Chilean Journal of Statistics
Vol. 4, No. 1, April 2013, 99–113

Bayesian Statistics
Research Paper

Bayesian reference analysis for the Poisson-exponential lifetime distribution

Vera L.D. Tomazella[1], Vicente G. Cancho[2], and Francisco Louzada[2,*]

[1]Department of Statistics, Universidade Federal de São Carlos, São Carlos-SP, Brazil
[2]Department of App. Math. and Stat., Universidade de São Paulo, São Carlos-SP, Brazil

(Received: 01 March 2012 · Accepted in final form: 21 January 2013)

Abstract

In this paper we propose an objective Bayesian estimation methodology for the parameter of the Poisson-exponential lifetime distribution, which is a new two-parameter lifetime distribution with increasing failure rate. This distribution has an important position on the latent complementary risk problem scenario. We also perform a simulation study in order to analyze the frequentist coverage probabilities of credible interval derived from non-subjective posteriors. The developed procedures are illustrated on a real data set.

Keywords: Bayesian inference · Complementary risks · Exponential distribution · Poisson distribution · Survival analysis.

Mathematics Subject Classification: Primary 62N01 · Secondary 62N02.

1. Introduction

Survival analysis in presence of complementary risks (CR) is a statistical modeling concept which aims to account for situations where the risks are latent in the sense that there is no information about which factor was responsible for the component failure, which can be often observed in field data, such as public health, actuarial science, biomedical studies, demography and industrial reliability; see Basu and Klein (1982); Louzada-Neto (1999); Davison and Louzada-Neto (2000); Cancho et al. (2011). On many occasions this information is not available or it is impossible that the true cause of failure be specified by an expert. In reliability studies, the components can be totally destroyed in the experiment. Further, the true cause of failure can be masked from our view. In modular systems, the need to keep a system running means that a module that contains many components can be replaced without the identification of the exact failing component. Goetghebeur and Ryan (1995) addressed the problem of assessing covariate effects based on a semi-parametric proportional hazards structure for each failure type when the failure type is unknown for some individuals. Reiser et al. (1995) considered statistical procedures for analyzing masked data, but their procedure can not be applied when all observations have an unknown cause of failure. Lu and Tsiatis (2001) present a multiple imputation method for estimating regression coefficients for risk modeling with missing cause of failure. A comparison of two partial likelihood approaches for this situation is presented in Lu and Tsiatis (2005).

*Corresponding author. Francisco Louzada. ICMC-USP. Av. Trabalhador Socarlense 400, CEP 13.566-590, São Carlos, SP, Brazil. Email: louzada@icmc.usp.br

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

Simplistically, in reliability, we observe only the maximum component lifetime of a parallel system. That is, the observable quantities for each component are the maximum lifetime value to failure among all risks, and the cause of failure. Full statistical procedures and extensive literature are available to deal with these problems and interested readers can refer to Lawless (2003), Crowder et al. (1991) and Cox and Oakes (1984). In the classical complementary risk scenarios the lifetime associated with a particular risk is not observable; rather we observe only the maximum lifetime value among all risks.

The exponential distribution (ED) provides a simple, elegant and close form solution to many problems in lifetime testing and reliability studies. However, the ED does not provide a reasonable parametric fit for some practical applications where the underlying hazard rates are nonconstant, presenting monotone shapes. In recent years, in order to overcome such a problem, new classes of models were introduced based on modifications of the ED. Gupta and Kundu (1999) proposed a generalized ED. This extended family can accommodate data with increasing and decreasing failure rate functions. Kus (2007) proposed another modification of the ED with a decreasing failure rate function. While Barreto-Souza and Cribari-Neto (2009) generalizes the distribution proposed by Kus (2007) by including a power parameter in his distribution. In this paper, we propose a new distribution family based on the ED with increasing failure rate function. Its genesis is stated on a complementary risk problem base (Basu and Klein, 1982) in presence of latent risks, in the sense that there is no information about which factor was responsible for the component failure and only the maximum lifetime value among all risks is observed. The proposed distribution can be seem as an counterpart of the distribution proposed by Kus (2007), since he only observe the minimum lifetime value among all risks, while on our formulation is only the the maximum lifetime value among all risks is observed.

The model presented in this article will be considered under an objective Bayesian perspective. From this perspective, the outcome of any inference problem is the posterior distribution of the quantity of interest, which combines the information provided by the data with available prior information. It has been often recognized that there is a pragmatically important need for a form of prior to posterior analysis which captures, in a well-defined sense, the notion that the prior should have a minimal effect, relative to the data, on the posterior inference. For instance, we can cite the Jeffreys (1946) and the reference Bernardo (1979) priors, which have such purpose. The Jeffreys prior is invariant in the sense of yielding properly transformed priors under reparametrization. Also, it has proved to be remarkably successful in one-dimensional problems. Jeffreys himself, however, noticed difficulties with the method when the parameter is multidimensional and then provided ad hoc modifications to the prior.

Reference priors introduced by Bernardo (1979) and further developed by Berger and Berger and Bernardo (1989, 1992a,b,c) is a method to achieve posterior distributions which produces objective Bayesian inference, meaning that inferential statements depend only on the postulated model and the available data. Moreover, there is the requirement that the prior distribution is minimally informative in a precise information-theoretic sense. The information provided by the data should dominate the prior information, reflecting the vague nature of the prior knowledge. The driving idea is to maximize the expected Kullback-Leibler divergence of the posterior distribution with respect to the prior. Starting from a reference prior, the reference posterior is a consequence of a formal application of the Bayes theorem. Reference analysis provides posterior distributions with some nice properties, such as invariance, consistent marginalization and consistent sampling properties. See Bernardo (2005) for a recent discussion of these ideas.

