Accurate Inference for the Mean of the Poisson-Exponential Distribution

Wei Lin [1]**, Xiang Li** [2]**, Augustine Wong** [2]

[1] Department of Mathematics and Statistics, Thompson Rivers University, 805 TRU Way, Kamloops, British Columbia, Canada V2C 0C8.

[2] Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3.

Abstract. Although the random sum distribution has been well-studied in probability theory, inference for the mean of such distribution is very limited in the literature. In this paper, two approaches are proposed to obtain inference for the mean of the Poisson-Exponential distribution. Both proposed approaches require the log-likelihood function of the Poisson-Exponential distribution, but the exact form of the log-likelihood function is not available. An approximate form of the log-likelihood function is then derived by the saddlepoint method. Inference for the mean of the Poisson-Exponential distribution can either be obtained from the modified signed likelihood root statistic or from the Bartlett corrected likelihood ratio statistic. The explicit form of the modified signed likelihood root statistic is derived in this paper, and a systematic method to numerically approximate the Bartlett correction factor, hence the Bartlett corrected likelihood ratio statistic is proposed. Simulation studies show that both methods are extremely accurate even when the sample size is small.

Corresponding Author: Wei Lin (wlin@tru.ca)
Xiang Li (lovedaershou@126.com)
Augustine Wong (august@yorku.ca)

Likelihood Root.

MSC: 62E20, 62F03, 62F05, 62F40.

1 Introduction

In probability theory, a compound Poisson distribution is defined by $S = \sum_{i=1}^{N} X_i$ where X_i is a sequence of identical and independently distributed (iid) random variables and N follows a Poisson distribution with X_i independent of N. As an example for the motivation of the compound Poisson distribution, let X be the claims that an insurance company has to pay to its clients. The arrival times of these payments, which is a counting process, follow a specific distribution. When the underlying counting process is a Poisson process, and the claims are random variables, then the amount of total claims is a compound Poisson process. The compound Poisson distribution is not only used in actuarial science and insurance for modeling the distribution of total claim, but also used in reliability theory, queuing theory, molecular sequence analysis, etc.

Although the compound Poisson distribution appears in many fields, the explicit distribution of S is, generally, not available in closed form. Barbour and Chryssaphinou (2001) review and discuss Stein's method to derive bounds on the error bound in compound Poisson approximation. For the same problem, Bero (2003) uses Kerstan's method to derive new bounds for the total variation distance. More recently, Barbour *et al.* (2010) provide an information-theoretic treatment of the problem of approximating the distribution. As well, Alya and Low (2013) derive the approximated density for S when $X_1, \ldots, X_N$ are iid random variables from the Exponential distribution.

This paper is motivated by the previous insurance example. Our aim is to obtain inference for the mean parameter in the compound Poisson distribution with a sum of Exponentially distributed random variables. Specifically, we assume that the amount of a claim, X_i, is Exponentially distributed with mean β and the number of claims, N, is following a Poisson distribution with mean λ. Then $S = X_1 + \cdots + X_N$ is the aggregate claims generated by the portfolio for the period under study. This specific compound Poisson process is often referred to as the Poisson-Exponential distribution. It is important to note that if $N = 0$ is realized, then S must be zero.

Since S follows the Poisson-Exponential distribution, the moment generating functi-

on (mgf) can be obtained as

$$m_S(t) \quad = \quad m_N\left[\log m_X(t)\right].$$

Hence, $E(S)$ is

$$\psi = E(S) = E(N)E(X) = \lambda\beta.$$

The main objective of this paper is to obtain inference of ψ based on an observed sample of aggregate claims $(s_1, s_2, \ldots, s_k)$. We are especially interested in the case where sample size, k, is small.

In the literature, there are limited studies on the inference of the Poisson-Exponential distribution. Thmazella *et al.* (2013) propose a Bayesian inference method for the shape parameter of the newly proposed Poisson-Exponential distribution by using Markov chain Monte Carlo simulation. Ghribi and Masmoudi (2013) introduce the compound Poisson model in the Bayesian network framework and apply implicit inference to this model for learning the parameters of the network. Both are Bayesian methods and pay no attention to the inference of the mean parameter. In this paper, we address the problem of mean parameter inference of the Poisson-Exponential distribution. First, we review the approximation to obtain the density function of S, and hence, the corresponding log-likelihood function. Then two likelihood-based inference approaches are proposed. Finally, we study the performance of the proposed inference methods when sample size is small by simulation studies.

The outline of the rest of the paper is as follows: in Section 2, we review the saddlepoint approximation of the density function of S. In Section 3, the first-order and third-order likelihood inference methods are reviewed and specifically applied to obtaining inference of the Poisson-Exponential distribution. Moreover, an algorithm to estimate the Bartlett correction factor is proposed in Section 4. To compare the accuracy of the methods discussed in this paper, simulation studies are conducted, results are reported and discussed in Section 5. Further concluding remarks are given in Section 6.

