Iranian Journal of Operations Research
Vol. 10, No. 2, 2019, pp. 62-77

Approximating Bayes Estimates by Means of the Tierney Kadane, Importance Sampling and Metropolis-Hastings within Gibbs Methods in the Poisson-Exponential Distribution: A Comparative Study

F. Bastan [1], S.M.T.K. MirMostafaee [2,*]

Here, we work on the problem of point estimation of the parameters of the Poisson-exponential distribution through the Bayesian and maximum likelihood methods based on complete samples. The point Bayes estimates under the symmetric squared error loss (SEL) function are approximated using three methods, namely the Tierney Kadane approximation method, the importance sampling method and the Metropolis-Hastings within Gibbs algorithm. The interval estimators are also obtained. The performance of the point and interval estimators are compared with each other by means of a Monte Carlo simulation. Several conclusions are given at the end.

Keywords: *Bayesian inference, Importance sampling method, Metropolis-Hastings within Gibbs algorithm, Monte Carlo simulation, Poisson-exponential distribution, Tierney Kadane approximation.*

Manuscript was received on 25/07/2019, revised on 30/10/2019 and accepted for publication on 13/12/2019.

1. Introduction

The exponential distribution is one of the popular lifetime distributions discussed by many authors and researchers. It has a simple probability function and enjoys simple mathematical properties. However, it suffers from a constant hazard rate function, which makes this distribution inappropriate for many lifetime situations. Thus, several scientists tried to extend this distribution to a more flexible distribution. Three old and famous extensions of the exponential distributions are the gamma, Weibull and generalized exponential distributions.

Recently, Cancho et al. [2] introduced the Poisson-exponential (PE) distribution by adding a shape parameter to the exponential distribution. The hazard rate function (hrf) of the PE distribution can be increasing as well. The motivation behind introducing the PE distribution can be explained as follows. Consider a parallel system whose number of components is not fixed. In other words, the number of the components, denoted by M, is a positive discrete random variable. Let $T_1, T_2, \cdots, T_M$ denote the lifetimes of the components and therefore $X = \max(T_1, T_2, \cdots, T_M)$ be the lifetime of the whole system. Now, suppose that M follows a zero–truncated Poisson distribution with parameter θ and the T_i are independent and identically distributed (iid) according to an exponential distribution with parameter λ. In addition, the T_i and M are independent. Then, the random variable X follows a PE distribution with parameters θ and λ. Therefore, the PE distribution can be used to model a parallel system, whose components' lifetimes are iid exponential random variables and the number of

* Corresponding Author.
[1] Department of Statistics, University of Mazandaran, Babolsar, Iran, Email: firozeh.bastan@gmail.com.
[2] Department of Statistics University of Mazandaran, Babolsar, Iran, Email: m.mirmostafaee@umz.ac.ir.

components follows a zero-truncated Poisson distribution. For more details, see Cancho et al. [2] and Louzada-Neto et al. [12].

The probability density function (pdf) of the PE distribution can be expressed as follows:

$$f(x) = \frac{\theta\lambda e^{-\lambda x-\theta e^{-\lambda x}}}{1-e^{-\theta}} \tag{1}$$

where $x > 0$, $\theta > 0$ and $\lambda > 0$. We note that θ and λ are the shape and scale parameters, respectively. As θ approaches zero, the PE distribution converges to an exponential distribution with parameter λ. The cumulative distribution function (cdf) of X is also given by

$$F(x) = 1 - \frac{1-e^{-\theta e^{-\lambda x}}}{1-e^{-\theta}}, \qquad x > 0, \qquad \lambda > 0, \qquad \theta > 0. \tag{2}$$

We write $X \sim \mathrm{PE}(\theta, \lambda)$, if the pdf and cdf of X can be written as (1) and (2), respectively. Moreover, the hrf of PE distribution is given by

$$h(x) = \frac{\theta\lambda \mathrm{e}^{-\lambda x-\theta \mathrm{e}^{-\lambda x}}}{1-\mathrm{e}^{-\theta \mathrm{e}^{-\lambda x}}}, \qquad x > 0, \qquad \lambda > 0, \qquad \theta > 0.$$

Figure 1 shows the PE pdfs and hrfs for selected values of θ and λ. From Figure 1, we see that the pdf can be either decreasing or unimodal and the hrf is increasing.

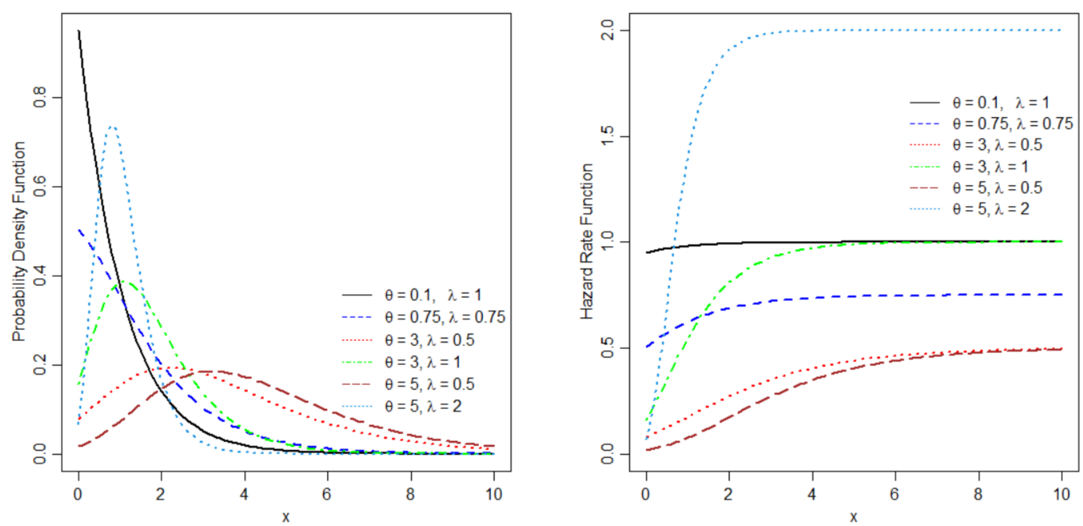

Figure 1. Pdfs (left panel) and hrfs (right panel) of the PE distribution for selected values of the parameters.

