Revista Colombiana de Estadística - Applied Statistics
January 2023, volume 46, issue 1, pp. 93–110
http://doi.org/10.15446/rce.v46n1.95989

Objective Prior Distributions to Estimate the Parameters of the Poisson-Exponential Distribution

Distribuciones previas objetivas para estimar los parámetros de la distribución Poisson-Exponencial

Fernando A. Moala[a], Gustavo Moraes[b]

Departamento de Estatística, Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, São Paulo, Brasil

Abstract

In this paper, a set of important objective priors are examined for the Bayesian estimation of the parameters present in the Poisson-Exponential distribution PE. We derived the multivariate Jeffreys prior and the Maximal Data Information Prior. Reference prior and others priors proposed in the literature are also analyzed. We show that the posterior densities resulting from these approaches are proper although the respective priors are improper. Monte Carlo simulations are used to compare the efficiencies and to assess the sensitivity of the choice of the priors, mainly for small sample sizes. This simulation study shows that the mean square error, mean bias and coverage probability of credible intervals under Gamma, Jeffreys' rule and Box & Tiao priors presented equal results, whereas Jeffreys and Reference priors showed the best results. The MDIP prior had a worse performance in all analyzed situations showing not to be indicated for Bayesian analysis of the PE distribution. A real data set is analyzed for illustrative purpose of the Bayesian approaches.

***Key words*:** Bayesian; Poisson-Exponential; Jeffreys; MDIP; Objective; Prior.

Resumen

En este artículo, se examina un conjunto de importantes priori objetivas para la estimación bayesiana de los parámetros de la distribución Poisson-Exponencial (PE). Derivamos la priori Jeffreys multivariada y la *Maximal Data Information Prior*. También se analizan la priori de Referencia y otras

[a]Professor. E-mail: femoala@fct.unesp.br
[b]Professor. E-mail: gustavo.perez@unesp.br

prioris propuestas en la literatura. Mostramos que las distribuciones posterioris resultantes de estos enfoques son adecuadas, aunque las respectivas prioris son impropias. Las simulaciones de Monte Carlo se utilizan para comparar las eficiencias, para evaluar la sensibilidad de la elección de las prioris, principalmente para tamaños de muestra pequeños. Este estudio de simulación muestra que los errores cuadráticos medios, el sesgo medio y la probabilidad de cobertura de los intervalos creíbles bajo la Gamma, regla de Jeffreys y Box & Tiao mostraron resultados iguales, mientras que los prioris de Jeffreys y Reference mostraron los mejores resultados. El priori MDIP tuvo un peor desempeño en todas las situaciones analizadas mostrando no estar indicado para el análisis bayesiano de la distribución PE. Se analiza un conjunto de datos reales con fines ilustrativos de los enfoques bayesianos.

***Palabras clave*:** Bayesiano; Jeffreys; MDIP; Objetiva; Poisson-Exponencial; Priori.

1. Introduction

The Poisson-Exponential (PE) is a lifetime distribution with increasing failure rate introduced by Cancho et al. (2011). This model is derived in a complementary risks scenario where the lifetime associated with a particular risk is not observable, rather we observe only the maximum lifetime value among all risks. This distribution can be used to engineering problems where after increasing the failure rate the function may become stabilized.

As is well known, the choice of a prior distribution is the fundamental part of any Bayesian analysis. Objective priors, generally also called noninformative priors, refer to the case where relatively little information is available a priori, that is, information about model parameters is not considered substantial compared to information from a data set or in situations where a researcher is not able to express his/her prior opinion into a prior distribution.

Because there is not a precise definition about the concept of noninformative prior, there are in the Bayesian literature several forms of formulating noninformative priors, for instance, Jeffreys (1967), MDIP (Zellner, 1977, 1984), Tibshirani (1987), reference (Bernardo, 1979) and many others. Therefore, a study to derive the priors for a distribution fitted for the experimental data and check if these proposed priors lead to the same posterior inference results is of great practical interest. Furthermore, it is desirable to compare the different priors to check if any of them are preferable, especially for small samples.

Rodrigues et al. (2018) studied different non-Bayesian methods of estimation for the parameters of PE distribution and Louzada-Neto et al. (2011) provide a Bayesian inference by using a Jeffreys's rule prior representing weak information for the parameters of PE distribution. Tomazella et al. (2013) also present a Bayesian analysis for the parameters of PE distribution but using the Reference prior distribution proposed by Bernardo (1979). Singh et al. (2014) proposed the Bayesian approach under symmetric and asymmetric loss functions comparing them with the maximum likelihood procedure to estimate the parameters of Poisson-Exponential distribution for complete sample. Belaghi et al. (2019) obtain the estimates under the maximum likelihood approach, and Bayesian estimates using different loss functions for estimation and prediction problems when the lifetime data following the Poisson-Exponential distribution are observed under type-II censoring.

In this paper a more complete set of important objective priors representing a situation of absense or weak information of the parameters of the Poisson-Exponential distribution are examined. We derived the multivariate Jeffreys prior proposed by Jeffreys (1967) and the Maximal Data Information Prior (MDIP) proposed by Zellner (1977). Others types of objective priors proposed in the literature are also analyzed. We are also interested in selecting an objective prior that best represents a state of little knowledge a priori about the parameters.

In Bayesian analysis with objective priors, it should be justified that the posterior densities are proper. In this paper, all the posterior densities of the parameters of PE distribution will result proper posterior distributions although the priors are improper. In addition, a simulation study is performed using different sample sizes and we examine the bias, mean square error and frequentist coverage probabilities in order to compare the performance of the proposed priors.

One purpose of Bayesian inference is to obtain the marginal posterior densities because they provide complete information about parameters of interest such as Bayes estimator, mode and credible intervals. For this, we need to integrate the joint posterior density with respect to each parameter. However, since the marginal posterior densities cannot be obtained in a closed form through the integration, the Markov Chain Monte Carlo (MCMC) techniques, in special Metropolis.Hastings algorithm, to generate samples of values of θ and λ from the joint posterior distributions is also carried out.