The paper is organized as follows. In Section 2, we present the new Poisson-exponential (PE) distribution and discuss its properties, where most of the review material is based on Cancho et al. (2011). In Section 3 we carry out Bayesian inference for this model. Some measures of model selection are presented in Section 3.3 and an comparison of the PE model with the exponential model is presented in Section 4. In Section 3.4 a simulation study is presented. In Section 4 the methodology is illustrated in a real data set. Some final comments are presented in Section 5.

2. Formulation of the Model

Following Cancho et al. (2011), the PE model can be derived as follows. Let M be a random variable denoting the number of CR related to the occurrence of an event of interest. Further, assume that M has a zero truncated Poisson distribution with probability mass function given by

$$\mathrm{P}(M=m) = \frac{e^{-\theta}\theta^{m}}{m![1-e^{-\theta}]},\ m=1,2,\ldots,\ \theta>0. \tag{1}$$

Let T_j $(j=1,2,\ldots)$ denote the time-to-event due to the jth CR, hereafter lifetime. Given $M=m$, the random variables T_j, $j=1,\ldots,m$ are assumed to be independent and identically distributed with a common distribution function with pdf given by

$$f(t;\lambda) = \lambda e^{-\lambda t},\quad t>0,\quad \lambda>0. \tag{2}$$

In the latent CR scenario, the number of causes M and the lifetime T_j associated with a particular cause are not observable, but only the maximum lifetime Y among all causes is usually observed. So, the component lifetime is defined as

$$Y = \max\left(T_1,\ldots,T_M\right). \tag{3}$$

The following result shows the distribution of Y.

Proposition 2.1 If the random variable Y is defined as in Equation (3), then considering Equations (1) and (2), Y is distributed according to a Poisson-exponential (PE) distribution with pdf given by

$$f(y) = \frac{\theta\lambda e^{-\lambda y-\theta e^{-\lambda y}}}{1-e^{-\theta}},\quad y>0. \tag{4}$$

Proof The conditional density function of Equation (3) given $M=m$ is given by

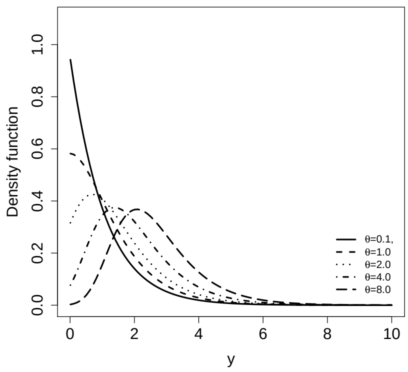
$$f(y|M=m,\lambda) = m\lambda(1-e^{-\lambda y})^{m-1}e^{-\lambda y},\ y>0,\ m=1,\ldots.$$

Then, the marginal pdf of Y is given by

$$
\begin{aligned}
f(y) =& \sum_{m=1}^{\infty} m\lambda(1-e^{-\lambda y})^{m-1}e^{-\lambda y} \times \frac{\theta^m e^{-\theta}}{m![1-e^{-\theta}]} \\
=& \frac{\theta\lambda e^{-\theta-\lambda y}}{1-e^{-\theta}} \sum_{m=1}^{\infty} \frac{\left(\theta[1-e^{-\lambda y}]\right)^{m-1}}{(m-1)!} \\
=& \frac{\theta\lambda e^{-\lambda y-\theta e^{-\lambda y}}}{1-e^{-\theta}}.
\end{aligned}
$$

■

From Equation (4), the parameter λ controls the scale of the distribution while the parameter θ controls its shape; see Figure 1. As θ approaches zero, the PE distribution converges to an exponential distribution with parameter λ. Figure 1 (left panel) shows the PE probability density functions for some fixed values of θ. The PE density function is decreasing if $0 < \theta < 1$ and unimodal for $\theta \geq 1$. The modal value λe^{-1} is obtained at $y = \log\theta/\lambda$. The parameters of the PE model have a direct interpretation in terms of complementary risks. The parameter θ represents the mean of the number of complementary risks, while λ denotes the lifetime failure rate.



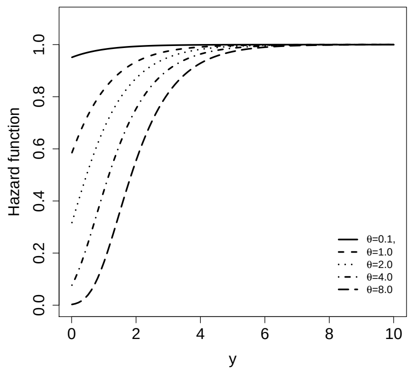


Figure 1. Left panel: probability density function of the PE distribution. Right panel: hazard function of the PE distribution. The parameters were fixed at $\lambda = 1$ and $\theta = 0.1, 1, 2, 4, 8$.

The pth quantile of the PE distribution is expressed by (Cancho et al., 2011),

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

as well as, the survival (or reliability) function of the PE model is given by

$$S(y) = \frac{1 - e^{-\theta e^{-\lambda y}}}{1 - e^{-\theta}}, \quad y > 0. \tag{5}$$

From Equations (4) and (5) it is easy to verify that the hazard function is given

$$h(y) = \frac{\theta\lambda e^{-\lambda y - \theta e^{-\lambda y}}}{1 - e^{-\theta e^{-\lambda y}}}, \quad y > 0. \tag{6}$$

The hazard function given in Equation (6) is increasing. The right panel of Figure 1 shows some hazard function shapes for some fixed values of θ. The initial and long-term hazard values are both finite and are given by $h(0) = \lambda\theta[e^{\theta} - 1]$ and $h(\infty) = \lambda$.

In comparison with the Kus (2007) formulation we follows an opposite way, since he defines the component lifetime as $Y = \min(T_1, \ldots, T_M)$, while, according to Equation (3), we are considering $Y = \max(T_1, \ldots, T_M)$.