2 Saddlepoint Approximation for the Poisson-Exponential Distribution

Let $Y_1, \ldots, Y_n$ be iid random variables with a known mgf $m_Y(t)$. By inverting $m_Y(t)$, Daniels (1954) proposes using the saddlepoint method to approximate the density

function of $\bar{Y} = \sum_{i=1}^{n} Y_i/n$. The approximated density function takes the form in below

$$f_{\bar{Y}}(\bar{y}) = c \left\{ \frac{n}{K_Y^{(2)}(\hat{t})} \right\}^{1/2} \exp \left\{ n \left[K_Y(\hat{t}) - \hat{t}\bar{y} \right] \right\},$$

where c is the normalizing constant, $K_Y(t) = \log m_Y(t)$ is the cumulant generation function,

$$K_Y^{(j)}(t) = \frac{d^j K_Y(t)}{dt^j},$$

is the jth derivative of the cumulant generating function, and $\hat{t}$ is the saddlepoint satisfying $K_Y^{(1)}(\hat{t}) = \bar{y}$. Moreover, based on this approximated density function and complex integration, Lugannani and Rice (1980) derive the approximated distribution function of $\bar{Y}$, which takes the form

$$F_{\bar{Y}}(\bar{y}) = \Phi(\hat{w}) + \phi(\hat{w}) \left\{ \frac{1}{\hat{w}} - \frac{1}{\hat{u}} \right\} + O(n^{-3/2}),$$

where $\phi()$ and $\Phi()$ are the density and cumulative distribution functions of the standard normal distribution,

$$\hat{w} = \text{sgn}(\hat{t}) \left\{ 2n \left[\hat{t}\bar{y} - K_Y(\hat{t}) \right] \right\}^{1/2}, \quad \text{and} \quad \hat{u} = \hat{t} \left[n K_Y^{(2)}(\hat{t}) \right]^{1/2}.$$

Since there is a singularity point at $\hat{t} = 0$, Daniels (1987) derives the limiting value of $F_{\bar{Y}}(\bar{y})$ at the singularity point:

$$F_{\bar{Y}}(\bar{y}) = \frac{1}{2} + \frac{K_Y^{(3)}(0)}{6\sqrt{2\pi} \left[K_Y^{(2)}(0) \right]^{3/2}}.$$

Consider the random sum problem with N being a Poisson random variable with mean λ, and $X_1, \ldots, X_N$ being iid Exponentially distributed with mean β. For $N = n \neq 0$, and $t \neq 1/\beta$, we have

$$m_N(t) = \exp \left\{ \lambda(e^t - 1) \right\}, \quad \text{and} \quad m_X(t) = \frac{1}{1 - \beta t}.$$

Therefore, the mgf of the aggregate claims is

$$m_S(t) = m_N \left[\log m_X(t) \right] = \exp \left\{ \lambda \left[\frac{1}{1 - \beta t} - 1 \right] \right\},$$

and the corresponding cumulant generating function is

$$K_S(t) = \log m_S(t) = \lambda \left[\frac{1}{1 - \beta t} - 1 \right].$$

By applying the saddlepoint method, Alya and Low (2013) derive the approximated density for the aggregate claims S,

$$f_S(s) = \frac{c}{\sqrt{K_S^{(2)}(\hat{t})}} \exp \left\{ K_S(\hat{t}) - \hat{t}s \right\}, \tag{2.1}$$

where

$$K_S^{(1)}(t) = \frac{\lambda \beta}{(1 - \beta t)^2}, \quad K_S^{(2)}(t) = \frac{2\lambda \beta^2}{(1 - \beta t)^3}, \tag{2.2}$$

and $\hat{t}$ satisfies $K_S^{(1)}(\hat{t}) = s$, which results in

$$\hat{t} = \frac{1}{\beta} \left[1 - \left(\frac{\lambda \beta}{s} \right)^{1/2} \right]. \tag{2.3}$$

Based on the approximated density and applying the Lugganni and Rice method, Alya and Low (2013) examine the accuracy of the cumulative distribution of S for given λ and β. But in practice, λ and β are unknown, so is the mean parameter, $\psi = \lambda \beta$. Obtaining inference ψ is of great interest to many insurance companies. However, there is a paucity of literature on this problem. Two inferential methods for ψ are proposed in the next two sections. The first method derived in Section 3 is based on the third-order likelihood-based method discussed in Fraser and Reid (1995), and a new numerical algorithm to approximate the Bartlett correction factor of the likelihood ratio method (Bartlett, 1937) is given and discussed in Section 4.

3 Inference for the Mean Aggregate Claims based on the Third-order Likelihood-based Method

In parametric statistical inference, data are modeled as a realization of a variable y with the density function $f(y; \theta)$. We assume that $f(y; \theta)$ has a known form and θ is an unknown parameter taking value in a finite-dimensional Euclidean space Θ. As

defined in Kalbfleisch (1985), the likelihood function based on a random sample, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, is

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; \mathbf{y}) = a \prod_{i=1}^{n} f(y_i; \theta), \ \theta \in \Theta \subseteq \mathcal{R}^d, \tag{3.1}$$

where $a > 0$ is an arbitrary constant. Reid (2010) notes that the value of the likelihood function is only meaningful in relative terms. The likelihood function presents almost everything that a model has to say about the observed data. Among the possible values of θ, the maximum likelihood value, $\hat{\theta}$, is the one which maximizes the likelihood function. For computational convenience, it is often easier to maximize the log-likelihood function,

$$\ell(\theta) = \ell(\theta; \mathbf{y}) = \log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log\{f(y_i; \theta)\} + \log(a), \ \theta \in \Theta, \tag{3.2}$$

which yields the same $\hat{\theta}$ for each a. Hence, without loss of generality, a is set to be 1 hereafter.