The Bayesian estimation for the PE distribution based on complete samples has been studied by Louzada-Neto et al. [12], Singh et al. [22] and Tomazella et al. [24]. Louzada-Neto et al. [12] and Singh et al. [22] found the Bayes estimators using a non-informative prior for the scale parameter and a gamma prior for the shape parameter. Tomazella et al. [24] focused on the Bayesian analysis using a joint reference prior for the parameters. As we will see later, an important problem that arises in Bayesian estimation of the parameters of the PE distribution is that the integrals pertaining to the

Bayes estimates do not seem to possess explicit forms. Therefore, one may use a suitable method to approximate these integrals. Louzada-Neto et al. [12], Singh et al. [22] and Tomazella et al. [24] applied the Metropolis-Hastings within Gibbs sampling method in order to handle this problem. However, there exist some other approximation methods that can also be implemented. Here, we are to apply two other methods, namely the Tierney Kadane and importance sampling methods to deal with the mentioned problem and then compare the methods by means of a Monte Carlo simulation. Here, we note that Arabi Belaghi et al. [1], who worked on the Bayesian estimation of the parameters of the PE distribution based on type-II censored order statistics, employed the importance sampling method as well as the Lindley approximation method. However, it seems that the Tierney Kadane approximation method for approximating Bayesian estimates of parameters of the PE distribution has not been investigated yet. We do not apply the Lindley method here because of its complexity.

The outline of the remaining parts of paper can be summarized as follows. Section 2 is devoted to maximum likelihood (ML) estimation. The information matrix and the asymptotic confidence intervals for the parameters are also obtained. In Section 3, we discuss the Bayesian estimation of the parameters under the SEL function. As we mentioned earlier and will see later, it seems that the integrals pertaining to the Bayes point estimates cannot be expressed explicitly. Therefore, we suggest using three methods to approximate these integrals. In this regard, we propose the Tierney Kadane approximation method, the importance sampling approximation method and the Metropolis-Hastings within Gibbs algorithm. The credible intervals for the parameters are also obtained using the method described by Chen and Shao [4]. We compare the performance of the point and interval estimators by means of a Monte Carlo simulation in Section 4. Finally, our concluding remarks are given in Section 5.

2. Maximum Likelihood Estimation

The maximum likelihood method is a most popular classical method of estimation of unknown parameters. Let $\boldsymbol{x} = (x_1, \cdots, x_n)$ be a set of an observed random sample of size n from PE(θ, λ). Then, the likelihood function is given by

$$L(\theta, \lambda|\boldsymbol{x}) = \frac{\theta^n \lambda^n \, e^{-\lambda \sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} e^{-\lambda x_i}}}{(1 - e^{-\theta})^n}. \tag{3}$$

The log-likelihood function of the parameters is given by

$$\ell(\theta, \lambda) = n \log(\theta\lambda) - \lambda \sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} e^{-\lambda x_i} - n \log\left(1 - e^{-\theta}\right).$$

The ML estimates of the parameters will be obtained by means of maximizing $\ell(\theta, \lambda)$ with respect to (w.r.t.) θ and λ; see Louzada-Neto et al. [12] for the nonlinear equations which can help us obtain the ML estimates; In our work here, we compute the ML estimates by using the *optim* function in R [16].

Though the main purpose of this paper does not involve interval estimation, we state here how we can obtain asymptotic confidence intervals for the unknown parameters. Let $\hat{\theta}_{MLE}$ and $\hat{\lambda}_{MLE}$ denote the ML estimators of θ and λ, respectively. Then, under the regularity conditions (see Lehmann and Casella [11]), as n tends to infinity, we have

$$\left((\hat{\theta}_{MLE}, \hat{\lambda}_{MLE}) - (\theta, \lambda)\right) \to N_2\left(\mathbf{0}, \frac{1}{n} I^{-1}(\theta, \lambda)\right),$$

Where by $\to$ we mean convergence in distribution, $N_2(.,.)$ denotes the two-variate normal distribution, $\mathbf{0}$ is a two-dimensional vector whose elements are both zero and $I^{-1}(\theta, \lambda)$ is the inverse of the information matrix, $I(\theta, \lambda)$; see Louzada-Neto et al. [12] who derived the elements of the information matrix.

The unknown parameters that appear in the elements of $I^{-1}(\theta, \lambda)$ may be replaced by their corresponding ML estimators to get an estimator of $I^{-1}(\theta, \lambda)$, which may be denoted as $\hat{I}^{-1}(\theta, \lambda)$. Therefore, the estimators of the asymptotic variances of $\hat{\theta}_{MLE}$ and $\hat{\lambda}_{MLE}$, that are denoted by $\widehat{Var}(\hat{\theta}_{MLE})$ and $\widehat{Var}(\hat{\lambda}_{MLE})$, respectively, are given by the first and second diagonal elements of $\frac{1}{n}\hat{I}^{-1}(\theta, \lambda)$, respectively. Now, the modified asymptotic $100(1-\alpha)\%$ two-sided equi-tailed confidence intervals (MATE CIs) for θ and λ are given by

$$\left(\max\left\{0, \hat{\theta}_{MLE} - z_{\frac{\alpha}{2}}\sqrt{\widehat{Var}(\hat{\theta}_{MLE})}\right\},\ \hat{\theta}_{MLE} + z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_{MLE})}\right),$$

and

$$\left(\max\left\{0, \hat{\lambda}_{MLE} - z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\lambda}_{MLE})}\right\},\ \hat{\lambda}_{MLE} + z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\lambda}_{MLE})}\right),$$

respectively, where $z_{\alpha/2}$ is the right $(\alpha/2)$-th quantile of the standard normal distribution.

Let

$$J^{-1}(\theta, \lambda) = \begin{bmatrix} -\frac{\partial^2 \ell(\theta,\lambda)}{\partial \theta^2} & -\frac{\partial^2 \ell(\theta,\lambda)}{\partial \theta \partial \lambda} \\ -\frac{\partial^2 \ell(\theta,\lambda)}{\partial \lambda \partial \theta} & -\frac{\partial^2 \ell(\theta,\lambda)}{\partial \lambda^2} \end{bmatrix}^{-1},$$

where

$$-\frac{\partial^2 \ell(\theta,\lambda)}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{ne^{-\theta}}{(1-e^{-\theta})^2}, \qquad -\frac{\partial^2 \ell(\theta,\lambda)}{\partial \theta \partial \lambda} = -\sum_{i=1}^{n} x_i e^{-\lambda x_i}, \qquad -\frac{\partial^2 \ell(\theta,\lambda)}{\partial \lambda^2} = \frac{n}{\lambda^2} + \theta \sum_{i=1}^{n} x_i^2 e^{-\lambda x_i}.$$

Suppose that $\hat{J}^{-1}(\theta, \lambda)$ is the estimate of $J^{-1}(\theta, \lambda)$ being obtained by replacing the unknown parameters appearing in the elements of $J^{-1}(\theta, \lambda)$ with their corresponding ML estimates. In practical situations, we can use $\hat{J}^{-1}(\theta, \lambda)$ instead of the observed matrix $\frac{1}{n}\hat{I}^{-1}(\theta, \lambda)$.