The outline of the remaining sections is organized as follows. In Section 2, the Poisson-Exponential, its properties and the expected Fisher information matrix derived by Cancho et al. (2011) was reviewed; in Sections 3 and 5, we derive the Jeffreys and MDIP priors, respectively. Section 4 reviews the Reference prior developed by Tomazella et al. (2013) and Section 6 presents other proposed priors. Section 7 illustrates and discusses the results from the simulation performance and in Section 8 we introduce an applied example provided by Lawless (2003) to illustrate the Bayesian approach proposed. Finally in section 9, we have presented the conclusions.

2. The Poisson-Exponential Distribution

Let be T representing the lifetime of a component under the Poisson-Exponential distribution, denoted by PE, then the density is given by

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

for all $t > 0$ and depending on the shape and scale parameters $\theta > 0$ and $\lambda > 0$, respectively.

The survival and hazard functions associated to (1) are given, respectively, by,

$$S(t) = \frac{1 - \exp\{-\theta e^{-\lambda t}\}}{1 - e^{-\theta}}, \quad t > 0$$

and

$$h(t) = \frac{\theta\lambda \exp\{-\lambda t - \theta e^{-\lambda t}\}}{1 - \exp\{-\theta e^{-\lambda t}\}}, \quad t > 0.$$

The pth quantile of the PE is given by

$$t_p = \frac{1}{\lambda} \log\, \theta - \log\Big[- \log\Big(p - e^{-\theta}(p-1) \Big) \Big], \quad 0 < p < 1.$$

According to Cancho et al. (2011), the raw moments of T are given by,

$$\mu_k = E(T^k) = \frac{\theta\Gamma(k+1)}{\lambda^k(1-e^{-\theta})} F_{k+1,k+1}([1,\ldots,1],[2,\ldots,2],-\theta), \tag{2}$$

where the generalized hypergeometric function, denoted by $F_{pq}(\boldsymbol{a},\boldsymbol{b},\theta)$, is defined as

$$F_{pq}(\boldsymbol{a},\boldsymbol{b},\theta) = \sum_{j=0}^{\infty} \frac{\theta^j \prod_{i=1}^{p} \Gamma(\mathrm{a}_i + j)\big(\Gamma(\mathrm{a}_i)^{-1}\big)}{\Gamma(j+1)\prod_{i=1}^{p} \Gamma(b_i + j)\big(\Gamma(b_i)^{-1}\big)}, \tag{3}$$

with $\boldsymbol{a} = (a_1,\, a_2, \ldots, a_p)$ and $\boldsymbol{b} = (b_1,\, b_2, \ldots, b_q)$.

From (2), we have the mean and variance given, respectively, by

$$E(T) = \frac{\theta}{\lambda(1-e^{-\theta})} F_{22}([1,1],[2,2],-\theta) \tag{4}$$

and

$$var(T) = \frac{\theta}{\lambda^2(1-e^{-\theta})}\Big[F_{33}([1,1,1],[2,2,2],-\theta) - \frac{\theta}{1-e^{-\theta}} F_{22}([1,1].[2,2],-\theta)\Big].$$

Suppose we have independent identically distributed lifetimes $t_1, t_2, \ldots, t_n$ from PE. The likelihood function for the parameters θ and λ, based on the random sample $\boldsymbol{t} = (t_1,\, t_2, \ldots, t_n)$, is given by

$$L(\theta,\lambda|\,\boldsymbol{t}) \propto \Big(\frac{\theta\lambda}{1-e^{-\theta}}\Big)^n \exp\left\{ -\lambda \sum_{i=1}^{n} t_i - \theta \sum_{i=1}^{n} e^{-\lambda t_i} \right\}.$$

Cancho et al. (2011) provide the conditions which are needed in order to obtain the existence and uniqueness of the MLE when the other parameter is known. They also provide the Fisher information matrix given by

$$I(\theta,\lambda) = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix}, \tag{5}$$

with

$$I_{11} = -E\left(\frac{\partial^2}{\partial\theta^2} \ln L\right) = \frac{n}{\theta^2} - \frac{ne^{\theta}}{(e^{\theta}-1)^2},$$

$$I_{12} = -E(\frac{\partial^2}{\partial\theta\partial\lambda} \ln L) = -\frac{n\theta}{4\lambda(1-e^{-\theta})} F_{22}([2,2],[3,3],-\theta)$$

and

$$I_{22} = -E(\frac{\partial^2}{\partial\lambda^2} \ln L) = \frac{n}{\lambda^2}\Big(1 + \frac{\theta^2}{4(1-e^{-\theta})} F_{33}([2,2,2],[3,3,3],-\theta)\Big). \qquad (6)$$

3. Jeffreys Prior

A well known weak prior to represent a situation with little information available a priori about the parameters was proposed by Jeffreys (1967). Since then Jeffreys prior has played an important role in Bayesian inference. His prior is derived from Fisher Information matrix $I(\theta,\, \lambda)$ as

$$\pi(\theta,\lambda) \propto \sqrt{det\, I(\theta,\, \lambda)}\,. \qquad (7)$$

Box & Tiao (1973) give an explaining of the derivation of the noninformative Jeffreys priors in terms of "data translated" likelihood.

Jeffreys prior is widely used due to its invariance property under one-to-one transformations of parameters.

Theorem 1. *Jeffreys prior for $(\theta,\, \lambda)$ parameters of PE is given by:*

$$\pi(\theta,\lambda) \propto \frac{1}{\lambda}\sqrt{\Big(\frac{1}{\theta^2} - \frac{e^{\theta}}{(e^{\theta}-1)^2}\Big)\Big(1 + \frac{\theta^2}{4(1-e^{-\theta})} F_{33}([2,2,2],[3,3,3],-\theta)\Big) - \Big(\frac{\theta}{4(1-e^{-\theta})} F_{22}([2,2],[3,3],-\theta)\Big)^2} \qquad (8)$$

***Proof*.** Immediate from (5) and (7). □

As this prior is an improper prior then it should be justified that the posterior density is proper.