For a vector parameter model, we assume $\boldsymbol{\theta} = (\psi, \lambda)$ with interest parameter ψ of dimension one and nuisance parameter vector λ. A simple way to eliminate the nuisance parameter is to use the profile log-likelihood function for ψ:

$$\ell_p(\psi) = \max_{\lambda} \ell(\psi, \lambda) = \ell(\psi, \tilde{\lambda}_\psi) = \ell(\tilde{\boldsymbol{\theta}}_\psi), \tag{3.3}$$

where $\tilde{\boldsymbol{\theta}}_\psi = (\psi, \tilde{\lambda}_\psi)$ is the constrained maximum likelihood estimate when ψ is fixed. In practise, likelihood inference for ψ is typically based on the following quantities:

$$\text{Signed likelihood root,} \quad r(\psi) = \text{sign}(\hat{\psi} - \psi) \sqrt{2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\}} \tag{3.4}$$

$$= \text{sign}(\hat{\psi} - \psi) \sqrt{2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}}_\psi)\}},$$

$$\text{Score quantity,} \quad s(\psi) = j_p(\hat{\psi})^{-1/2} \left(\frac{\partial \ell_p(\psi)}{\partial \psi} \right), \tag{3.5}$$

$$\text{Wald quantity,} \quad q(\psi) = j_p(\hat{\psi})^{1/2}(\hat{\psi} - \psi), \tag{3.6}$$

where $j_p(\hat{\psi})$ is the observed information, which is the negative second derivative of the profile log-likelihood function evaluated at the $\hat{\psi}$, and the sign(x) function takes "+" if x is positive, "−" if x is negative, and "0" if $x = 0$.

Reid (2010) gives a detailed review on the asymptotic distribution of these quantities. More specifically, each of the above quantities has an asymptotic standard normal distribution with first order accuracy, $O(n^{-1/2})$. They are referred to as first-order inference quantities. Murphy and Van der Vaart (2000) present a rigorous proof of the likelihood ratio statistic, in the presence of nuisance parameter, is asymptotically distributed as the Chi-square distribution. The asymptotic distribution of (7) follows when the parameter of interest is of dimension one.

It is well-known that the accuracy of first-order inference quantities may be poor or questionable when the sample size n is not large. In the literature, many improvements based on the signed likelihood root, $r(\psi)$, have been proposed and developed.

An important variant of the signed likelihood root is the modified signed likelihood root proposed by Barndorff-Nielsen (1986, 1991), denoted by $r^*(\psi)$,

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log \frac{Q(\psi)}{r(\psi)}, \tag{3.7}$$

where $r(\psi)$ is defined in (3.4). A general form of $Q(\psi)$ is derived by Fraser *et al.* (1999) and given as

$$Q(\psi) = \frac{|\varphi(\hat{\boldsymbol{\theta}}) - \varphi(\tilde{\boldsymbol{\theta}}_\psi) \quad \varphi_\lambda(\tilde{\boldsymbol{\theta}}_\psi)|}{|\varphi_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})|} \left\{ \frac{|j_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})|}{|j_{\lambda\lambda}(\tilde{\boldsymbol{\theta}}_\psi)|} \right\}^{1/2}, \tag{3.8}$$

where $\varphi(\boldsymbol{\theta})$ is the exact or approximated canonical parameter in the likelihood function, and

$$j_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad j_{\lambda\lambda}(\tilde{\boldsymbol{\theta}}_\psi) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \lambda \partial \lambda^T}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_\psi},$$

are the observed full and nuisance information function, respectively. The modified signed likelihood root, $r^*(\psi)$, converges to the standard normal distribution with third-order accuracy, $O(n^{-3/2})$. Hence, the p-value function of ψ defined in Fraser (2017) can be approximated by

$$p(\psi) = \Phi\left(r^*(\psi)\right) \quad \text{and} \quad p(\psi) = \Phi\left(r(\psi)\right) + \phi\left(r(\psi)\right) \left\{ \frac{1}{r(\psi)} - \frac{1}{Q(\psi)} \right\},$$

using the modified signed likelihood root and the Lugannani and Rice formula, respectively. Both approximations have third-order accuracy and are equivalent up to third-order accuracy (Fraser and Reid , 1995). Thus, as discussed in Fraser (2017), either approximation of the p-value function can be used to obtain inference for ψ.