2.1 Moments

A general expression for the rth ordinary moment $\mu'_r = \mathrm{E}[Y^r]$ of the PE distribution can be obtained analytically by considering the generalized hypergeometric function denoted by $F_{p,q}(\boldsymbol{a}, \boldsymbol{b}, \theta)$ and defined as $F_{p,q}(\boldsymbol{a}, \boldsymbol{b}, \theta) = \sum_{j=0}^{\infty}[\theta^j \prod_{i=1}^{p} \Gamma(a_i + j)\Gamma(a_i)^{-1}]/[\Gamma(j + 1)\prod_{i=1}^{q} \Gamma(b_i + j)\Gamma(b_i)^{-1}]$, where $\boldsymbol{a} = [a_1, \ldots, a_p]$, p is the number of terms of $\boldsymbol{a}$, $\boldsymbol{b} = [b_1, \ldots, b_q]$, and q is the number of terms of $\boldsymbol{b}$.

In our case, following Cancho et al. (2011), suppose Y is a PE random variable with parameters $\lambda > 0$ and $\theta > 0$, and density function given by Equation (4); then

$$\mu'_r = \mathrm{E}[Y^r] = \frac{\theta\Gamma(r+1)}{\lambda^r[1 - e^{-\theta}]} F_{r+1,r+1}([1, \ldots, 1], [2, \ldots, 2], -\theta), \tag{7}$$

where $F_{p,q}(\boldsymbol{a}, \boldsymbol{b}, \theta)$ is the generalized hypergeometric function. The proof of the Equation 7 is obtained by direct integration; see Cancho et al. (2011).

Considering Equation (7), the mean and variance of the distribution are given, respectively, by

$$\begin{aligned}
\mathrm{E}[Y] &= \frac{\theta}{\lambda[1 - e^{-\theta}]} F_{2,2}\left([1, 1], [2, 2], -\theta\right), \\
\mathrm{Var}[Y] &= \frac{\theta}{\lambda^2[1 - e^{-\theta}]} \left[F_{3,3}\left([1, 1, 1], [2, 2, 2], -\theta\right) - \frac{\theta}{[1 - e^{-\theta}]} F_{2,2}\left([1, 1], [2, 2], -\theta\right)^2 \right].
\end{aligned}$$

The skewness and kurtosis of the PE distribution can be computed as

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \quad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3,$$

where μ_2, μ_3 and μ_4 are the second, third and fourth central moments, respectively, and can be represented in terms of the generalized hypergeometric functions:

$$\begin{aligned}
\mu_2 =& \frac{1}{\lambda^2}\left[\frac{\theta F_{3,3}(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{1 - e^{-\theta}} - \frac{\theta^2 F_{2,2}^2(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})^2}\right], \\
\mu_3 =& \frac{1}{\lambda^3}\left[\frac{2\theta^3 F_{2,2}^3(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})^3} - \frac{3\theta^2 F_{2,2}(\boldsymbol{a}, \boldsymbol{b}, -\theta) F_{3,3}(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})^2} + \frac{12\theta F_{4,4}(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})}\right], \\
\mu_4 =& \frac{1}{\lambda^4}\left[\frac{-3\theta^4 F_{2,2}^4(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})^4} + \frac{6\theta^3 F_{2,2}(\boldsymbol{a}, \boldsymbol{b}, -\theta) F_{3,3}(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})^3} - \right. \\
& \left. \frac{48\theta^2 F_{2,2}(\boldsymbol{a}, \boldsymbol{b}, -\theta) F_{4,4}(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})^2} + \frac{24\theta F_{5,5}(\boldsymbol{a}, \boldsymbol{b}, -\theta)}{(1 - e^{-\theta})}\right],
\end{aligned}$$

where $\boldsymbol{a} = [1, \ldots, 1]$ and $\boldsymbol{b} = [2, \ldots, 2]$. It should be stressed that vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ have variable dimension in the above formulas. The skewness and kurtosis are both independent of the scale parameter. Figure 2 shows a graphical representation of skewness and kurtosis. It is observed that both skewness and kurtosis are decreasing functions of θ, moreover, the limiting value of the skewness is approximately 1.139712 and the limiting value of the kurtosis is approximately 2.400126.

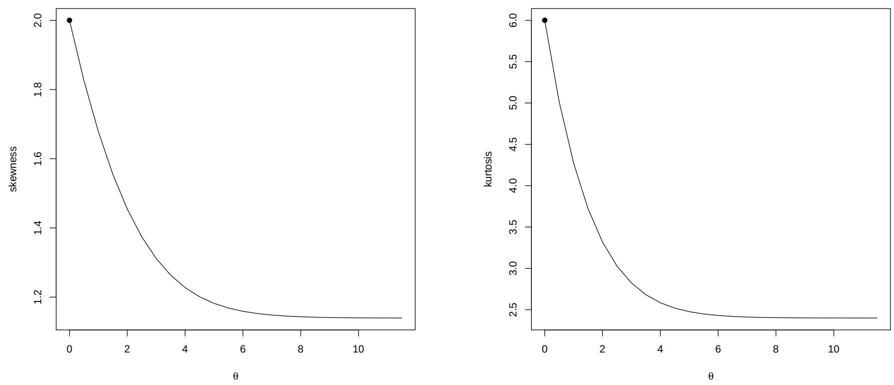


Figure 2. Left panel: skewness of the PE distribution. Right panel: kurtosis of the PE distribution.

3. Reference Analysis

The declared objective of reference Bayesian analysis introduced by Bernardo (1979) and further developed by Berger and Bernardo (1992a) and Berger and Bernardo (1992b) is to specify a prior distribution such that, even for moderate sample sizes, the information provided by the data should dominate the prior information because of the "vague" nature of the prior knowledge.