Corollary 1. *The joint posterior density $p(\theta,\lambda\,|\,\boldsymbol{t})$ for parameters $(\theta,\, \lambda)$ under Jeffreys prior (8) is proper.*

***Proof*.** Since $e^{-\lambda t} \leq 1$ for all $t_i > 0$, $i = 1, \dots, n$, and $\lambda > 0$, it follows that

$$\int_0^{\infty}\int_0^{\infty} p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \int_0^{\infty}\int_0^{\infty}\Big(\frac{\theta\lambda}{1-e^{-\theta}}\Big)^n \exp\Big\{-\lambda\sum_{i=1}^{n} t_i\Big\} e^{-n\theta}\pi(\theta,\lambda)d\theta d\lambda =$$

$$\int_0^{\infty}\Big[\int_0^{\infty}\lambda^{\,n-1} e^{-\lambda\sum_{i=1}^{n} t_i} d\lambda\Big]\Big(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\Big)^n \varphi(\theta) d\theta$$

where

$$\varphi(\theta)=\sqrt{\Big(\frac{1}{\theta^2}-\frac{e^\theta}{(e^\theta-1)^2}\Big)\Big(1+\frac{\theta^2}{4(1-e^{-\theta})}F_{33}([2,2,2],[3,3,3],-\theta)\Big)-\Big(\frac{\theta}{4(1-e^{-\theta})}F_{22}([2,2],[3,3],-\theta)\Big)^2}. \tag{9}$$

Taking the integration with respect to λ, we have $\int_0^\infty \lambda^{n-1}e^{-\lambda\sum_{i=1}^n t_i}d\lambda = \frac{\Gamma(n)}{\left(\sum_{i=1}^n t_i\right)^n}$ and hence,

$$\int_0^\infty\int_0^\infty p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \frac{\Gamma(n)}{\left(\sum_{i=1}^n t_i\right)^n}\int_0^\infty\Big(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\Big)^n\varphi(\theta)d\theta.$$

Now, since $\frac{\theta e^{-\theta}}{1-e^{-\theta}}\leq 1$, then $\left(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\right)^n\leq\frac{\theta e^{-\theta}}{1-e^{-\theta}}$ and

$$\int_0^\infty\Big(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\Big)^n\varphi(\theta)d\theta \leq \int_0^\infty\frac{\theta e^{-\theta}}{1-e^{-\theta}}\varphi(\theta)d\theta = 0.311677. \tag{10}$$

The last integral in (10) is evaluated numerically by using Mathematica. Therefore

$$\int_0^\infty\int_0^\infty p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \frac{\Gamma(n)}{\left(\sum_{i=1}^n t_i\right)^n}<\infty.$$

□

4. Reference Prior

Another well-known class of noninformative priors is the reference prior proposed by Bernardo (1979) and further improved by Berger & Bernardo (1992).

The idea is to derive a prior $\pi(\phi)$ that maximizes the expected posterior information about the parameters provided by independent replications of an experiment relative to the information in the prior. A natural measure of the expected information about ϕ provided by data $\boldsymbol{x}$ is given by

$$I(\phi)=E_{\boldsymbol{x}}[K(p(\phi\mid\boldsymbol{x}),\pi(\phi))]$$

where

$$K(p(\phi\mid\boldsymbol{x}),\pi(\phi))=\int_\Phi p(\phi\mid\boldsymbol{x})\log\frac{p(\phi\mid\boldsymbol{x})}{\pi(\phi)}d\phi \tag{11}$$

is the Kullback-Leibler distance. Thus, the reference prior is defined as the prior $\pi(\phi)$ that maximizes the expected Kullback-Leibler distance between the posterior density $p(\phi|\boldsymbol{x})$ and the prior density $\pi(\phi)$, taken over the experimental data.

The prior density $\pi(\phi)$ which maximizes the functional (11) is found through calculus of variation and, the solution is not explicit. However, when the posterior $p(\phi \mid \boldsymbol{x})$ is asymptotically normal, this approach leads to Jeffreys prior for a single parameter situation. If on the other hand, we are interested in one of the parameters, being the remaining parameters nuisances, the situation is quite different, and the appropriated reference prior is not a multivariate Jeffreys prior. Bernardo argues that when nuisance parameters are present, the reference prior should depend on which parameters are considered to be of primary interest. The reference prior in this case is derived as follows. We will present here the two-parameters case in details. For the multiparameter case, Berger & Bernardo (1992).

Let $\theta = (\theta_1, \theta_2)$ be the whole parameter, θ_1 being the parameter of interest and θ_2 the nuisance parameter. The algorithm is given as follows:

Step 1: Determine $\pi_2(\theta_2| \, \theta_1)$, the conditional reference prior for θ_2 assuming that θ_1 is given,

$$\pi_2(\theta_2| \, \theta_1) = \sqrt{\mathrm{I}_{22}(\theta_1, \theta_2)},$$

where $\mathrm{I}_{22}(\theta_1, \theta_2)$ is the (2,2)-entry of the Fisher Information Matrix.

Step 2: Normalize $\pi_2(\theta_2| \, \theta_1)$.