3. Bayesian Estimation

In the Bayesian analysis, it is assumed that the unknown parameters are random variables with a joint prior density function. The prior density function and its hyperparameters can be determined based on the past knowledge and experience. When no prior knowledge is available, non-informative priors can be implemented for the purpose of Bayesian inference. Here, we assume that θ and λ are independent gamma random variables with prior densities as follows (see Arabi Belaghi et al. [1]):

$$\pi(\theta|a_1,b_1)=\frac{b_1^{a_1}\theta^{a_1-1}e^{-b_1\theta}}{\Gamma(a_1)} \quad \text{and} \quad \pi(\lambda|a_2,b_2)=\frac{b_2^{a_2}\lambda^{a_2-1}e^{-b_2\lambda}}{\Gamma(a_2)} \tag{4}$$

where the hyperparameters a_1, b_1, a_2 and b_2 are all positive and known. The gamma prior possesses explicit expressions for its mean and variance and consequently we can include our prior information about the mean and/or variance of the parameter easier. From (4), the joint prior pdf of θ and λ is

$$\pi(\theta,\lambda)=\pi(\lambda|a_2,b_2)\,\pi(\theta|a_1,b_1),\;\; \theta>0,\;\; \lambda>0. \tag{5}$$

Let $\boldsymbol{x}=(x_1,\cdots,x_n)$ be an observed vector of a random sample of size n from PE(θ,λ). Then, from (1) and (5), the joint posterior pdf of θ and λ given $\boldsymbol{x}$ is given by

$$\pi(\theta,\lambda|\boldsymbol{x})=\frac{\theta^{n+a_1-1}\lambda^{n+a_2-1}e^{-\theta\left(b_1+\sum_{i=1}^n e^{-\lambda x_i}\right)-\lambda\left(b_2+\sum_{i=1}^n x_i\right)}}{\mathrm{C}_0(1-e^{-\theta})^n}, \tag{6}$$

where

$$C_0=\int_0^\infty\int_0^\infty \frac{\theta^{n+a_1-1}\lambda^{n+a_2-1}e^{-\theta\left(b_1+\sum_{i=1}^n e^{-\lambda x_i}\right)-\lambda\left(b_2+\sum_{i=1}^n x_i\right)}}{(1-e^{-\theta})^n}\mathrm{d}\theta\mathrm{d}\lambda.$$

Here, we derive the Bayes estimates under the SEL function, which is symmetric. Let $\tilde{\theta}$ be an estimator of parameter θ. Then, the SEL function is defined as $Loss_1(\tilde{\theta},\theta)=(\tilde{\theta}-\theta)^2$. The Bayesian point estimate under the SEL function is the posterior mean of the parameter given the informative data. Therefore, in our case, the Bayes estimates of θ and λ are (provided that they exist)

$$\tilde{\theta}_S=E(\theta|\boldsymbol{x})=\int_0^\infty\int_0^\infty \frac{\theta^{n+a_1}\lambda^{n+a_2-1}e^{-\theta\left(b_1+\sum_{i=1}^n e^{-\lambda x_i}\right)-\lambda\left(b_2+\sum_{i=1}^n x_i\right)}}{\mathrm{C}_0(1-e^{-\theta})^n}\mathrm{d}\lambda\mathrm{d}\theta, \tag{7}$$

$$\tilde{\lambda}_S=E(\lambda|\boldsymbol{x})=\int_0^\infty\int_0^\infty \frac{\theta^{n+a_1-1}\lambda^{n+a_2}e^{-\theta\left(b_1+\sum_{i=1}^n e^{-\lambda x_i}\right)-\lambda\left(b_2+\sum_{i=1}^n x_i\right)}}{\mathrm{C}_0(1-e^{-\theta})^n}\mathrm{d}\lambda\mathrm{d}\theta, \tag{8}$$

respectively.

It seems that the integrals (7) and (8) do not have explicit forms and therefore numerical techniques may be used to approximate these integrals. In what follows, we explain three procedures for approximating these integrals, namely the Tierney Kadane approximation method, the importance sampling method and the Metropolis-Hastings within Gibbs method.

3.1. Tierney Kadane Approximation Method

The Tierney Kadane approximation method was proposed by Tierney and Kadane [23]. We want to approximate $E(u(\theta,\lambda)|\boldsymbol{x})$. In this regard, we define the functions L_0 and L_*, respectively, as follows:

$$L_0(\theta,\lambda)=\frac{1}{n}\left[\log\big(L(\theta,\lambda|\boldsymbol{x})\big)+\log(\pi(\theta,\lambda))\right],$$
$$L_*(\theta,\lambda)=L_0(\theta,\lambda)+\frac{1}{n}\log(u(\theta,\lambda)).$$

Assume that $(\hat{\theta}_0,\hat{\lambda}_0)$ and $(\hat{\theta}_*,\hat{\lambda}_*)$ are the maximizing points of $L_0(\theta,\lambda)$ and $L_*(\theta,\lambda)$, respectively, such that $\hat{\theta}_0$ and $\hat{\lambda}_0$ are the unique maximizers of $L_0(\theta,\lambda)$ and are obtained by solving the equations $\frac{L_0(\theta,\lambda)}{\partial\theta}=0$ and $\frac{L_0(\theta,\lambda)}{\partial\lambda}=0$ simultaneously and $\hat{\theta}_*$ and $\hat{\lambda}_*$ are also obtained by solving the equations $\frac{L_*(\theta,\lambda)}{\partial\theta}=0$ and $\frac{L_*(\theta,\lambda)}{\partial\lambda}=0$ simultaneously. Assume further that Σ^0 and Σ^* are minus the inverse Hessians of L_0 and L_* at $(\hat{\theta}_0,\hat{\lambda}_0)$ and $(\hat{\theta}_*,\hat{\lambda}_*)$, respectively. Then, the approximate value of $E(u(\theta,\lambda)|\boldsymbol{x})$ is obtained from the following relation

$$\hat{E}(u(\theta,\lambda)|\boldsymbol{x})=\sqrt{\frac{|\Sigma^*|}{|\Sigma^0|}}\exp\left[n\left(L_*(\hat{\theta}_*,\hat{\lambda}_*)-L_0(\hat{\theta}_0,\hat{\lambda}_0)\right)\right].$$