Applying the saddlepoint method to the Poisson-Exponential aggregate claims problem, the log-likelihood function of parameters (λ, β) for a given sample of size m is

$$\ell(\lambda, \beta; s_1, \ldots, s_m) = 2\Big(\sum_i^m \sqrt{s_i}\Big)\frac{\lambda}{\sqrt{\lambda\beta}} - \Big(\sum_i^m s_i\Big)\frac{1}{\beta} - m\lambda - \frac{m}{4}\log(\lambda\beta) + \frac{m}{2}\log(\lambda), \quad (3.9)$$

where $s_i = X_{i,1} + \ldots + X_{i,Ni}$ for $i = 1, \ldots, m$. Details of the derivation is in Appendix A. Since the parameter of interest is $\psi = \beta\lambda$, the log-likelihood function in terms of (ψ, λ) can be re-written as

$$\ell(\psi, \lambda; s_1, \ldots, s_m) = 2\Big(\sum_i^m \sqrt{s_i}\Big)\frac{\lambda}{\sqrt{\psi}} - \Big(\sum_i^m s_i\Big)\frac{\lambda}{\psi} - m\lambda - \frac{m}{4}\log(\psi) + \frac{m}{2}\log(\lambda). \quad (3.10)$$

By maximizing the log-likelihood function with respect to $\boldsymbol{\theta} = (\psi, \lambda)$, the overall maximum likelihood estimate is

$$\hat{\boldsymbol{\theta}} = (\hat{\psi}, \hat{\lambda}) = \Big(\bar{s}, \frac{\bar{s}}{4(\bar{s} - u\sqrt{\bar{s}})}\Big),$$

where $\bar{s} = \sum_{i=1}^m s_i/m$ and $u = \sum_{i=1}^m \sqrt{s_i}/m$. Moreover, for a given ψ, the constrained maximum likelihood estimate is $\tilde{\boldsymbol{\theta}}_\psi = (\psi, \tilde{\lambda}_\psi)$, where

$$\tilde{\lambda}_\psi = \frac{\psi}{2(\psi - 2u\sqrt{\psi} + \bar{s})}.$$

Hence, $r(\psi)$ can be obtained from (3.4). Notice that, for this case, the exact canonical parameter is directly available from the log-likelihood function, and is

$$\varphi'(\boldsymbol{\theta}) = \Big(\frac{\lambda}{\sqrt{\psi}}, \frac{\lambda}{\psi}\Big).$$

With the above canonical parameter and estimates, $Q(\psi)$ can be obtained from (3.8). Finally, the p-value function of ψ is approximated by either the modified signed likelihood root or the Lugganani and Rice formula with third-order accuracy.

4 Simulated Bartlett Corrected Likelihood Ratio Method

For ψ being a scalar parameter, Reid (2010) showed that $r(\psi)$ is asymptotically normally distributed, and Murphy and Van der Vaart (2000) derived that the generalized

likelihood ratio statisitic is asympotically distributed as a Chi-square distribution with 1 degree of freedom. Hence, we have

$$w(\psi) = [r(\psi)]^2 = 2\left[\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}}_\psi)\right], \tag{4.1}$$

is asymptotically distributed as Chi-square distribution with 1 degree of freedom. Thus, the p-value function for ψ can be approximated by

$$p(\psi) = P\left(\chi_1^2 \leq w(\psi)\right).$$

Bartlett (1937) proposed the transformation to $w^*(\psi)$, which takes the form

$$w^*(\psi) = \frac{w(\psi)}{B}, \tag{4.2}$$

such that the mean of the transformed statistic is the mean of the limiting distribution. This transformed statistic is known as the Bartlett corrected likelihood ratio statistic and B is referred to as the Bartlett correction factor. Bartlett (1937) showed that

$$p(\psi) = P\left(\chi_1^2 \leq w^*(\psi)\right),$$

has fourth-order accuracy. The simplest choice of B is the mean of the likelihood ratio statistic. However, except in a few well-defined cases, B is not available in explicit closed form. Obtaining the asymptotic expansion of B is also very complicated, and hence, it restricts its popularity in practice despite the high accuracy.

We propose the following algorithmic way to obtain B numerically.

Given: $(y_1, \ldots, y_n)$ is a sample of size n from a distribution with known density function $f(\cdot; \boldsymbol{\theta})$.

Interest: Inference concerning $\psi = \psi(\boldsymbol{\theta})$.

Obtain: We can obtain the overall maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, the constrained maximum likelihood estimate $\tilde{\boldsymbol{\theta}}_\psi$, and the likelihood ratio statistic $w(\psi)$.

Step 1:   Simulate M samples of size n from $f(\cdot; \tilde{\boldsymbol{\theta}}_\psi)$.

Step 2:   For each simulated sample, obtain the likelihood ratio statistic. As a result, we have $w_1(\psi), \ldots, w_M(\psi)$.

Step 3:   Calculate

$$\bar{w}(\psi) = \frac{\sum_{i=1}^{M} w_i(\psi)}{M},$$

which is an estimate of the mean of the likelihood ratio statistic. Hence, we have $\hat{B} = \bar{w}(\psi)$.

Step 4:   The simulated Bartlett corrected likelihood ratio statistic is

$$w^*(\psi) = \frac{w(\psi)}{\bar{w}(\psi)},$$

which has a limiting χ_1^2 distribution. Thus, the p-value function is $p(\psi) = P\left(\chi_1^2 \leq w^*(\psi)\right)$.

For the mean of Poisson-Exponential distribution, $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}}_\psi$, and $w(\psi)$ are obtained in Section 3. Thus, it is extremely easy to apply the proposed algorithm to obtain the p-value function of ψ.