Case $\pi_2(\theta_2| \, \theta_1)$ is improper, choose a sequence of subsets $\Omega_1 \subseteq \Omega_2 \subseteq \cdots \to \Omega$ on which $\pi_2(\theta_2| \, \theta_1)$ is proper. Define

$$c_{\mathrm{m}}(\theta_1) = \frac{1}{\int_{\Omega_m} \pi_2(\theta_2| \, \theta_1) d\theta_2}$$

and

$$p_m(\theta_2| \, \theta_1) = c_m(\theta_1)\pi_2(\theta_2| \, \theta_1) 1_{\Omega_m}(\theta_2).$$

Step 3: Find the marginal reference prior for θ_1, i.e., the reference prior for the experiment formed by marginalizing out with respect to $p_m(\theta_2| \, \theta_1)$. We obtain

$$\pi_m(\theta_1) \propto exp\Big\{ \frac{1}{2} \int_{\Omega_m} p_m(\theta_2| \, \theta_1) log \left\| \frac{det \; I(\theta_1, \theta_2)}{I_{22}(\theta_1, \theta_2)} \right\| d\theta_2 \Big\}.$$

Step 4: Compute the reference prior for (θ_1, θ_2) when θ_2 is a nuisance parameter:

$$\pi(\theta_1, \theta_2) = \lim_{m \to \infty} \left(\frac{c_m(\theta_1)\pi_m(\theta_1)}{c_m(\theta_1^*) \; \pi_m(\theta_1^*)} \right) \pi(\theta_2 \mid \theta_1),$$

where θ_1^* is any fixed point with positive density for all π_m.

Tomazella et al. (2013) derive reference prior and prove that the corresponding posteriori density is proper.

Theorem 2. *Reference prior for* (θ, λ) *parameters of* PE *is given by:*

$$\pi(\theta, \lambda) \propto \frac{1}{\lambda} \Big(1 + \frac{\theta^2}{4(1 - e^{-\theta})} F_{33}([2, 2, 2], [3, 3, 3], -\theta) \Big) \varphi(\theta), \tag{12}$$

where $\varphi(\theta)$ is given by (9).

Proof. See Tomazella et al. (2013). □

Note that Reference prior is the product of Jeffreys prior and the term $\Big(1 + \frac{\theta^2}{4(1-e^{-\theta})} F_{33}([2,2,2],[3,3,3],-\theta)\Big)$.

Corollary 2. *The joint posterior density $p(\theta,\lambda \,|\, \boldsymbol{t})$ for parameters $(\theta,\ \lambda)$ under Reference prior given in (12) is proper.*

Proof. See Tomazella et al. (2013). □

5. Maximal Data Information Prior (MDIP)

It is of interesting that the data gives more information about the parameter than the information on the prior density, otherwise, there would not be justification for the realization of the experiment. Thus, we wish a prior density $\pi(\phi)$ that provides the gain in the information supplied by the data the largest as possible relative to the prior information of the parameter, that is, maximizes the information on the data. With this idea, Zellner (1977), Zellner (1984) and Zellner & Min (1992) derived a prior which maximize the average information in the data density $f(x \,|\, \phi)$ relative to that one in the prior. Let

$$H(\phi) = \int_{\mathrm{a}}^{\mathrm{b}} f(x \,|\, \phi) \ln f(x \,|\, \phi) dx$$

be the negative entropy, the measure of the information in $f(x \,|\, \phi)$. Thus, the following functional criterion is employed in the MDIP approach:

$$G[\pi(\phi)] = \int_{\mathrm{a}}^{\mathrm{b}} H(\phi)\pi(\phi)d\phi - \int_{\mathrm{a}}^{\mathrm{b}} \pi(\phi) \ln \pi(\phi) d\phi,$$

which is the prior average information in the data density minus the information in the prior density. $G[\pi(\phi)]$ is maximized by selection of $\pi(\phi)$ subject to $\int_{\mathrm{a}}^{\mathrm{b}} \pi(\phi)d\phi = 1$.

The solution is then a proper prior density given by

$$\pi(\phi) = k \exp\Big\{H(\phi)\Big\}, \qquad a \leq \phi \leq b, \tag{13}$$

where $k^{-1} = \int_{\mathrm{a}}^{\mathrm{b}} \exp\Big\{H(\phi)\Big\} d\phi$ is the normalizing constant.

Therefore, the MDIP is a prior that leads to an emphasis on the information in the data density or likelihood function, that is, its information is weak in comparison with data information. More details of the construction of this prior can be found in Zellner (1984).

Zellner (1990) and Zellner (1996) show several interesting properties of MDIP and additional conditions that can also be imposed to the approach refleting given initial information. However, the MDIP has restrictive invariance properties.

We suppose that we do not have much prior information available about θ and λ. Therefore, under this condition, the prior distribution MDIP for the parameters (θ, λ) of PE density (1) is also appropriated for our inference problems.

Theorem 3. *MDIP prior for (θ, λ) parameters of PE is given by*

$$\pi_Z(\theta,\lambda) \propto \frac{\lambda\theta}{1-e^{-\theta}} \exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22}([1,1],[2,2],-\theta) - e^{-\theta}\Big)\Big\}. \qquad (14)$$

Proof. Firstly, we have to evaluate the negative entropy of the distribution $f(t \mid \theta, \lambda)$,

$$H(\theta,\lambda) = \int_0^\infty \ln\Big(\frac{\theta\lambda\exp\{-\lambda t - \theta e^{-\lambda t}\}}{1-e^{-\theta}}\Big) f(t)dt,$$

and after some algebras,

$$H(\theta,\lambda) = \ln\Big(\frac{\theta\lambda}{1-e^{-\theta}}\Big) - \lambda E(T) - \theta E(e^{-\lambda T}).$$

From mean of the PE in (4) we have

$$H(\theta,\lambda) = \ln\Big(\frac{\theta\lambda}{1-e^{-\theta}}\Big) - \frac{\theta}{(1-e^{-\theta})} F_{22}([1,1],[2,2],-\theta) - \theta E(e^{-\lambda T}),$$

with $F_{22}([1,1],[2,2],-\theta)$ defined in (3).