In our case for the PE distribution, we have

$$L_0(\theta,\lambda)=\log(\theta\lambda)-\frac{\lambda}{n}\sum_{i=1}^{n}x_i-\frac{\theta}{n}\sum_{i=1}^{n}e^{-\lambda x_i}-\log\big(1-e^{-\theta}\big)$$
$$+\frac{(a_1-1)\log(\theta)+(a_2-1)\log(\lambda)-\theta b_1-\lambda b_2}{n}+A,$$

where

$$A=\frac{a_1\log(b_1)+a_2\log(b_2)-\log\big(\Gamma(a_1)\big)-\log\big(\Gamma(a_2)\big)}{n}.$$

Thus, $\hat{\theta}_0$ and $\hat{\lambda}_0$ are obtained by solving the following nonlinear equations simultaneously:

$$\frac{\partial L_0(\theta,\lambda)}{\partial\theta}=\frac{n+a_1-1}{n\theta}-\frac{e^{-\theta}}{1-e^{-\theta}}-\frac{b_1+\sum_{i=1}^{n}e^{-\lambda x_i}}{n}=0,$$
$$\frac{\partial L_0(\theta,\lambda)}{\partial\lambda}=\frac{n+a_2-1}{n\lambda}+\frac{\theta\sum_{i=1}^{n}x_ie^{-\lambda x_i}-\sum_{i=1}^{n}x_i-b_2}{n}=0.$$

Now, to approximate the Bayes estimate of θ under the SEL function, i.e., $E(\theta|\boldsymbol{x})$, we set $u(\theta,\lambda)=\theta$. As a result, we have

$$L_{*S}(\theta,\lambda)=L_0(\theta,\lambda)+\frac{1}{n}\log(\theta).$$

Then, $\hat{\theta}_{*S}$ and $\hat{\lambda}_{*S}$ are obtained by solving the following nonlinear equations simultaneously:

$$\frac{\partial L_{*S}(\theta,\lambda)}{\partial\theta}=\frac{n+a_1}{n\theta}-\frac{e^{-\theta}}{1-e^{-\theta}}-\frac{b_1+\sum_{i=1}^{n}e^{-\lambda x_i}}{n}=0,$$

$$\frac{\partial L_{*S}(\theta,\lambda)}{\partial\lambda}=\frac{n+a_2-1}{n\lambda}+\frac{\theta\sum_{i=1}^{n}x_i e^{-\lambda x_i}-\sum_{i=1}^{n}x_i-b_2}{n}=0.$$

Let

$$A_{11}(\theta,\lambda)=-\frac{\partial^2 L_0(\theta,\lambda)}{\partial\theta^2}=\frac{n+a_1-1}{n\theta^2}-\frac{e^{-\theta}}{(1-e^{-\theta})^2},$$
$$A_{12}(\theta,\lambda)=-\frac{\partial^2 L_0(\theta,\lambda)}{\partial\theta\partial\lambda}=-\frac{\sum_{i=1}^{n}x_i e^{-\lambda x_i}}{n},$$
$$A_{22}(\theta,\lambda)=-\frac{\partial^2 L_0(\theta,\lambda)}{\partial\lambda^2}=\frac{n+a_2-1}{n\lambda^2}+\frac{\theta\sum_{i=1}^{n}x_i^2 e^{-\lambda x_i}}{n},$$
$$A_{11}^{*}(\theta,\lambda)=-\frac{\partial^2 L_{*S}(\theta,\lambda)}{\partial\theta^2}=\frac{n+a_1}{n\theta^2}-\frac{e^{-\theta}}{(1-e^{-\theta})^2}.$$

The approximate Bayes estimate of θ under the SEL function based on the Tierney Kadane method is given by

$$\tilde{\theta}_{T-K}=\sqrt{\frac{A_{11}(\hat{\theta}_0,\hat{\lambda}_0)A_{22}(\hat{\theta}_0,\hat{\lambda}_0)-\left[A_{12}(\hat{\theta}_0,\hat{\lambda}_0)\right]^2}{A_{11}^{*}(\hat{\theta}_{*S},\hat{\lambda}_{*S})A_{22}(\hat{\theta}_{*S},\hat{\lambda}_{*S})-\left[A_{12}(\hat{\theta}_{*S},\hat{\lambda}_{*S})\right]^2}}\exp\left[n\left(L_{*S}(\hat{\theta}_{*S},\hat{\lambda}_{*S})-L_0(\hat{\theta}_0,\hat{\lambda}_0)\right)\right].$$

Similarly, we can derive the approximate Bayes estimates of λ under the SEL function based on the Tierney Kadane approximation method.

3.2. Importance Sampling Method

Another well-known method for approximating Bayesian estimates is called the importance sampling method. As mentioned earlier, this method was also used by Arabi Belaghi et al. [1] to approximate the Bayesian estimates of parameters of the PE distribution based on type-II censored data. Given the vector of random sample $\boldsymbol{x}$, the joint posterior density function of θ and λ, relation (6), can be rewritten as

$$\pi(\theta,\lambda|\boldsymbol{x})=g_1(\lambda|\boldsymbol{x})g_2(\theta|\lambda,\boldsymbol{x})h(\theta,\lambda,\boldsymbol{x}),$$

where $g_1(\lambda|\boldsymbol{x})$ is a gamma density function with parameters $n+a_2$ and $b_2+\sum_{i=1}^{n}x_i$, $g_2(\theta|\lambda,\boldsymbol{x})$ is a gamma density function with parameters $n+a_1$ and $b_1+\sum_{i=1}^{n}\mathrm{e}^{-\lambda x_i}$ and $h(\theta,\lambda,\boldsymbol{x})$ is given by

$$h(\theta,\lambda,\boldsymbol{x})=\frac{\Gamma(n+a_1)\Gamma(n+a_2)}{\left(b_1+\sum_{i=1}^{n}\mathrm{e}^{-\lambda x_i}\right)^{n+a_1}\left(b_2+\sum_{i=1}^{n}x_i\right)^{n+a_2}C_0(1-\mathrm{e}^{-\theta})^{n}}.$$

Now, the Bayesian estimates of the parameters can be approximated using Algorithm 1 below.