5   Numerical Studies

Let us consider a sample of five aggregate claims:

$$3.24, \ 4.39, \ 1.30, \ 11.51, \ 4.15.$$

Assume the Poisson-Exponential model and our aim is to obtain inference for the mean aggregate claims. Figure 1 plots the p-value functions obtained by the signed likelihood root method (r), the likelihood ratio method (w), the modified signed likelihood root method (r^*) and the simulated Bartlett corrected likelihood ratio method (w^*), respectively. From the plot, the signed likelihood root method and the likelihood ratio method give exactly the same result because former is the signed root of the later, and

they are different from the modified signed likelihood root method and the simulated Bartlett corrected likelihood ratio method.
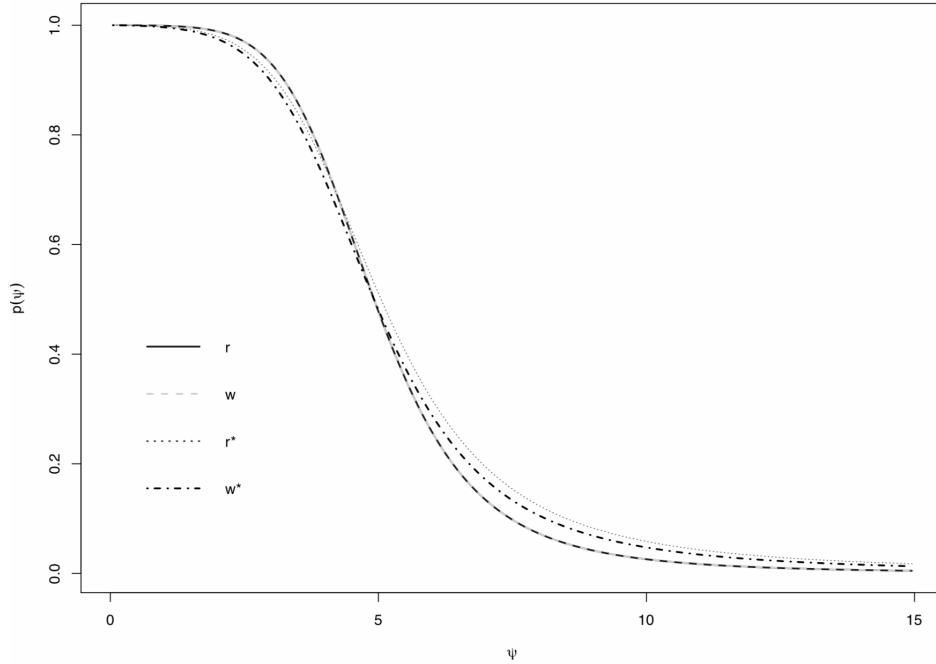

Figure 1: The p-value function plot for four inference methods. r represents the signed likelihood root method, w represents the likelihood ratio statistic, r^* represents the modified signed likelihood root method and w^* represents the simulated Bartlett corrected likelihood ratio method. Note that r and w coincide with one another.

In order to compare the accuracy of the four methods, simulation studies are performed. Moreover, the coverage probability of the confidence interval is used as the criterion for determining whether a method is accurate or not. Furthermore, the coverage probability is used to examine how close the proportion of samples not being rejected at α level of significance is to the chosen nominal coverage level $1 - \alpha$. Table 1 records the results and Figure 2 gives the visual idea of accuracy from a simulation with the following parameterization set-up:

1. N is distributed as Poisson($\lambda = 10$).

2. $X_1, \ldots, X_N$ are identical and independently distributed as exponential with mean $\beta = 0.5$, and $S = X_1 + X_2 \ldots + X_N$.

3. 5000 simulated samples of S of size n are generated, where $n = 2, 3, 5, 10, 20, 50$.

4. To obtain $\hat{B}$, we use $M = 1000$.

5. Report the proportion of $p(\psi = \lambda, \beta = 5) > 1 - \alpha = 0.95$ for each method discussed in this paper.

It is clear that the simulated Bartlett corrected likelihood ratio method gives results closest to the nominal level $1 - \alpha = 0.95$ even when the sample size, k, is as small as 2. It is followed closely by the modified signed lilelihood root method. The other two methods, the likelihood ratio method and the signed likelihood root method, do not give satisfactory results, especially when the sample size is small.

Table 1: The coverage probability of the 95% confidence interval at different sample size

n	$r(\psi)$	$w(\psi)$	$r^*(\psi)$	$w^*(\psi)$
2	0.7660	0.7648	0.9258	0.9572
3	0.8558	0.8552	0.9480	0.9554
5	0.9056	0.9044	0.9510	0.9500
10	0.9356	0.9354	0.9528	0.9490
20	0.9460	0.9454	0.9560	0.9504
40	0.9524	0.9516	0.9558	0.9506
50	0.9532	0.9528	0.9566	0.9512