Now, to evaluate the expectance $E(e^{-\lambda T}) = \int_0^\infty e^{-\lambda t} \frac{\theta\lambda\exp\{-\lambda t - \theta e^{-\lambda t}\}}{1-e^{-\theta}} dt$ consider the transformation $u = e^{-\lambda t}$ then $E(e^{-\lambda T}) = \frac{1}{\theta} + \frac{1}{1-e^{-\theta}}$.

Hence, the entropy of PE is obtained as

$$H(\theta,\lambda) = \ln\Big(\frac{\theta\lambda}{1-e^{-\theta}}\Big) - \frac{\theta}{(1-e^{-\theta})} F_{22}([1,1],[2,2],-\theta) + \frac{\theta e^{-\theta}}{1-e^{-\theta}} - 1. \qquad (15)$$

From (13) and (15), the MDIP prior for parameters of PE is obtained. □

Corollary 3. *The joint posterior density for parameters (θ, λ) under MDIP prior (14) is proper.*

Proof. The joint posterior density for parameters (θ, λ) is given by

$$p(\theta,\lambda\,|\,\boldsymbol{t}) \propto \Big(\frac{\theta\lambda}{1-e^{-\theta}}\Big)^{n+1} \exp\Big\{-\lambda\sum_{i=1}^n t_i - \theta\sum_{i=1}^n e^{-\lambda t_i}\Big\} \exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22} - e^{-\theta}\Big)\Big\},$$

where $F_{22} = F_{22}([1,1],[2,2],-\theta)$.

Since $e^{-\lambda t} \leq 1$ for all $t_i > 0$, $i = 1, \ldots, n$, and $\lambda > 0$, it follows that

$$\int_0^\infty \int_0^\infty p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \int_0^\infty \Big[\int_0^\infty \lambda^{n+1} e^{-\lambda \sum_{i=1}^n t_i} d\lambda\Big]\Big(\frac{\theta}{1-e^{-\theta}}\Big)^{n+1} e^{-n\theta} \exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22}-e^{-\theta}\Big)\Big\}d\theta.$$

From $\int_0^\infty \lambda^{n+1} e^{-\lambda \sum_{i=1}^n t_i} d\lambda = \frac{\Gamma(n+2)}{\left(\sum_{i=1}^n t_i\right)^{n+2}}$ we have

$$\int_0^\infty \int_0^\infty p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \frac{\Gamma(n+2)}{\left(\sum_{i=1}^n t_i\right)^{n+2}} \int_0^\infty \Big(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\Big)^{n+1} \exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22}-e^{-\theta}\Big)\Big\}d\theta.$$

Due to $\frac{\theta e^{-\theta}}{1-e^{-\theta}} \leq 1$ then

$$\int_0^\infty \int_0^\infty p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \frac{\Gamma(n+2)}{\left(\sum_{i=1}^n t_i\right)^{n+2}} \int_0^\infty \Big(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\Big) \exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22}-e^{-\theta}\Big)\Big\}d\theta.$$

From numerical integration we have

$$\int_0^\infty \Big(\frac{\theta e^{-\theta}}{1-e^{-\theta}}\Big) exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22}-e^{-\theta}\Big)\Big\}d\theta = 0.835182,$$

and therefore

$$\int_0^\infty \int_0^\infty p(\theta,\lambda\,|\,\boldsymbol{t})d\theta d\lambda \leq \frac{\Gamma(n+2)}{\left(\sum_{i=1}^n t_i\right)^{n+2}} < \infty.$$

□

6. Other proposed priors

A possible simplification of Jeffreys prior is to consider a noninformative prior from $\pi(\theta,\, \lambda)= \pi(\lambda|\, \theta)\pi(\theta)$. Using the Jeffreys' rule, we have

$$\pi(\theta,\lambda) \propto \sqrt{E\Big(\frac{\partial^2}{\partial\lambda^2}\ln\, L\Big)}\,\pi(\theta), \tag{16}$$

where $E\Big(\frac{\partial^2}{\partial\lambda^2}\ln\, L\Big)$ is given by (6) and $\pi(\theta)$ is a noniformative proper prior.

Louzada-Neto et al. (2011) consider $\pi(\theta)$ given by the Gamma distribution with parameters "a" and "b".

Theorem 4. *The Jeffreys' rule prior is given by*

$$\pi(\theta,\lambda) \propto \frac{1}{\lambda}\theta^{a-1}e^{-b\theta}. \tag{17}$$

Proof. In fact, from element I_{22} given in (6) and prior (16) with $\pi(\theta)$ given by a gamma distribution with hyperparameters "a" and "b" the considered prior is obtained. □

Let us denote the prior (17) as Jeffreys' rule.

Corollary 4. *The joint posterior density $p(\theta,\lambda\,|\,\boldsymbol{t})$ for parameters (θ,λ) under Jeffreys' rule prior given in (17) is proper.*

Proof. See Louzada-Neto et al. (2011). □

Theorem 5. *Another noninformative prior distribution is assumed considering independence between θ and λ, given by*

$$\pi(\theta,\lambda) \propto \frac{1}{\theta\lambda}. \tag{18}$$

Let us denote the prior (18) as Box & Tiao prior.

Corollary 5. *The joint posterior density $p(\theta,\lambda \mid \boldsymbol{t})$ for parameters $(\theta,\ \lambda)$ under Box & Tiao prior given in (18) is proper.*

Proof. The proof is similar to that considered by Louzada-Neto et al. (2011). □

Other prior specifications also could be used, as independent informative Gamma distributions, that is,

$$\pi_\theta(\theta) \sim Gamma(a_\alpha,\, b_\alpha) \tag{19}$$

and

$$\pi_\lambda(\lambda) \sim Gamma(a_\lambda,\, b_\lambda), \tag{20}$$

where a_θ, b_θ, a_λ and b_λ are known hyperparameters and $Gamma(\mathrm{a},\, b)$ denotes a gamma distribution with mean a/b and variance a/b^2.