An important feature in the Berger-Bernardo approach to construct a non-informative prior is the different treatment for interest and nuisance parameters. When there are nuisance parameters (typical case in this paper), one must establish an ordered parametrization with the parameter of interest singled out and then follow the procedure below.

Proposition 3.1 Let $f(\mathbf{x}|\theta,\lambda)$, $(\theta,\lambda) \in \Theta \times \Lambda \subseteq \mathfrak{R} \times \mathfrak{R}$ be a probability model with two real-valued parameters θ and λ, where θ is the quantity of interest. Let $I(\theta,\lambda)$ the corresponding 2x2 Fisher's matrix in terms of θ and λ, and let $V(\theta,\lambda) = I^{-1}(\theta,\lambda)$. Suppose that the joint posterior distribution of (θ,λ) is asymptotically normal with covariance matrix $V(\hat{\theta},\hat{\lambda})$, where $\hat{\theta}$ and $\hat{\lambda}$ are the corresponding consistent estimators of θ and λ. It follows that:

(i) the conditional reference prior of λ given θ is

$$\pi(\lambda|\theta) \propto I_{22}(\theta,\lambda)^{1/2} \qquad\qquad \lambda \in \Lambda(\theta);$$

(ii) if $\pi(\lambda|\theta)$ is not proper, a compact approximation $\{\Lambda_i(\theta),\ i = 1, 2, \dots,\}$ to $\Lambda(\theta)$ is required, and the reference prior of λ given θ is

$$\pi_i(\lambda|\theta) = \frac{I_{22}(\theta,\lambda)^{1/2}}{\int\limits_{\Lambda_i(\theta)} I_{22}(\theta,\lambda)^{1/2} d\lambda}, \qquad\qquad \lambda \in \Lambda_i(\theta);$$

(iii) the sequence of priors can be obtained as

$$\pi_i(\theta) \propto e^{\int_{\Lambda_i(\theta)} \pi_i(\lambda|\theta)\log\left(v_{11}^{1/2}(\theta,\lambda)\right)d\lambda},$$

where $v_{11}^{1/2}(\theta,\lambda) = I_\theta(\phi,\lambda) = I_{11} - I_{12}I_{22}^{-1}I_{21}$;

(iv) the reference posterior distribution of θ given data $\{y_1,\ldots,y_n\}$ is

$$\pi(\theta|y_1,\ldots,y_n) \propto \pi(\theta)\left\{\int_{\Lambda(\theta)}\left(\prod_{l=1}^{n} p(y_l|\theta,\lambda)\right)\pi(\lambda|\theta)\,d\lambda\right\}.$$

Proof See a heuristic justification in Bernardo (2005). ■

Corollary 3.2 If the nuisance parameter space $\Lambda(\theta) = \Lambda$ is independent of θ, and the functions $v_{11}^{-1/2}(\theta,\lambda)$ and $I_{22}^{1/2}(\theta,\lambda)$ factorize in the form

$$(v_{11}(\theta,\lambda))^{-1/2} = f_1(\theta)\,g_1(\lambda), \qquad (I_{22}(\theta,\lambda))^{1/2} = f_2(\theta)\,g_2(\lambda).$$

Then

$$\pi(\theta) \propto f_1(\theta), \qquad \pi(\lambda|\theta) \propto g_2(\lambda).$$

Thus, the reference prior relative the ordered parametrization (θ,λ) is given by

$$\pi(\lambda,\theta) = f_1(\theta)\,g_2(\lambda),$$

and there is no need for compact approximation, even if the conditional reference prior $\pi(\lambda|\theta)$ is not proper; Bernardo (2005).

Proof See proof of Theorem 12 in Bernardo (2005). ■

3.1 Prior and posterior densities

According to Bernardo (2005), let $\boldsymbol{y}^k = (y_1,\ldots y_k)$ k-independent replications of the PE model and consider $h(\boldsymbol{\vartheta}) = 1$ a positive function. Then, $q(\boldsymbol{\vartheta}, y_1,\ldots y_k) \propto L(\boldsymbol{\vartheta})h(\boldsymbol{\vartheta})$ is an asymptotic approximation of posterior distribution. Under certain regularity conditions whereas there is a maximum likelihood estimator $(\hat{\theta}(\boldsymbol{y}^k), \hat{\lambda}(\boldsymbol{y}^k))$ we have that the posterior density $q(\boldsymbol{\vartheta}, y_1,\ldots y_k)$ is approximately normal k-dimensional, i.e,

$$q(\boldsymbol{\vartheta}, y_1,\ldots y_k) \sim \mathrm{N}_k\left((\hat{\theta},\hat{\lambda}), nI(\hat{\theta},\hat{\lambda})\right),$$

where $I(\hat{\theta},\hat{\lambda})$ is the Fisher information matrix; see Bernardo (2005).

Considering that the posterior distribution is asymptotically normal, then the reference prior only depends on Fisher information matrix. Here we derive the reference prior considering the approach of one nuisance parameter described above.