A visual comparison of the accuracy level of the simulation result is presented in Figure 2. We observed that, as sample size increases, all the results are converging to the nominal value. However, when the sample size is small, the simulated Bartlett corrected likelihood method is extremely accurate with the modified signed likelihood root method closely following. This observation supports the theory that the former has fourth-order accuracy, $O(n^{-2})$, whereas the latter has third order of accuracy, $O(n^{-3/2})$. Moreover, the other two methods have only first-order accuracy, $O(n^{-1/2})$. Furthermore, the simulated Bartlett corrected likelihood method can easily be implemented into standard statistical software, such as R.
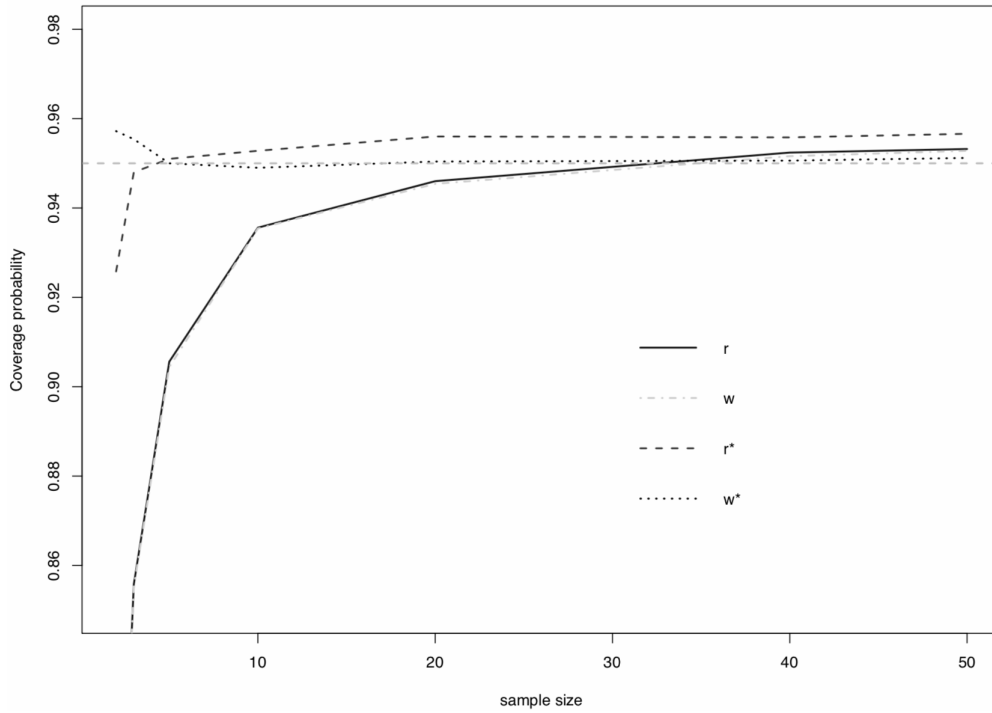
Figure 2: Coverage probability plot of 95% confidence interval. The horizontal dash line is the 95% nominal level. r represents the signed likelihood root method, w represents the likelihood ratio method, r^* represents the modified signed likelihood root method and w^* represents the simulated Bartlett corrected likelihood ratio method. Note that r and w coincide with one another.

In Appendix B, we reported results from a sample of other simulation studies with different choices of parameter values that we have performed. The results are similar to what we have presented here. More simulation studies have been carried out and the results are available upon request from authors.

6   Conclusion and Discussion

In this paper, two methods are proposed to obtain inference for the mean of the Poisson-Exponential distribution. One is based on the modified signed likelihood root method, where all equations are derived. The other method is an algorithm we developed to

approximate the Bartlett corrected likelihood ratio method. From the simulations we conducted, we observed that the simulated Bartlett corrected likelihood ratio method is extremely accurate even when the sample size is extremely small. A drawback of the modified signed likelihood root method is the complication in deriving the required statistics. Theoretically, the Bartlett corrected likelihood ratio method has an order of accuracy $O(n^{-2})$, while the modified signed likelihood root method has an order of accuracy $O(n^{-3/2})$ only. Our study confirms that the approximated Bartlett corrected likelihood ratio method proposed does achieve a better accuracy level than the modified signed likelihood ratio method in this case. Another advantage of the Bartlett corrected likelihood ratio method is that it can be applied to a vector parameter of interest, whereas, the modified signed likelihood ratio method is restricted to the scalar parameter of interest.

Appendix

Appendix A

Derivation of equations (12) and (13).

The density function of S is approximated by the saddlepoint approximation method and is presented in (1). Since $\hat{t}$ in (3), we have

$$1 - \beta\hat{t} = 1 - \beta\frac{1}{\beta}\left[1 - \left(\frac{\lambda\beta}{s}\right)^{1/2}\right] = \left(\frac{\lambda\beta}{s}\right)^{1/2}$$