Since shape θ and scale λ parameters of PE assumes values greater than zero then we can pre-establish as usual a Gamma prior distribution with shape and scale hyperparameters $a > 0$ and $b > 0$, respectively. Values of $a \to 0$ and $b \to 0$ suggest a vague (absence of information) prior distribution. Thus, the hyperparameters in (19) and (20) could be choosen such as 0.01 or 0.001 to provide no prior information.

Table 1 displays the joint prior distributions obtained for the different approaches shown in this paper.

7. Simulated Data

This section presents a simulation in order to compare and choose a prior distribution proposed in this paper which better represent a situation of weak information about the parameters.

Table 1: Joint prior distributions

Prior	$\boldsymbol{\pi(\theta,\lambda)}$
Jeffreys	$\frac{1}{\lambda}\varphi(\theta)$
Reference	$\frac{1}{\lambda}\varphi(\theta)\Big(1+\frac{\theta^2}{4(1-e^{-\theta})}F_{33}([2,2,2],[3,3,3],-\theta)\Big)$
MDIP	$\frac{\lambda\theta}{1-e^{-\theta}}exp\Big\{-\frac{\theta}{(1-e^{-\theta})}\Big(F_{22}([1,1],[2,2],-\theta)-e^{-\theta}\Big)\Big\}$
Gamma	$\lambda^{a-1}e^{-b\lambda}\theta^{c-1}e^{-d\theta}$
Jeffreys' rule	$\frac{1}{\lambda}\theta^{a-1}e^{-b\theta}$
Box & Tiao	$\frac{1}{\lambda\theta}$

The simulated data are generate from PE distribution with parameter values $\theta = 5$ and $\lambda = 2$ for different sample sizes, as $n = 10$ (small), 30, 50 (moderate), 100 and 200 (large).

We report the average estimators, the mean square error (MSE) and the mean bias, over 1000 generated samples for each sample size. The efficiency for the estimators was compared according to these measures. The results are reported in Tables 2 and 3.

We also need to appeal to numerical procedures to extract characteristics of marginal posterior distributions such as Bayesian estimators and credible intervals. We can use the MCMC algorithm to obtain a sample of θ and λ from the joint posterior. The chain is run for 105 000 iterations with a burn-in period of 5000 and jumping 100.

Table 2: Simulation results for different sample sizes and parameters $\theta = 5$.

n	Estimate	Gamma	Jeffreys'rule	Jeffreys	Reference	MDIP	Box&Tiao
	Bias	4.7966	4.8092	3.5161	3.0793	20.4176	6.6193
10	MSE	82.9010	84.8641	27.8572	21.3021	738.9883	440.7152
	Estimator	7.0448	7.0711	7.5547	6.8291	24.9537	8.9153
	Bias	1.6384	1.6420	1.5211	1.4486	2.3645	1.6602
30	MSE	5.1780	5.1551	4.7750	4.2047	21.0929	5.3872
	Estimator	5.2461	5.2460	5.7086	5.4990	6.7998	5.2830
	Bias	1.1173	1.1218	1.0748	1.0493	1.2731	1.1279
50	MSE	2.1494	2.1604	2.1012	1.9480	3.1424	2.2008
	Estimator	5.1577	5.1594	5.3817	5.2644	5.7675	5.1773
	Bias	0.7402	0.7412	0.7428	0.7266	0.8212	0.7418
100	MSE	0.9019	0.9023	0.9198	0.8755	1.1420	0.9042
	Estimator	5.1775	5.1800	5.2746	5.2168	5.4502	5.1855
	Bias	0.4984	0.4996	0.4979	0.4944	0.5221	0.4896
200	MSE	0.4066	0.4091	0.4108	0.4012	0.4605	0.3866
	Estimator	5.0781	5.0772	5.1240	5.0976	5.2050	5.0734

Some of the points are quite clear from the results. As expected, when the sample size increases, it is observed that the performances of all estimators become better and closer, the MSE and biases decrease, although slowly.

TABLE 3: Simulation results for different sample sizes and parameters $\lambda = 2$.

n	Estimate	Gamma	Jeffreys'rule	Jeffreys	Reference	MDIP	Box & Tiao
	Bias	0.5627	0.5616	0.4813	0.4490	1.3930	0.5900
10	MSE	0.5161	0.5198	0.4129	0.3538	3.1809	0.6065
	Estimator	1.9667	1.9708	2.2300	2.1430	3.3163	2.0055
	Bias	0.2877	0.2876	0.2643	0.2589	0.3414	0.2893
30	MSE	0.1417	0.1421	0.1225	0.1168	0.2234	0.1438
	Estimator	1.9848	1.9855	2.0687	2.0383	2.2195	1.9899
	Bias	0.2120	0.2111	0.2020	0.1999	0.2251	0.2119
50	MSE	0.0729	0.0724	0.0679	0.0658	0.0874	0.0732
	Estimator	1.9987	1.9992	2.0369	2.0196	2.1164	2.0013
	Bias	0.1460	0.1459	0.1455	0.1437	0.1567	0.1460
100	MSE	0.0337	0.0339	0.0340	0.0330	0.0399	0.0338
	Estimator	2.0226	2.0232	2.0383	2.0298	2.0771	2.0241
	Bias	0.0973	0.0979	0.0973	0.0967	0.1007	0.0963
200	MSE	0.0150	0.0151	0.0150	0.0148	0.0161	0.0146
	Estimator	2.0049	2.0044	2.0120	2.0080	2.0307	2.0042

The results from Tables 2 and 3 show that the PE distribution is not indicated for samples with size $n < 30$ due to the poor estimates obtained whatever prior used. In addition, the Reference prior shown better performance than the others.

The results of the estimation of parameter λ given in Table 3 show practically insignificant differences even for sample sizes $n < 30$ and the values of Bias and MSE decrease very slowly when n increases.

Therefore, the conclusion of this analysis is that the shape parameter θ must be considered for the choice of the best prior to be used.