Algorithm 1:
Step 1. Generate λ_1 from $g_1(\lambda|\boldsymbol{x})$ and then given λ_1, generate θ_1 from $g_2(\theta|\lambda_1,\boldsymbol{x})$.
Step 2. Repeat Step 1, M times to obtain $(\theta_1,\lambda_1),\cdots,\ (\theta_M,\lambda_M)$, where M is a large number.
Step 3. The approximate Bayes estimates of θ and λ under the SEL function are given by

$$\tilde{\theta}_{I-S} = \frac{\sum_{i=1}^{M} \theta_i h(\theta_i,\lambda_i,\boldsymbol{x})}{\sum_{i=1}^{M} h(\theta_i,\lambda_i,\boldsymbol{x})}, \quad \text{and} \quad \tilde{\lambda}_{I-S} = \frac{\sum_{i=1}^{M} \lambda_i h(\theta_i,\lambda_i,\boldsymbol{x})}{\sum_{i=1}^{M} h(\theta_i,\lambda_i,\boldsymbol{x})},$$

respectively.

Equivalently, the approximate Bayes estimates of θ and λ under the SEL function based on the importance sampling method can be rewritten as

$$\tilde{\theta}_{I-S} = \frac{\sum_{i=1}^{M} \theta_i h^*(\theta_i,\lambda_i,\boldsymbol{x})}{\sum_{i=1}^{M} h^*(\theta_i,\lambda_i,\boldsymbol{x})}, \quad \text{and} \quad \tilde{\lambda}_{I-S} = \frac{\sum_{i=1}^{M} \lambda_i h^*(\theta_i,\lambda_i,\boldsymbol{x})}{\sum_{i=1}^{M} h^*(\theta_i,\lambda_i,\boldsymbol{x})},$$

respectively, where

$$h^*(\theta,\lambda,\boldsymbol{x}) = \frac{1}{\left(b_1 + \sum_{i=1}^{n} \mathrm{e}^{-\lambda x_i}\right)^{n+a_1} (1-\mathrm{e}^{-\theta})^n}.$$

3.3. Metropolis-Hastings within Gibbs Method

The Metropolis-Hastings algorithm is one of the most popular methods being applied by researchers to deal with complicated integrals related to Bayesian inference. The algorithm was first proposed by Metropolis et al. [13] and then generalized by Hastings [7]. Using a Metropolis-Hastings algorithm, one tries to simulate a Markov chain whose stationary distribution is approximately the same as the posterior distribution of interest. To this end, the algorithm generates numbers from a proposal distribution and updates them step by step so that the generated sample can be considered sensibly as a sample from the target posterior distribution. Here, since we have two parameters, the univariate proposal distribution of each parameter can be updated at a time by the newest generated value of that parameter and thus, we use a method that is somehow similar to the Gibbs sampling. Therefore, the method is called the "Metropolis-Hastings within Gibbs method", but note that it is naturally different from the Gibbs sampling; see, for example, Carlin and Louis [3] for more details.

Louzada-Neto et al. [12], Singh et al. [22] and Tomazella et al. [24] applied the Metropolis-Hastings within Gibbs sampling method in order to approximate the Bayes estimates of the parameters of the PE distribution. Louzada-Neto et al. [12] and Tomazella et al. [24] utilized a transformation procedure to apply the Metropolis-Hastings within Gibbs more efficiently. Here, we follow their procedure and describe the procedure in detail. The conditional posterior density of θ given λ and $\boldsymbol{x}$ is

$$\pi(\theta|\lambda,\boldsymbol{x}) \propto \frac{\theta^{n+a_1-1} e^{-\theta\left(b_1+\sum_{i=1}^{n} e^{-\lambda x_i}\right)}}{(1-e^{-\theta})^n}.$$

Moreover, the conditional posterior density of λ given θ and $\boldsymbol{x}$ is

$$\pi(\lambda|\theta,\boldsymbol{x}) \propto \lambda^{n+a_2-1} e^{-\theta \sum_{i=1}^{n} e^{-\lambda x_i} - \lambda\left(b_2+\sum_{i=1}^{n} x_i\right)}.$$

These conditional densities do not belong to the known distributions (note that if the conditional distributions belong to the known distributions and we could generate samples directly from them, then there will be no need to use the Metropolis-Hastings algorithm and we can use the Gibbs sampling procedure). Now, let $\eta = \log(\theta)$ and $\gamma = \log(\lambda)$, where $-\infty < \eta < \infty$ and $-\infty < \gamma < \infty$. The joint distribution of η and γ is given by

$$\pi(\eta,\gamma|\boldsymbol{x}) = \frac{e^{\eta(n+a_1)+\gamma(n+a_2)}e^{-e^{\gamma}\left(b_2+\sum_{i=1}^{n}x_i\right)}e^{-e^{\eta}\left(b_1+\sum_{i=1}^{n}e^{-e^{\gamma}x_i}\right)}}{C_0(1-e^{-e^{\eta}})^n}.$$

Thus, the conditional densities of η and γ are given by

$$\pi(\eta|\gamma,\boldsymbol{x}) \propto \frac{e^{\eta(n+a_1)}e^{-e^{\eta}\left(b_1+\sum_{i=1}^{n}e^{-e^{\gamma}x_i}\right)}}{(1-e^{-e^{\eta}})^n},$$

and

$$\pi(\gamma|\eta,\boldsymbol{x}) \propto e^{\gamma(n+a_2)}e^{-e^{\gamma}\left(b_2+\sum_{i=1}^{n}x_i\right)}e^{-e^{\eta}\sum_{i=1}^{n}e^{-e^{\gamma}x_i}},$$

respectively.

Following Tomazella et al. [24], we use the random walk procedure to generate samples from $\pi(\eta,\gamma|\boldsymbol{x})$. We can also use the thinning approach, namely we discard all the generated samples but every k-th generated pairs of numbers, to reduce the autocorrelation of the chain. So, we employ Algorithm 2 below.

Algorithm 2:
Step 1. Start with a guess vector (η_0,γ_0) and set $q=1$.
Step 2. Generate a random normal variate Z with the mean zero and variance τ_1^2. Given the vector $(\eta_{q-1},\gamma_{q-1})$, set $\eta_q=\eta_{q-1}+\sigma Z$ with probability

$$P=\min\left\{1,\frac{\pi(\eta_q|\gamma_{q-1},\boldsymbol{x})}{\pi(\eta_{q-1}|\gamma_{q-1},\boldsymbol{x})}\right\};$$

Otherwise, set $\eta_q=\eta_{q-1}$.
Step 3. Generate another random normal variate Z with the mean zero and variance τ_2^2. Given the vector (η_q,γ_{q-1}), set $\gamma_q=\gamma_{q-1}+\sigma Z$ with probability

$$P=\min\left\{1,\frac{\pi(\gamma_q|\eta_q,\boldsymbol{x})}{\pi(\gamma_{q-1}|\eta_q,\boldsymbol{x})}\right\};$$

Otherwise, set $\gamma_q=\gamma_{q-1}$.
Step 4. Set $q=q+1$ and repeat Steps 2 and 3, N times, where N is a large number.
Step 5. For $q=1,\dots,N$, set $\theta_q=e^{\eta_q}$ and $\lambda_q=e^{\gamma_q}$. Then, the generated sample from $\pi(\theta,\lambda|\boldsymbol{x})$ will be $\left((\theta_{T+1},\lambda_{T+1}),(\theta_{T+k+1},\lambda_{T+k+1}),\dots,(\theta_{T+k(N'-1)+1},\lambda_{T+k(N'-1)+1})\right)$, where T is a burn-in period, k is the thinning parameter and N' is the size of the generated sample. Let us denote the generated sample by $\left((\theta_1',\lambda_1'),(\theta_2',\lambda_2'),\dots,(\theta_{N'}',\lambda_{N'}')\right)$, for the sake of simplicity.