The likelihood function of $\boldsymbol{\vartheta} = (\theta,\lambda)$ based on the observed sample of size n, $\boldsymbol{y} = (y_1,\ldots,y_n)$, from the PE distribution is given by

$$L(\boldsymbol{\vartheta}) = e^{n\log(\theta\lambda) - \lambda\sum_{i=1}^{n} y_i - \theta\sum_{i=1}^{n} e^{-\lambda y_i} - n\log(1-e^{-\theta})}. \tag{8}$$

From Equation (8), the Fisher information matrix are given by,

$$\mathbf{I}(\boldsymbol{\vartheta}) = n \begin{bmatrix} \frac{1}{\theta^2} - \frac{e^{\theta}}{(e^{\theta}-1)^2} & -\frac{\theta F_{2,2}([2,2],[3,3],-\theta)}{4\lambda(1-e^{-\theta})} \\ -\frac{\theta F_{2,2}([2,2],[3,3],-\theta)}{4\lambda(1-e^{-\theta})} & \frac{1}{\lambda^2}\left[1+\frac{\theta^2 F_{3,3}([2,2,2],[3,3,3],-\theta)}{4(1-e^{-\theta})}\right] \end{bmatrix}. \tag{9}$$

From Corollary 3.2 and Equation (9), the joint reference prior of θ and λ is given by

$$\pi(\lambda, \theta) \propto \frac{\pi(\theta)}{\lambda}, \tag{10}$$

where

$$\pi(\theta) \propto \sqrt{\varphi(\theta)}\left(1 + \frac{\theta^2 F_{3,3}([2,2,2],[3,3,3],-\theta)}{4(1-e^{-\theta})}\right)^{-1/2} \tag{11}$$

and

$$\varphi(\theta) = \left[\frac{1}{\theta^2} - \frac{e^{\theta}}{(e^{\theta}-1)^2}\right]\left[1 + \frac{\theta^2 F_{3,3}([2,2,2],[3,3,3],-\theta)}{4(1-e^{-\theta})}\right] - \frac{\theta^2 F_{2,2}^2([2,2],[3,3],-\theta)}{16(1-e^{-\theta})^2}.$$

Combining the likelihood function, $L(\boldsymbol{\vartheta})$ in Equation (8) and the prior specification given in Equation (10), the joint posterior distribution for θ, λ is given by,

$$\pi(\theta, \lambda|\mathbf{y}) \propto \frac{\theta^n \lambda^{n-1}}{(1-e^{-\theta})^n} e^{-\lambda\sum_{i=1}^n y_i - \theta\sum_{i=1}^n e^{-\lambda y_i}}\pi(\theta). \tag{12}$$

Theorem 3.3 Under the PE model, with prior given in Equation (11), it follows that the joint posterior distribution given in Equation (12) is proper.

Proof Since $e^{-\lambda y_i} \le 1$, for all $y_i > 0$, $(i = 1, \ldots, n)$ and $\lambda > 0$, it follows that

$$\begin{aligned}\int_0^\infty\int_0^\infty \pi(\boldsymbol{\theta}, \lambda|\mathbf{y})d\theta d\lambda &\le \int_0^\infty\int_0^\infty \frac{\theta^n\lambda^{n-1}e^{-n\theta}e^{-\lambda\sum_{i=1}^n y_i}}{(1-e^{-\theta})^n}\pi(\theta)d\theta d\lambda \\ &= \int_0^\infty \frac{\theta^n e^{-n\theta}}{(1-e^{-\theta})^n}\pi(\theta)\int_0^\infty \lambda^{n-1}e^{-\lambda\sum_{i=1}^n y_i}d\lambda d\theta \\ &= \frac{\Gamma(n)}{\left(\sum_{i=1}^n y_i\right)^n}\int_0^\infty \frac{\theta^n e^{-n\theta}}{(1-e^{-\theta})^n}\pi(\theta)d\theta \\ &\le \frac{\Gamma(n)}{\left(\sum_{i=1}^n y_i\right)^n} < \infty,\end{aligned}$$

since $\theta e^{-\theta}/(1-e^{-\theta}) \le 1$ for $\theta > 0$ and $\int_0^\infty \theta^n e^{-n\theta}/(1-e^{-\theta})^n\pi(\theta)d\theta = 0.272897$ (obtained numerically). ■

3.2 Computation

In the Bayesian approach, the target distribution for inference is the posterior of the parameters of interest. Thus, we need to obtain the marginal posterior densities of such a parameter. The posterior distribution is proper considering the reference prior given in Equation (10). However, irrespective to the prior distribution chosen, the marginal posterior distributions for the parameters of the proposed model are analytically intractable. We then consider the use of Markov chain Monte Carlo methods (MCMC), e.g. Gibbs Sampling and Metropolis-Hastings algorithm; see, e.g., Chib and Greenberg (1995).

The Gibbs sampler is an iterative procedure of a broad class of methods generically named Markov Chain Monte Carlo (MCMC). Many practical aspects of MCMC are described in Gelfand and Smith (1990) and Gamerman and Lopes (2006). This method is applicable in situations where one is not able to generate samples directly from the joint posterior density. It requires the full conditional densities for generating samples. The full conditional densities for θ and λ, obtained from Equation (12), are given by

$$\pi(\theta|\lambda, \mathbf{y}) \propto \theta^n e^{-\theta \sum_{i=1}^n e^{-\lambda y_i} - n \log(1-e^{-\theta})} \pi(\theta)$$

and

$$\pi(\lambda|\theta, \mathbf{y}) \propto \lambda^{n-1} e^{-\lambda \sum_{i=1}^n y_i - \theta \sum_{i=1}^n e^{-\lambda y_i}}.$$

Since the conditional densities above do not belong to any known parametric family, in order to generate our samples we then implement a Metropolis-Hasting algorithm within Gibbs iterations; see Chib and Greenberg (1995). For example, to implement the Metropolis-Hastings algorithm for the parameter θ, we consider a target distribution $g_\theta(\theta) = \pi(\theta|\lambda, \boldsymbol{y})$, and under given model, $\theta > 0$ we consider the transformation $\eta = \log(\theta)$, where, $-\infty < \eta < \infty$. Then, $g_\eta(\eta) = g_\theta(\eta)e^\eta$.

