained from the samples under $\text{ZIP}(\lambda, \pi)$, $\pi = 0.1, 0.3, 0.5, 0.7, 0.9$, and $\lambda = 1, 3, 5$.

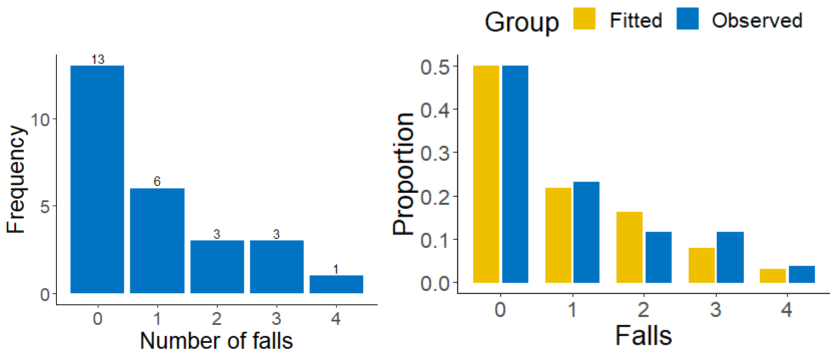

Figure 6 Frequency distribution of the number of meteorite falls (left) and fitted and observed proportions of meteorite falls (right).

6. Conclusions

In this paper, three interval estimations were derived for the Poisson parameter in a zero–inflated Poisson distribution, in which parameter π is unknown. The formulae of WCIs using either the joint likelihood or profile likelihood were mathematically derived and proved to be equivalent because $J^{-1}(\hat{\lambda}_P) = I^{-1}(\hat{\lambda}_P) = J^{11}(\hat{\lambda}_{ML}, \hat{\pi}_{ML}) = I^{11}(\hat{\lambda}_{ML}, \hat{\pi}_{ML})$. Also, the formulae of two score intervals using profile likelihood were derived in terms of inequalities.

The comprehensive simulation study was conducted. Three criteria involving the coverage probability and average length of confidence intervals estimated by the Monte Carlo method were to compare their performances. In conclusion, two score intervals employing either observed or expected Fisher information were found to be comparable but they both outperformed the Wald confidence interval, especially in a small sample size with a low value of λ and a high value of π; however, a shorter average length is a good compensation for WCIs. Overall, if λ is low, the score intervals are preferable for all π's, and if λ is high (not less than 5), all confidence intervals are equally advisable.

Acknowledgements

This study was supported by Thammasat University Research Fund, Contract No. TUFT 030/2563.

References

Balin SJ, Cascalho M. The rate of mutation of a single gene. Nucleic Acids Res. 2010; 38(5): 1575-1582.

Beckett S, Jee J, Ncube T, Pompilus S, Washington Q, Singh A, Pal N. Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. Involve. 2014; 7(6): 751-767.

Brent RP. Algorithms for Minimization without Derivatives. New York: Dover Publications; 2013.

Cameron AC, Trivedi PK. Regression analysis of count data. New York: Oxford University Press; 2013.

David M, Jemna D. Modeling the frequency of auto insurance claims by means of Poisson and negative binomial models. Sci Ann Econ. 2015; 62(2): 151-168.

DeGroot MH, Schervish MJ. Probability and statistics. Boston: Pearson Education; 2018.

Ibrahim R, Ye H, L'Ecuyer P, Shen H. Modeling and forecasting call center arrivals: A literature survey and a case study. Int J Forecast. 2016; 32(3): 865-874.

Lambert D. Zero-inflated Poisson regression, with an applicant to defects in manufacturing. Technometrics. 1992; 34(1): 1-14.

Lee AH, Wang K, Yau KW. Analysis of zero-inflated Poisson data incorporating extent of exposure. Biom J. 2001; 43(8): 963-975.

Murphy SA, Van Der Vaart AW. On profile likelihood. J Am Stat Assoc. 2000; 95(450): 449-465.

Numna S. Analysis of extra zero counts using zero-inflated Poisson models. Master [Thesis]. Thailand: Prince of Songkla University; 2009.

Paneru K, Padgett RN, Chen H. Estimation of zero-inflated population mean: a bootstrapping approach. J Mod Appl Stat Methods. 2018; 17(1): eP2557.

Rohde CA. Introductory statistical inference with the likelihood function. Cham: Springer International Publishing; 2014.

Ridout M, Demetrio CGB, Hinde J. Models for count data with many zeros. IBC98: Proceedings of the XIXth International Biometric Conference; 1998 Dec 14-18; South Africa. 1998. pp.179-192.

Sarul LS, Sahin S. An application of claim frequency data using zero inflated and hurdle models in general insurance. J Bus Econ Fin. 2015; 4(4): 732-743.

Schwartz J, Giles DE. Bias-reduced maximum likelihood estimation of the zero-inflated Poisson distribution. Commun Stat - Theory Methods. 2016; 45(2): 465-478.

Sakthivel KM, Rajitha CS. Estimation of zero-inflation parameter in zero-inflated Poisson model. Int J Math Trends Technol. 2018; 56(2): 135-140.

Srisuradetchai P, Junnumtuam S. Wald Confidence intervals for the parameter in a Bernoulli component of zero-inflated Poisson and zero-altered Poisson models with different link functions. Sci Technol Asia. 2020; 25(2): 1-14.

Tlhaloganyang BP, Sakia RM. Zero inflated Poisson distribution in equidispersed data with excessive zeros. Res J Math Stat Sci. 2020; 8(1): 31-34.

Unhapipat S, Chen JY, Pal N. Small sample inferences on the sharpe ratio. Am J Math Manag. 2016; 35(2): 105-123.

Unhapipat S, Tiensuwan M, Pal N. Bayesian predictive inference for zero-inflated Poisson (ZIP) distribution with applications. Am J Math Manag. 2018; 37(1): 66-79.

Vandenbroek J. A score test for zero inflation in a Poisson distribution. Biometrics. 1995; 51(2): 738-743.

Wagh YS, Kamalja KK. Zero-inflated models and estimation in zero-inflated Poisson distribution. Commun Stat - Simul Comput. 2018; 47(8): 2248-2265.

Waguespack D, Krishnamoorthy K, Lee M. Test and confidence interval for the mean of a zero-inflated Poisson distribution. Am J Math Manag. 2020; 39(4): 383-390.

Xie M, He B, Goh TN. Zero-inflated Poisson model in statistical process control. Comput Stat Data Anal. 2001; 38(2): 191-201.

Xu HY, Xie M, Goh TN. Objective Bayes analysis of zero-inflated Poisson distribution with application to healthcare data. IIE Trans. 2014; 46(8): 843-852.