Here, we set $(\eta_0,\gamma_0)=\left(\log(\hat{\theta}),\log(\hat{\lambda})\right)$, where $\hat{\theta}$ and $\hat{\lambda}$ are the ML estimates of θ and λ, respectively. In addition, we take the scale parameter σ to be 2 (see Tomazella et al. [24]). We discuss how to choose τ_1^2 and τ_2^2 in the Appendix. Now, given the generated sample $\left((\theta_1',\lambda_1'),(\theta_2',\lambda_2'),\dots,(\theta_{N'}',\lambda_{N'}')\right)$, the approximate Bayes of θ and λ under the SEL function based on the Metropolis-Hastings within Gibbs method are given by

$$\tilde{\theta}_{MHGibbs}=\frac{1}{N'}\sum_{i=1}^{N'}\theta_i', \quad \text{and} \quad \tilde{\lambda}_{MHGibbs}=\frac{1}{N'}\sum_{i=1}^{N'}\lambda_i',$$

respectively.

We can also find credible intervals for θ and λ using the method described by Chen and Shao [4]. To this end, first sort the generated sample $(\theta'_1, \dots, \theta'_{N'})$ to get $\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(N')}$. Construct the intervals $(\theta_{(j)},\ \theta_{(j+[(1-\alpha)N'])})$, for $j = 1, \dots, N' - [(1-\alpha)N']$, where $[x]$ is the integer part of x. Then, the Chen and Shao shortest credible interval (CSS CrI) for θ is given by the shortest interval among all the $(\theta_{(j)},\ \theta_{(j+[(1-\alpha)N'])})$. Similarly, we can find the CSS CrI for λ.

4. Simulation

Our main goal here, as pointed out earlier, is to compare three methods of approximating the integrals that are related to Bayes estimates of the parameters of the PE distribution. These methods are the Tierney Kadane method, importance sampling procedure and the Metropolis-Hastings within Gibbs algorithm, as described earlier. We also compare MATE CIs with CSS CrIs in terms of average width (AW) and coverage probability (CP). To this end, we conduct a simulation study. We consider two sample sizes $n = 70$ and 150. We also use the standard exponential priors (the standard exponential distribution is a special case of the gamma distribution), namely we set $a_1 = b_1 = a_2 = b_2 = 1$. We take the values of the scale parameter to be $\lambda = 0.5, 1$ and 2, and the values of the shape parameter to be $\theta = 3$ and 5. For the importance sampling method, we take $M = 3000$ which seems to be sufficiently large. In applying the Metropolis-Hastings within Gibbs method, we implemented the thinning technique. For the Metropolis-Hastings within Gibbs method, the settings are as follows: $N = 20000k + T$, where $T = 2000$ is the burn-in period and k, the thinning parameter, is mostly taken to be 2 but also taken to be 3 or 4 for some iterations. So, the size of each generated Markov chain Monte Carlo (MCMC) sample is $N' = 20000$. In addition, the convergence of the MCMC samples generated by the Metropolis-Hastings within Gibbs method are checked using the Geweke's test (see Geweke [5]), Raftery and Lewis's diagnostic (see Raftery and Lewis [17,18]) and Heidelberger and Welch's convergence diagnostic (see Heidelberger and Welch [10]). It is worth mentioning that Heidelberger and Welch [10] combined the works of Heidelberger and Welch [8], Heidelberger and Welch [9], Schruben [19], Schruben et al. [20] and Schruben et al. [21]. The results of our simulation are based on $J = 1000$ Monte Carlo replicates. The generated Metropolis-Hastings within Gibbs Markov chains for $\theta = 3$, $\lambda = 2$ and $n = 70$ for one of the iterations are given in Figure 2, from which we can verify the convergence.

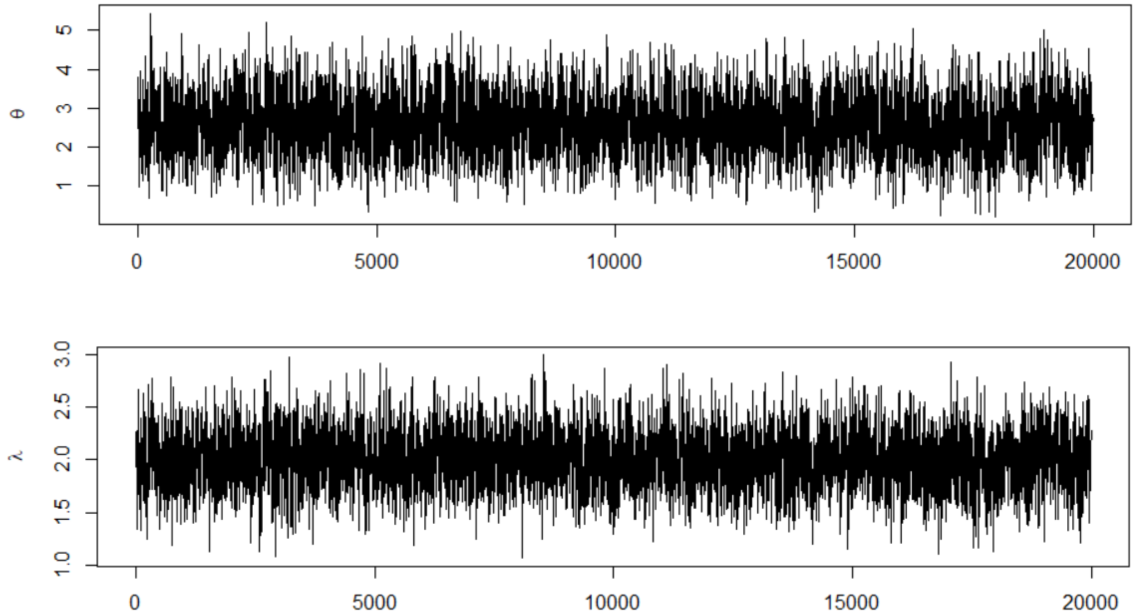

Figure 2. Metropolis-Hastings within Gibbs Markov chains for $\theta = 3, \lambda = 2$ and $n = 70$.