Instead of directly sampling θ, we generate η by a random-walk. That is, we consider transition kernels $q(\eta, w)$, mapping η to w such that $w \leftarrow \eta + \sigma z$, with $z \sim \mathrm{N}(0, \tau^2)$ and σ is parameter of scale that controle the rate of acceptance of algorithm. Following Carlin and Louis (2001), we choose τ^2 as the correspondent diagonal element of the inverse of the minus the matrix of second derivative of logarithm of joint posteriori distribution of $\eta = \log(\theta)$ and $\gamma = \log(\lambda)$, $\pi(\eta, \gamma|\boldsymbol{y})$, evaluated in the mode posteriori of $\pi(\eta, \gamma|\boldsymbol{y})$. The algorithm to generate η operates as follows:

(1) let η be the current value;
(2) generate a point η^* according to the transitional kernel $q(\eta, w)$;
(3) a move from η to η^* is made with probability $\min\{1, g_\eta(\eta^*)/g_\eta(\eta)\}$.

After we sample η we obtain $\theta = e^\eta$. We assume τ a same value in all step of algorithm.

In our case, as in Carlin and Louis (2001), we choose the scale parameter to be equal to two, i.e, $\sigma = 2$. However, other higher values were also considered but the convergence becomes slow and rate of acceptance low. To implement the Metropolis-Hastings algorithm for the parameter λ, we proceed in the same way as for the parameter θ, but considering a target distribution $g_\lambda(\lambda) = \pi(\lambda|\theta, \mathbf{y})$ and respective diagonal element of the minus the matrix of second derivative of logarithm of joint posteriori distribution. The Metropolis-Hasting algorithm within Gibbs is shown in Appendix A.

3.3 Model comparison strategies

In the literature, there are various methodologies which intend to analyze the suitability of a model, as well as selecting the best fitting among a collection of models. In this paper we shall consider some of these Bayesian model selection criterion, which penalize the number of parameters in the model. Namely, the deviance information criteria (DIC) (Spiegelhalter et al., 2002), the expected Akaike information criterion (EAIC) (Brooks, 2002), and the expected Bayesian (or Schwarz) information criterion (EBIC) (Carlin and

Louis, 2001). These criteria are based on the posterior mean of the deviance, $\mathrm{E}[D(\boldsymbol{\vartheta})]$, which is also a measure of fit and can be approximated from the MCMC output by $\mathrm{Dbar} = [1/B]\sum_{b=1}^{B} D(\boldsymbol{\vartheta}^{(b)})$, where the index b indicates the bth realization of a total of B realizations and $D(\boldsymbol{\vartheta}) = -2\sum_{i=1}^{n}\log(f(y_i|\boldsymbol{\vartheta}))$, where $f(\cdot)$ is the probability density of the PE distribution. The EAIC, EBIC and DIC criterion can be calculated using the MCMC output by means of $\widehat{\mathrm{EAIC}} = \mathrm{Dbar} + 2q$, $\widehat{\mathrm{EBIC}} = \mathrm{Dbar} + q\log(n)$ and $\widehat{\mathrm{DIC}} = \mathrm{Dbar} + \widehat{\rho_D} = 2\mathrm{Dbar} - \mathrm{Dhat}$, respectively, where q is the number of parameters in the model and ρ_D is the effective number of parameters, defined as $\mathrm{E}[D(\boldsymbol{\vartheta})] - D(\mathrm{E}[\boldsymbol{\vartheta}])$, where $D(\mathrm{E}[\boldsymbol{\vartheta}])$ is the deviance evaluated at the expected values of the posterior distributions, and can be estimated as $\mathrm{Dhat} = D([1/B]\sum_{b=1}^{B}\boldsymbol{\vartheta}^{(b)})$. Having to compare alternative models, the preferred model is the one with the smallest value of the criterion. In our case, for instance, we shall be interested in comparing the PE and the exponential models. This will be done in Section 4.

Moreover, in order to decide for the best model we can use the Bayes factors which is the relative weight of evidence for model M_1 compared to model M_2 given by

$$B_{12} = \frac{f(\boldsymbol{t}_0|M_1)}{f(\boldsymbol{t}_0|M_1)}, \tag{13}$$

where $\boldsymbol{t_0}$ denotes the actual observations and $f(\boldsymbol{t}_0|M_j)$ denotes the marginal density under model M_j, j=1,2; see Gelfand (1996); Louzada et al. (2007). The model M_1 is preferred over M_2 when $B_{12} > 1$; see Kass and Raftery (1995) for more details. We approximate the marginal densities in Equation (13) by their Monte Carlo estimates, obtained from the R generated samples, given by $[1/B]\sum_{b=1}^{B} f(\boldsymbol{t}_0|\boldsymbol{\vartheta}_j^{(b)}, M_j)$. For more details on Bayesian discrimination, interested readers may refer to Robert and Wraith (2009) for a survey of some recent approaches on Bayes factor used in Bayesian hypothesis testing and in Bayesian model choice, and to Marin and Robert (2010) for methods for Bayesian discrimination between embedded models.

3.4 Frequentist properties

Non-subjective posterior credible interval are often numerically very close, and sometimes identical, to the frequentist confidence intervals based on sufficient statistics; see Jaynes (1976). The analysis on the frequentist coverage probabilities of credible interval derived from non-subjective posteriors, in an attempt to verify whether or not they are "well calibrated" and it does provide some bridge between frequentist and Bayesian inference. Reference within this topic include Lindley (1958) and Datta and Ghosh (1996). This section presents some frequentist properties of the estimators of θ and λ based on the non-informative prior proposed here. We focus on the frequentist mean squared error and on the frequentist coverage probability of 95% credible intervals for different samples sizes, $n = 30, 50, 100, 150$ and 200. For each set up $1,000$ generated samples were considered.