Substituting $\hat{t}$, $K_S^{(2)}(\hat{t})$ and $K_S(t)$ into (1) we have

$$
\begin{aligned}
f_S(s) &= \frac{c}{\sqrt{K_S^{(2)}(\hat{t})}} \exp\left\{K_S(\hat{t}) - \hat{t}s\right\} \\
&= \frac{c}{\sqrt{\frac{2\lambda\beta^2}{(1-\beta t)^3}}} \exp\left\{\lambda\left[\frac{1}{1-\beta t} - 1\right] - \frac{1}{\beta}\left[1 - \left(\frac{\lambda\beta}{s}\right)^{1/2}\right]s\right\} \\
&= \frac{c}{\sqrt{2\lambda\beta^2(\lambda\beta/s)^{-3/2}}} \exp\left\{\lambda\left[\left(\frac{\lambda\beta}{s}\right)^{-1/2} - 1\right] - \frac{s}{\beta}\left[1 - \left(\frac{\lambda\beta}{s}\right)^{1/2}\right]\right\} \\
&= \frac{c}{\sqrt{2s^{-1}\frac{\sqrt{\lambda\beta}}{\lambda\sqrt{s}}}} \exp\left\{\lambda\left(\frac{s}{\lambda\beta}\right)^{1/2} - \lambda - \frac{s}{\beta} + \frac{\lambda\sqrt{s}}{\sqrt{\lambda\beta}}\right\} \\
&= \frac{c}{\sqrt{2s^{-1}}}\left(\frac{\sqrt{\lambda\beta}}{\lambda\sqrt{s}}\right)^{-1/2} \exp\left\{2\lambda\left(\frac{s}{\lambda\beta}\right)^{1/2} - \lambda - \frac{s}{\beta}\right\}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\ell(\lambda,\beta;s_i) &= \log a + \log f_S(s_i) = \log\left(\frac{c}{\sqrt{2s_i^{-1}}}\left(\frac{\sqrt{\lambda\beta}}{\lambda\sqrt{s_i}}\right)^{-1/2} \exp\left\{2\lambda\left(\frac{s_i}{\lambda\beta}\right)^{1/2} - \lambda - \frac{s_i}{\beta}\right\}\right) \\
&= \left(\log a + \log\frac{c}{\sqrt{2s_i^{-3/2}}}\right) - \frac{1}{4}\log(\lambda\beta) + \frac{1}{2}\log\lambda + 2\sqrt{s_i}\frac{\lambda}{\sqrt{\lambda\beta}} - \lambda - s_i/\beta,
\end{aligned}
$$

As explained in Section 3, without loss of generality, those terms that do not depend on parameters are set to 1 because they will not affect the rest of the likelihood-based calculations. Therefore, the log-likelihood function for a random sample $(s_1, s_2, \ldots, s_m)$ is

$$
\begin{aligned}
\ell(\lambda, \beta; s_1, s_2, \ldots, s_m) &= \sum_{i=1}^{m} \ell(\lambda, \beta; s_i) \\
&= \sum_{i=1}^{m} \left(-\frac{1}{4} \log(\lambda\beta) + \frac{1}{2} \log \lambda + 2\sqrt{s_i} \frac{\lambda}{\sqrt{\lambda\beta}} - \lambda - s_i/\beta \right) \\
&= -\frac{m}{4} \log(\lambda\beta) + \frac{m}{2} \log \lambda + 2\left(\sum_{i=1}^{m} \sqrt{s_i}\right) \frac{\lambda}{\sqrt{\lambda\beta}} - m\lambda - \frac{\sum_{i=1}^{m} s_i}{\beta}.
\end{aligned}
$$

which gives (12). With the reparameterization $\psi = \lambda\beta$, we have

$$
\ell(\lambda, \psi; s_1, s_2, \ldots, s_m) = -\frac{m}{4} \log(\psi) + \frac{m}{2} \log \lambda + 2\left(\sum_{i=1}^{m} \sqrt{s_i}\right) \frac{\lambda}{\sqrt{\psi}} - m\lambda - \sum_{i=1}^{m} s_i \frac{\lambda}{\psi}
$$

which gives (13).

Appendix B

Results of simulation studies from different choices of β and λ. The set-up of these studies is the same as the one presented in Section 5, but with different β and λ values. As observed from these plots, the simulated Bartlett corrected likelihood ratio method is extremely accurate with the modified signed likelihood root method closely following. The other two methods do not give satisfying results especially when the sample size is small.