Other criterion for comparison of the prior densities consists on checking the frequentist coverage probabilities of the posterior intervals. Tables 4 and 5 illustrate the coverage probabilities.

TABLE 4: Frequentist coverage probability of the 95% intervals for $\theta = 5$.

n	Gammas	Jeffreys'rule	Jeffreys	Reference	MDIP	Box&Tiao
10	0.90	0.89	0.94	0.95	0.65	0.88
30	0.93	0.93	0.96	0.95	0.91	0.93
50	0.94	0.94	0.95	0.96	0.93	0.95
100	0.94	0.94	0.94	0.95	0.93	0.94
200	0.95	0.95	0.94	0.94	0.93	0.95

In terms of coverage probability, the simulation study indicates that the Reference prior performs better than the other priors, mainly for the parameters θ. Tables 4 and 5 illustrate that the coverage probabilities are often smaller than the nominal level for sample size $n < 30$, although the Reference prior performs slightly better than the others for this case. We can also see that the MDIP prior performs most poorly.

Table 5: Frequentist coverage probability of the 95% intervals for $\lambda = 2$.

n	Gammas	Jeffreys'rule	Jeffreys	Reference	MDIP	Box&Tiao
10	0.93	0.94	0.94	0.97	0.64	0.93
30	0.93	0.93	0.94	0.95	0.91	0.93
50	0.95	0.96	0.95	0.95	0.91	0.94
100	0.96	0.95	0.94	0.95	0.92	0.95
200	0.95	0.96	0.94	0.96	0.94	0.96

Some of the points are quite clear from the numerical results.

As expected, a moderate large ($n \geq 50$) sample size is needed to achieve the desirable accuracy, and in this case the choice of the priors become irrelevant.

The same result cannot be completely achieved when n is a small value ($n < 30$). In this case, we conclude that Jeffreys and Reference priors provide better estimation than any other priors presented in the study and, consequently, both can represent a situation of weak information a priori. Furthermore, a comparison between Jeffreys and Reference priors shows that Reference prior slightly dominates Jeffreys when n is very small. Note that MDIP prior does not provide good estimates among all the considered class of priors.

Therefore, based on this simulated study we recommend the Bayesian approach with Reference prior as the best inference to estimate the parameters and these results are of great interest in applications of the Poisson-Exponential distribution.

8. An Example With Literature Data

The data set below was obtained from Lawless (2003). The data given arose in tests on endurance of deep groove ball bearings. The data are the number of million revolutions before failure for each of the 23 ball bearings in the life test and they are 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.44, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

In Figure 1, we have the plots of the histogram with fitted density, and empirical with fitted cumulative functions modeled by PE. From Figure 1, we observe a good fit of the PE distribution for the data (Lawless data). Thus, based on the plots, we can assume that PE distribution is appropriated to analyze this dataset.

Table 6 presents the posterior mean, standard deviation and credible interval considering each prior distribution for the parameters θ and λ. From this Table we see that the Reference prior is more appropriate for both θ and λ although the differences with the other priors are not so significant as observed also in the results obtained by the simulation given in Tables 3, 4 and 5. The exception again is the MDIP prior, whose results are also far from.

The graphical representations of the marginal posterior densities for the parameters θ and λ are shown in Figure 2. Comparing the marginal posterior densities we can see the posteriors for both parameter are quite similar. The plots of posteriors $p(\theta \,|\, \boldsymbol{t})$ are so close such that one choice is almost impossible.

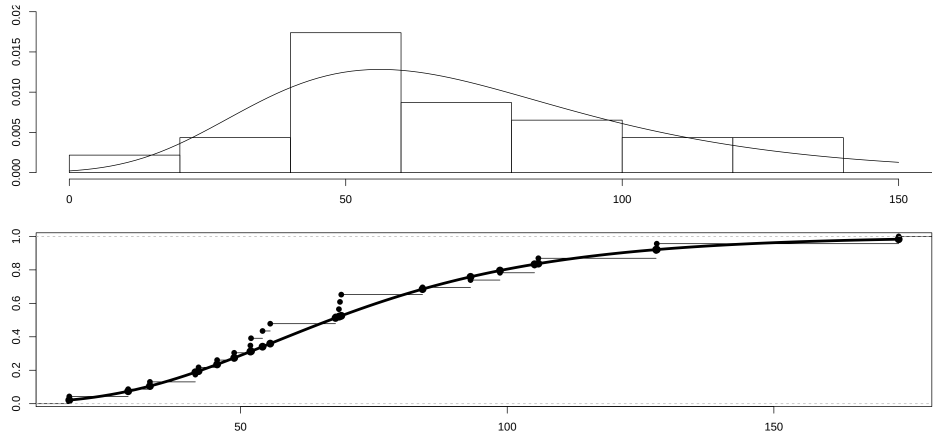

FIGURE 1: Plots of histogram and empirical cumulative function with respective fitted functions.

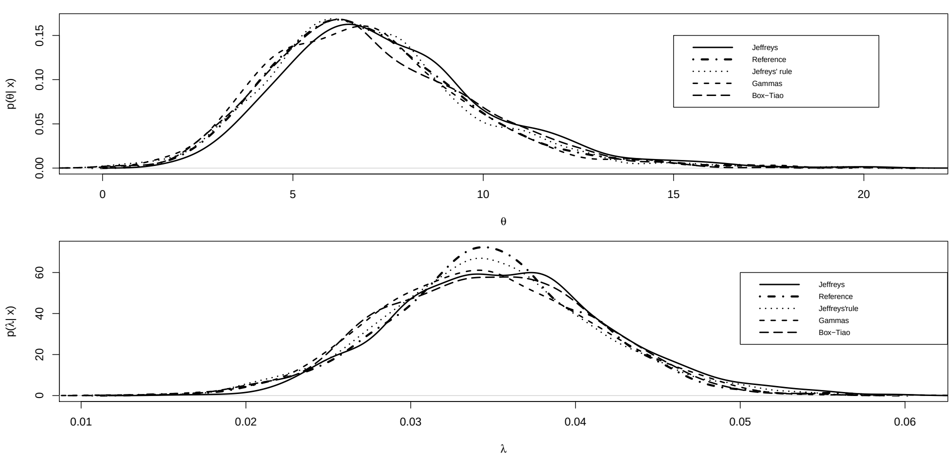

FIGURE 2: Marginal posterior densities of parameters θ and λ for the data (Lawless data).