In this study we considered the PE distribution given in Equation (4) with parameters $\theta = 5$ and $\lambda = 2$. For each simulated data set, we obtained the posterior summaries of the parameters. We simulated two parallel chains of size 10,000 for each parameter, disregarding the first 5,000 iterations to eliminate the effect of the initial values obtaining a sample of size 5,000. For each setup, we conducted 1,000 replicates and then we averaged the estimates of parameters, and calculate the mean square error (MSE), the coverage of the lower HPD bound (L), the coverage of the upper HDP bound (U) and the coverage of the 95% HPD intervals (C) for θ and λ. The results are summarized in Table 1. The empirical MSEs decrease as the sample size increases and the differences between the average estimates and the true values are almost always smaller than one empirical MSE.

Also, for instance, for the sample size 30, $L = 0$ and $U = 0.049$, indicates that 4.9% of the posterior estimates that are not covered by the 95% HPD are located above the upper bound. Although, balance between lower and upper HPD interval bounds is only observed for large samples, the empirical coverage probabilities are close to the nominal level even for small sample sizes.

Table 1. Means, empirical MSEs, coverage probabilities of the lower HPD bound (L), upper HDP bound (U) and the 95% HPD intervals (C) for θ and λ.

	θ					λ				
n	Mean	MSE	L	U	C	Mean	MSE	L	U	C
30	4.930	1.730	0.000	0.049	0.951	1.963	0.334	0.013	0.037	0.950
50	4.956	1.244	0.002	0.051	0.947	1.972	0.252	0.011	0.045	0.944
75	4.971	0.994	0.003	0.039	0.958	1.979	0.204	0.013	0.048	0.939
100	4.974	0.893	0.010	0.042	0.948	1.990	0.188	0.022	0.034	0.944
150	4.973	0.718	0.015	0.036	0.949	1.988	0.159	0.019	0.036	0.945
200	5.022	0.609	0.019	0.026	0.955	1.997	0.126	0.025	0.034	0.941

We also evaluate the performance of the EAIC criterion via Monte Carlo simulation study considering the samples generated as above. The idea is to verify the percentage of times the PE distribution, which originated the sample, was the best fitted model when compared to the exponential distribution. Table 2 summarizes the results of the simulation study. The percentage of times the PE distribution was the best fitted model increases with θ and sample size n. We point out that DIC and EBIC criterion were also evaluated via Monte Carlo simulation, but the results are very similar to the results obtained to the EAIC criterion and are then omitted.

4. Deep Groove Ball Bearings Data

We reanalyze the data extracted from Lawless (2003, p. 98), which arises in tests on endurance of deep groove ball bearings. The data consist of the number of million revolution before failure for each of the 23 ball bearings in a life test.

Firstly, in order to identify the shape of a lifetime data hazard function we shall consider a graphical method based on the total time on test (TTT) plot; see Aarset (1987). In its empirical version the TTT plot is given by $G(r/n) = [(\sum_{i=1}^{r} Y_{i:n}) - (n-r)Y_{r:n}]/(\sum_{i=1}^{r} Y_{i:n})$, where $r = 1, \dots, n$ and $Y_{i:n}$ represent the order statistics of the sample. It has been shown that the hazard function increases (decreases) if the TTT plot is concave (convex). Although the TTT plot is only a sufficient condition, not a necessary one for indicating the hazard function shape, it is used here as a crude indicator of its shape. Figure 3 (left panel) shows the TTT plot for the considered data, which is concave indicating an increasing hazard function, that can be accommodated by a PE distribution.

Table 2. Percentage of times the PE model, which originated the sample was the best fitted model.

	Sample size					
θ	30	50	75	100	150	200
0.1	19.3	24.3	21.7	23.3	22.7	21.0
1.0	38.7	46.3	55.5	63.8	74.6	81.0
2.0	71.7	85.9	92.7	97.7	99.7	99.9
4.0	99.0	99.9	100.0	100.0	100.0	100.0
8.0	100.0	100.0	100.0	100.0	100.0	100.0

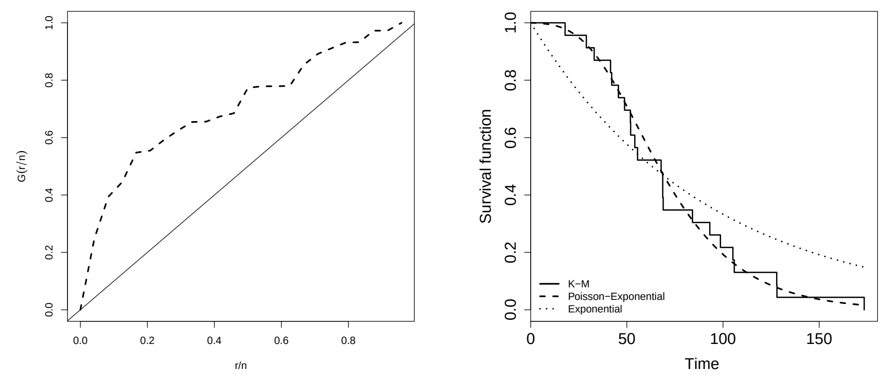


Figure 3. Left panel: empirical scaled TTT-Transform for the data. Right panel: Kaplan-Meier curve with estimated survival functions of the Poisson-exponential and exponential distributions.

Then, the PE and the exponential models were fitted the data via our Bayesian reference approach. We ran two parallel independent runs of the Gibbs sampler with size 20,000 for each parameter, disregarding the first 5,000 iterations to eliminate the effect of the initial values and, to avoid correlation problems, we considered a spacing of size 10, obtaining a sample of size 3,000. To monitor the convergence of the Gibbs sampler we resorted to the methods recommended by Cowles and Carlin (1996). MCMC computations were implemented in the R system (R, 2008); see Appendix. Table 3 shows the posterior summaries for the parameters of both distributions and the three model comparison criteria. The PE distribution outperforms the exponential one irrespective of the criterion to be used.

Table 3. Posterior means and 95% HDP intervals for the parameters, in brakets, and Bayesian comparison criteria.