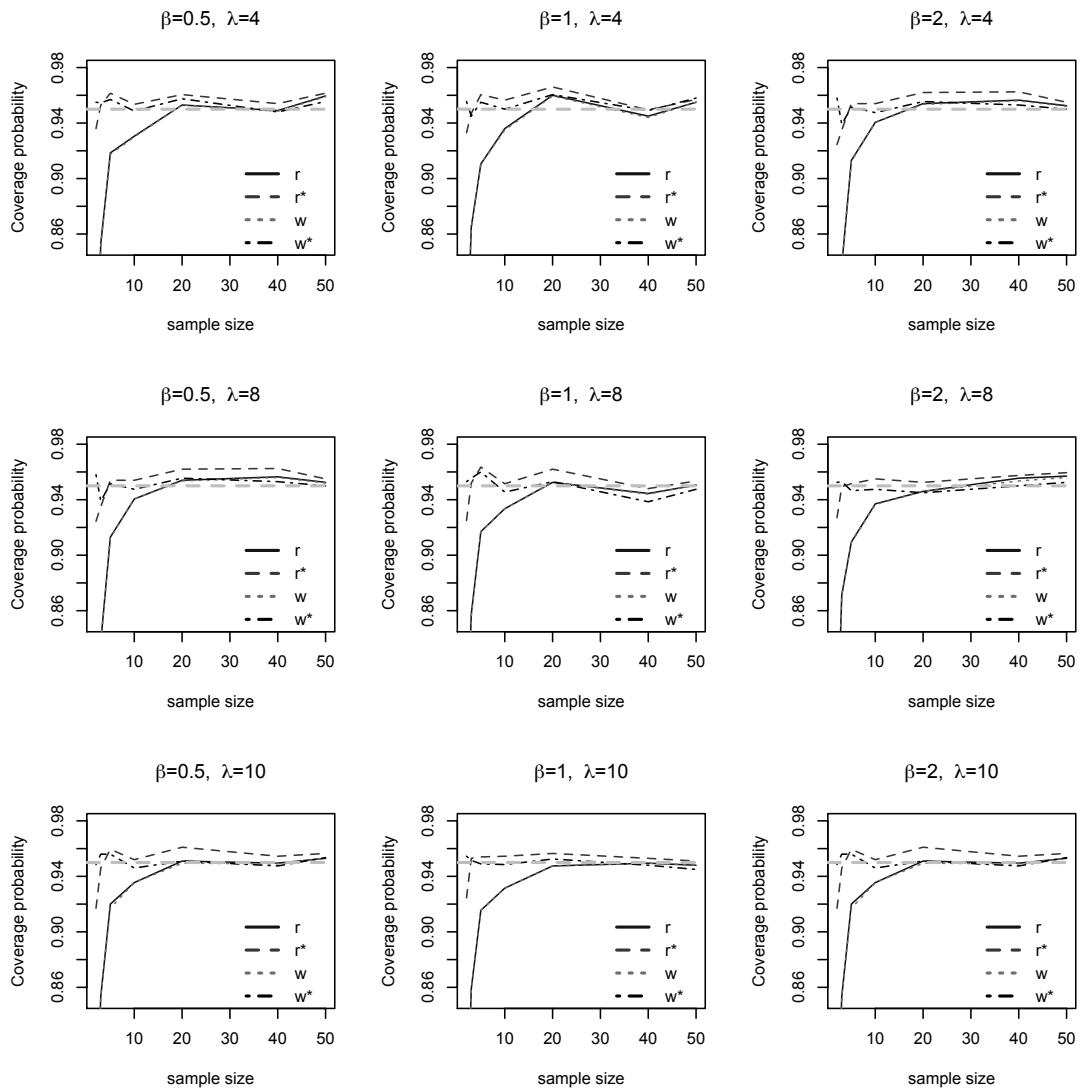Other simulation results are available upon request.

Figure 3: Coverage probability plot of 95% confidence interval. The horizontal dash line is the 95% nominal level. r represents the signed likelihood root method, w represents the likelihood ratio method, r represents the modified signed likelihood root method and w represents the simulated Bartlett corrected likelihood ratio method. Note that visually r and w coincide with one another.

References

Alya, A. M., and Low, H. C. (2013), Saddlepoint approximation to cumulative distribution function for poisson-exponential distribution. *Modern Applied Science*, **7**, 26-32.

Barbour, A. D., and Chryssaphinou, O. (2001), Compound poisson approximation: A user's guide. *The Annals of Applied Probability*, **11**, 964–1002.

Barbour, A. D., Johnson, O., Kontoyiannis, I., and Madiman, M. (2010), Compound poisson approximation via information functionals. *Electron. J. Prob.*, **15**, 1344–1369.

Barndorff-Nielsen, O. E. (1986), Inference on full and partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307–322.

Barndorff-Nielsen, O. E. (1991), Modified signed log likelihood ratio. *Biometrika*, **78**, 557–564.

Bartlett, M. S. (1937), Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London*, **160**, 268–282.

Bero, R. (2003), Ckerstan's method for compound poisson approximation. *The Annals of Probability*, **32**,1754–1771.

Daniels, H. E. (1954), Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, **25**, 631–650.

Daniels, H. E. (1987), Tail probability approximations. *International Statistical Review*, **55**, 37–48.

Fraser, D. A. S. (2017), p-values: The insight to modern statistical inference. *Annual Review of Statistics and Its Application*, **4**, 1–14.

Fraser, D. A. S., and Reid, N. (1995), Ancillaries and third-order significance. *Utilitas Mathematica*, **7**, 33–53.

Fraser, D. A. S., Reid, N., and Wu, J. (1999), A simple general formula for tail probabilities for frequentist and bayesian inference. *Biometrika*, **86**, 249–264.

Ghribi, A. and Masmoudi, A. (2013), Acompound poisson model for learning discrete bayesian networks. *Acta Mathematica Scientia*, **33**, 1767–1784.

Kalbfleisch, J. G. (1985), *Probability and Statistical Inference Volumne 2: Statistical Inference (2nd Edition)*. Springer-Verlag, New York.

Lugannani, R., and Rice, S. (1980), Saddlepoint approximation for the distribution of sums of random variables. *Advances in Applied Probability*, **12**, 475–490.

Murphy, S. A., and Van der Vaart, A. M. (2000), On profile likelihood. *Journal of the American Statistical Association*, **95**, 449–465.

Reid, N. (2010), Likelihood inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(5), 517–525.

Thmazella , V. L. D., Cancho, V. G., and Louzada V. (2013), Bayesian reference analysis for the poisson-exponential lifetime distribution. *Chilean Journal of Statistics*, **4**, 99–113.