Assume that $\hat{\theta}$ is an estimator of θ and $\hat{\theta}_i$ is the corresponding estimate of θ obtained in the i-th replication. Then, the estimated mean squared error (EMSE) of $\hat{\theta}$ is given by

$$EMSE(\hat{\theta}) = \frac{1}{J}\sum_{i=1}^{J}\left(\hat{\theta}_i - \theta\right)^2.$$

Similarly, we can define the EMSE of an estimator of λ. The simulation results regarding the point estimators are presented in tables 1 and 2 and the ones regarding the interval estimators are given in Table 3. From tables 1 and 2, we observe that there exist some cases in which the Bayes estimates being approximated based on the Tierney Kadane method perform better than the other estimates. This is also true for the other estimates specially for the ML estimates. From these tables, we cannot draw a general conclusion for identifying the best estimator. We cannot say which approximation is the best either, as we must compare the approximate Bayes estimates with the true values of the integrals not the true values of the parameters. From Table 3, we see that CSS CrIs have smaller AWs than the MATE CIs for most cases but MATE CIs have larger CPs in all the cases.

Table 1. The EMSEs of the point estimators of θ and λ when $n = 70$.

Point Estimators of θ				
(θ, λ)	ML Estimator	Tierney Kadane	Importance Sampling	Metropolis-Hastings within Gibbs
(3,0.5)	0.66491	0.72451	0.51979	0.71640
(3,1)	0.66405	0.79023	0.53763	0.76780
(3,2)	0.71863	0.91876	0.56926	0.89359
(5,0.5)	1.33691	1.13655	4.89455	1.14446
(5,1)	1.17753	1.19437	5.12868	1.20097
(5,2)	1.37910	1.18567	5.05687	1.19276
Point Estimators of λ				
(θ, λ)	ML Estimator	Tierney Kadane	Importance Sampling	Metropolis-Hastings within Gibbs
(3,0.5)	0.00400	0.00445	0.01274	0.00421
(3,1)	0.01543	0.24707	0.05317	0.01743
(3,2)	0.06805	0.08934	0.22127	0.08233
(5,0.5)	0.00317	0.00315	0.03015	0.00315
(5,1)	0.01104	0.01218	0.12250	0.01219
(5,2)	0.04931	0.05353	0.50689	0.05352

Table 2. The EMSEs of the point estimators of θ and λ when $n = 150$.

Point Estimators of θ				
(θ, λ)	ML Estimator	Tierney Kadane	Importance Sampling	Metropolis-Hastings within Gibbs
(3,0.5)	0.26950	0.29279	0.47261	0.29349
(3,1)	0.26990	0.30102	0.47837	0.30360
(3,2)	0.27709	0.30592	0.49955	0.30914
(5,0.5)	0.56022	0.51682	5.51200	0.51747
(5,1)	0.47218	0.45341	5.58900	0.45528
(5,2)	0.54667	0.51414	5.60340	0.51523
Point Estimators of λ				
(θ, λ)	ML Estimator	Tierney Kadane	Importance Sampling	Metropolis-Hastings within Gibbs
(3,0.5)	0.00173	0.00187	0.02017	0.00182
(3,1)	0.00744	0.00791	0.08136	0.00792
(3,2)	0.02923	0.03131	0.33416	0.03132
(5,0.5)	0.00126	0.00127	0.04225	0.00127
(5,1)	0.00496	0.00503	0.16982	0.00499
(5,2)	0.02157	0.02225	0.68597	0.02231

Table 3. The AWs and CPs of the interval estimators of θ and λ.

Interval Estimators of θ					
		$n = 70$		$n = 150$	
(θ, λ)		MATE CI	CSS CrI	MATE CI	CSS CrI
(3,0.5)	AW	3.0983	2.9866	2.0776	2.0550
	CP	0.9600	0.8840	0.9660	0.9390
(3,1)	AW	3.0998	2.9904	2.0752	2.0529
	CP	0.9590	0.8730	0.9620	0.9320
(3,2)	AW	3.0982	2.9693	2.0809	2.0584
	CP	0.9540	0.8350	0.9610	0.9340
(5,0.5)	AW	4.1709	3.6469	2.7517	2.5801
	CP	0.9660	0.8680	0.9440	0.9160
(5,1)	AW	4.1190	3.6151	2.7495	2.5769
	CP	0.9640	0.8820	0.9630	0.9190
(5,2)	AW	4.2164	3.6454	2.7572	2.5713
	CP	0.9590	0.8720	0.9520	0.9160
Interval Estimators of λ					
		$n = 70$		$n = 150$	
(θ, λ)		MATE CI	CSS CrI	MATE CI	CSS CrI
(3,0.5)	AW	0.2454	0.2470	0.1670	0.1680
	CP	0.9590	0.9400	0.9560	0.9540
(3,1)	AW	0.4929	0.4960	0.3340	0.3358
	CP	0.9540	0.9180	0.9520	0.9400
(3,2)	AW	0.9860	0.9878	0.6684	0.6724
	CP	0.9430	0.8920	0.9520	0.9380
(5,0.5)	AW	0.2073	0.2022	0.1402	0.1381
	CP	0.9410	0.9200	0.9540	0.9420
(5,1)	AW	0.4156	0.4059	0.2809	0.2767
	CP	0.9580	0.9310	0.9640	0.9400
(5,2)	AW	0.8297	0.8057	0.5612	0.5519
	CP	0.9480	0.9080	0.9500	0.9330

5. Concluding Remarks

We discussed three approximation methods for approximating the Bayesian estimates of the parameters of the PE distribution and then presented a simulation study for the purpose of comparison. We also compared the classical and Bayesian interval estimators. We could not draw a general conclusion that which one of the approximation methods always performed the best. Our computations were performed using Maple 17 and R [16] using the packages *nleqslv* (see Hasselman [6]) and *coda* (see Plummer et al. [14] and Plummer et al. [15]) in R.

Acknowledgements

The authors thank the referees for their valuable comments.

References

[1] Arabi Belaghi, R., Noori Asl, M., Gurunlu Alma, O., Singh, S. and Vasfi, M. (2019), Estimation and prediction for the Poisson-exponential distribution based on type-II censored data, *American Journal of Mathematical and Management Sciences*, 38(1), 96-115.