TABLE 6: Estimators and 95% condence intervals of θ and λ for different prior.

	θ			λ		
Prior	Estimator	sd	CI	Estimator	sd	CI
Gamma	6.9254	2.6523	(2.6606, 13.3033)	0.0344	0.0064	(0.0226, 0.0474)
Jeffreys' rule	7.0384	2.6962	(2.7252, 12.9593)	0.0346	0.0063	(0.0221, 0.0481)
Jeffreys	7.5568	2.8554	(3.2070, 14.3470)	0.0357	0.0065	(0.0242, 0.0501)
Reference	7.0820	2.6152	(2.9928, 13.3740)	0.0348	0.0060	(0.0226, 0.0465)
MDIP	13.7054	13.2475	(3.9285, 50.5502)	0.0428	0.0120	(0.0269, 0.0746)
Box&Tiao	7.0629	2.6331	(2.8454, 12.9287)	0.0346	0.0063	(0.0222, 0.0463)

9. Conclusions

In this paper, the objective priors are derived for the parameters of PE distribution. A study to check if these priors lead to the same posterior inference for small and moderate sample sizes is of great practical interest. This way, a simulation study was performed and it indicated that the Reference prior performs better than the other priors for both parameters θ and λ.

In all the evaluation criteria of the considered priors, Gamma, Jeffreys' rule and Box&Tiao presented equal results whereas Jeffreys and Reference priors showed close and better results. The MDIP prior had a worse performance in all analyzed situations showing not to be indicated for Bayesian analysis of the PE distribution.

The results from Tables 2 and 3 show that the Poisson-Exponential distribution should not be indicated for dataset with size $n < 30$ due to the poor estimates obtained whatever prior used.

In addition to the results obtained from the simulation study of objective priors treated in this paper for the specific case of the Poisson-Exponential distribution, it is also worth highlighting the advantages and disadvantages of using these priors in the general case, as we will see below.

Jeffreys prior is quite universal and invariant in the sense of yielding properly tranformed priors under reparametrization, however, Jeffreys himself noticed difficulties with the method when the parameter is multi-dimensional.

Reference prior provides one of the most successful general methods to derive noninformative prior distributions. In practice, however, reference priors are typically difficult to use. Undesirable properties include lack of invariance to reparameterization and nonuniqueness of prior due to the choice of the parameter of interest.

The MDIP prior provides a fresh view and operational results for the problem of selecting "diffuse" prior. Side conditions reflecting initial information that may be available can readily be used to derive it. Therefore, MDIP prior can be quite useful for problems where moments are known. As for the invariance, the MDIP prior is invariant for a class of transformations much smaller than the Jeffreys prior. A serious problem with the MDIP prior is that it can lead to an improper posterior density more often than other objective priors for reliability distributions, for instance with Gamma and Generalized Exponential distributions.

[Received: May 2022 — Accepted: November 2022]

References

Belaghi, R., Asl, M., Alma, O., Singh, S. & 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.

Berger, J. O. & Bernardo, J. M. (1992), 'On the development of the reference prior method', *Bayesian Statistics* **4**(4), 35–60.

Bernardo, J. M. (1979), 'Reference posterior distributions for bayesian inference', *Journal Royal Statistical Society* **41**(2), 113–147.

Box, G. E. P. & Tiao, G. C. (1973), *Bayesian inference in statistical analysis*.

Cancho, V. G., Louzada-Neto, F. & Barriga, G. D. C. (2011), 'The poisson-exponential lifetime distribution', *Computational Statistics and Data Analysis* **55**, 677–686.

Jeffreys, S. H. (1967), *Theory of probability*, number 3, London, Oxford U. Press.

Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, Vol. 2 edition, Wiley, New York.

Louzada-Neto, F., Cancho, V. G. & Barriga, G. D. C. (2011), 'The poisson–exponential distribution: a bayesian approach', *Journal of Applied Statistics* **38**(6), 1239–1248.

Rodrigues, G., Louzada, F. & Ramos, P. (2018), 'Poisson–exponential distribution: different methods of estimation', *Journal of Applied Statistics* **45**(1), 128–144.

Singh, S., Singh, U. & Kumar, M. (2014), 'Estimation for the parameter of poisson-exponential distribution under bayesian paradigm', *Journal of Data Science* **12**, 157–173.

Tibshirani, R. (1987), 'Noninformative priors for one parameters of many', *Biometrika* **76**, 604–608.

Tomazella, V. L. D., Cancho, V. G. & Louzada-Neto, F. (2013), 'Bayesian reference analysis for the poisson-exponential lifetime distribution', *Chilean Journal of Statistics* **4**(1), 99–113.

Zellner, A. (1977), Maximal data information prior distributions, *in* A. Aykac & C. Brumat, eds, 'New methods in the applications of bayesian methods', North-Holland, Amsterdam.

Zellner, A. (1984), 'Maximal data information prior distributions', *Basic Issues in Econometrics* .

Zellner, A. (1990), Bayesian methods and entropy in economics and econometrics, *in* J. W. T. Grandy & L. H. Schick, eds, 'Maximum Entropy and Bayesian Methods', Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 17–31.

Zellner, A. (1996), 'Models, prior information and bayesian analysis', *Journal of Econometrics* **75**(1), 51–68.

Zellner, A. & Min, C.-K. (1992), *Bayesian analysis, model selection and prediction*, University of Chicago, Graduate School of Business, Department of Economics.