[2] Cancho, V.G., Louzada-Neto, F. and Barriga, G.D.C. (2011), The Poisson-exponential lifetime distribution, *Computational Statistics and Data Analysis*, 55, 677-686.

[3] Carlin, B.P. and Louis, T.A. (2000), Bayes and empirical Bayes methods for data analysis, 2nd ed., Chapman & Hall/CRC, Boca Raton.

[4] Chen, M-H and Shao, Q-M (1999), Monte Carlo estimation of Bayesian credible and HPD intervals, Journal of Computational and Graphical Statistics, 8(1), 69-92.

[5] Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith A.F.M. (Ed), Bayesian Statistics 4, Clarendon Press, Oxford, UK, pp. 169-193.

[6] Hasselman, B. (2018), nleqslv: Solve systems of nonlinear equations, R package version 3.3.2, https://CRAN.R-project.org/package=nleqslv.

[7] Hastings, W.K. (1970), Monte Carlo sampling methods using Markov chains and their applications, Biometrika, 57(1), 97-109

[8] Heidelberger, P. and Welch, P.D. (1981), A spectral method for confidence interval generation and run length control in simulations, *Communications of the ACM*, 24(٤), 233-245.

[9] Heidelberger, P. and Welch, P.D. (1981), Adaptive spectral methods for simulation output analysis, *IBM Journal of Research and Development*, 25(٦), 860-876.

[10] Heidelberger, P. and Welch, P.D. (1983), Simulation run length control in the presence of an initial transient, *Operations Research*, 31(6), 1109-1144.

[11] Lehmann, E.L. and Casella, G. (1998), Theory of point estimation, 2nd ed., Springer-Verlag, New York.

[12] Louzada-Neto, F., Cancho, V.G. and Barriga, G.D.C. (2011), The Poisson-exponential distribution: a Bayesian approach, *Journal of Applied Statistics*, 38(6), 1239-1248.

[13] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953) , Equation of state calculations by fast computing machines, Journal of Chemical Physics, 21(6), 1087-1092

[14] Plummer, M., Best, N., Cowles, K. and Vines, K. (2006), CODA: Convergence diagnosis and output analysis for MCMC, *R News*, 6(1), 7-11.

[15] Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., Almond, R. and Magnusson, A. (2019), coda: Output analysis and diagnostics for MCMC, R package version 0.19-3, https:// CRAN.R-project.org/package=coda.

[16] R Core Team (2020), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.

[17] Raftery, A.E. and Lewis, S.M. (1992), Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo, *Statistical Science*, 7(٤), 493-497.

[18] Raftery, A.E. and Lewis, S.M. (1996), Implementing MCMC, In: Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.)Ed), Markov chain Monte Carlo in practice, Chapman & Hall/CRC, Boca Raton, pp. 115-130.

[19] Schruben, L.W. (1982), Detecting initialization bias in simulation output, *Operations Research*, 30(3), 569-590.

[20] Schruben, L., Singh, H. and Tierney, L. (1980), A test of initialization bias hypotheses in simulation output, Technical Report 471, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York.

[21] Schruben, L., Singh, H. and Tierney, L. (1983), Optimal tests for initialization bias in simulation output, *Operations Research*, 31(٦), 1167-1178.

[22] Singh, S.K., Singh, U. and Kumar, M. (2014), Estimation for the parameter of Poisson-exponential distribution under Bayesian paradigm, *Journal of Data Science*, 12(1), 157-173.

[23] Tierney, L. and Kadane, J.B. (1986), Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, 81(393), 82-86.

[24] Tomazella, V.L.D., Cancho, V.G. and Louzada, F. (2013), Bayesian reference analysis for the Poisson-exponential lifetime distribution, *Chilean Journal of Statistics*, 4(1), 99-113.

Appendix

Here, we discuss how to choose τ_1^2 and τ_2^2 in Algorithm 2. Suppose that η_* and γ_* maximize $\pi(\eta,\gamma|\boldsymbol{x})$. The logarithm of $\pi(\eta,\gamma|\boldsymbol{x})$ is given by

$$\log\pi(\eta,\gamma|\boldsymbol{x}) = -\log C_0 + \eta(n+a_1) + \gamma(n+a_2) - e^{\gamma}\left(b_2 + \sum_{i=1}^{n} x_i\right) - e^{\eta}\left(b_1 + \sum_{i=1}^{n} e^{-e^{\gamma}x_i}\right) - n\log\left(1 - e^{-e^{\eta}}\right).$$

Solving the following two nonlinear equations help us find η_* and γ_*:

$$\frac{\partial \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\eta} = n + a_1 - e^{\eta}\left(b_1 + \sum_{i=1}^{n} e^{-e^{\gamma}x_i}\right) - \frac{ne^{\eta-e^{\eta}}}{1-e^{-e^{\eta}}} = 0,$$
$$\frac{\partial \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\gamma} = n + a_2 - e^{\gamma}\left(b_2 + \sum_{i=1}^{n} x_i\right) + e^{\eta+\gamma}\sum_{i=1}^{n} x_i e^{-e^{\gamma}x_i} = 0.$$

Define matrix V as follows:

$$V = \begin{bmatrix} -\frac{\partial^2 \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\eta^2} & -\frac{\partial^2 \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\eta\partial\gamma} \\ -\frac{\partial^2 \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\gamma\partial\eta} & -\frac{\partial^2 \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\gamma^2} \end{bmatrix}_{(\eta,\gamma)=(\eta_*,\gamma_*)},$$

where

$$-\frac{\partial^2 \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\eta^2} = e^{\eta}\left(b_1 + \sum_{i=1}^{n} e^{-e^{\gamma}x_i}\right) + \frac{n\left(1-e^{\eta}-e^{-e^{\eta}}\right)e^{\eta-e^{\eta}}}{(1-e^{-e^{\eta}})^2},$$
$$-\frac{\partial^2 \log\pi(\eta,\gamma|\boldsymbol{x})}{\partial\gamma^2} = e^{\gamma}\left(b_2 + \sum_{i=1}^{n} x_i\right) - e^{\eta}\sum_{i=1}^{n}(1 - x_i e^{\gamma})x_i\, e^{\gamma - x_i e^{\gamma}},$$

$$-\frac{\partial \log \pi(\eta,\gamma|\boldsymbol{x})}{\partial\eta\partial\gamma} = -e^{\eta}\sum_{i=1}^{n} x_i\, e^{\gamma - x_i e^{\gamma}}.$$

We choose τ_1^2 and τ_2^2 as the first and second diagonal elements of V^{-1}, respectively, where V^{-1} is the inverse matrix of V (see Tomazella et al. [24]).