Model	λ	θ	DIC	EAIC	EBIC
Exponential	0.01343		244.800	245.940	247.090
	(0.0015 , 0.022202)				
PE	0.035	7.095	230.243	232.290	234.561
	(0.024 , 0.047)	(3.082, 13.125)			

The right panel of Figure 3 exhibits the Kaplan–Meier estimate of the survival function and the Bayesian estimates from the exponential and PE distributions. The poorest fit is achieved with the exponential distribution. At earlier times the PE model yields a closer concordance with the Kaplan–Meier estimates. Overall, confronting the PE model with the exponential distribution, since in Table 3 the estimate of θ is statistically different from zero, $0 \notin (3.082; 13.125)$, we have evidence in favor of the PE distribution. Further, Table 3 shows the posterior summaries for the parameters of both distributions and the three model comparison criteria. The PE distribution outperforms the exponential one irrespective of the criterion to be used.

5. Concluding Remarks

In this paper we present the PE distribution with a formal derivation of its pdf, showing its survival and failure rate functions, quantiles and moments (particularly, the mean, variance, skewness and kurtosis). The parameters of the PE distribution have a direct interpretation in terms of CR. Also we present a motivation for deriving the reference posterior distribution in the case of one nuisance parameter. We discuss Bayesian inference via MCMC, including a straightforwardly model comparison procedure. We have illustrated the theory considering simulated data and an important example extracted from Lawless (2003), suggesting that the PE distribution outperforms the exponential one, irrespective of the criteria used. The codes used in the paper can be obtained by emailing the authors.

Appendix A: Code of the Metropolis-Hasting algorithm within Gibbs

```
library(hypergeo)

##########################################################################
### psi function                                                       ###
##########################################################################
psi=function(theta){((((1/theta^2-exp(theta)/(exp(theta)-1)^2)*
(1+genhypergeo(c(2,2,2),c(3,3,3),-theta)*theta^2/(4*(1-exp(-theta)))-
(genhypergeo(c(2,2),c(3,3),-theta)*theta/(4*(1-exp(-theta))))^2))
}

###########################################################################
### Log_posteriori function                                             ###
###########################################################################
log_posteriori=function(vpar){
theta=exp(vpar[1])
lambda=exp(vpar[2])
adFunc=0
for(i in 1:n)
   {
vf=theta*lambda*exp(-lambda*t[i]-theta*exp(-lambda*t[i]))/(1-exp(-theta));
adFunc=adFunc+log(vf);
 }
adFunc=adFunc +log(theta)+0.5*log(abs(psi(theta)))-
0.5*log(1+genhypergeo(c(2,2,2),c(3,3,3),-theta)*theta^2/(4*(1.00000-exp(-theta))))
return(adFunc)
}


###########################################################################
### Gibbs with Metropolis-Hasting algorithm                             ###
### R: Iteration Number; burn: Burn in; vpar: vector parameters         ###
### log_posteriori: logarithm of posteriori density                     ###
###########################################################################
GM_H=function(vpar, log_posteriori, scale=2,R, burn=10){
fit=optim(vpar, log_posteriori,   hessian = TRUE, control = list(fnscale = -1),
method = "BFGS")
   start = fit$par
sD=sqrt(diag(-solve(fit$hessian)))
p = length(start)
vth = array(0, dim = c(R, p))
f0 = log_posteriori(start)
arate = array(0, dim = c(1, p))

th0 = start
th1 = th0
  for (i in -burn:R) {
        for (j in 1:p) {
            th1[j] = th0[j] + scale*rnorm(1) * sD[j]   # proposal
            f1 = log_posteriori(th1)
            proba = exp(f1 - f0)
         if (is.na(proba) == FALSE) {
            u = runif(1) < proba
            th0[j] = th1[j] * (u == 1) + th0[j] * (u == 0)
            f0 = f1 * (u == 1) + f0 * (u == 0)
            vth[i, j] = th0[j]
            arate[j] = arate[j] + u
        }
    }
  }
    arate = arate/(R-burn+1)
    saida = list(par = exp(vth), accept = arate)
   }


################################################################
## Example                                                   ###
## t:  data vector                                           ###
################################################################
n=length(t)
vpar=c(1,1/mean(t))    # start values
saida1=GM_H(vpar, log_posteriori,R=15000, burn=5000)
```

Acknowledgements

We are grateful to the Editorial Board and to the referees for their very valuable comments and suggestions, which led to improvements at the paper. The research was partially supported by CNPq and FAPESP of Brazil.

References

Aarset, M.V., 1987. How to identify a bathtub hazard rate. IEEE Transactions on Reliability, 36, 106–108.

Barreto-Souza, W., Cribari-Neto, F., 2009. A generalization of the exponential-Poisson distribution. Statistics and Probability Letters, 79, 2493–2500.

Basu, A., Klein, J., 1982. Some recent development in competing risks theory. In Crowley, J., Johnson, R.A., (eds.). Survival Analysis. Institute of Mathematical Statistics, Hayward, pp. 216–229.

Berger, J.O., Bernardo, J.M., 1989. Estimating a product of means: Bayesian analysis with reference priors. Journal of the American Statistical Association, 84, 200–207.

Berger, J.O., Bernardo, J.M., 1992a. Ordered group reference priors with applications to a multinomial problem. Biometrika, 79, 25–37.

Berger, J.O., Bernardo, J.M., 1992b. Reference priors in a variance components problem. In Goel, P.K., Iyengar, N.S., (eds.). Bayesian Analysis in Statistics and Econometrics. Springer, Berlin, pp. 323–340.

Berger, J.O., Bernardo, J.M., 1992c. On the development of reference priors. In Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., (eds.). Bayesian Statistics 4. Oxford University Press, Oxford, pp. 35–60.

Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference. Journal of The Royal Statistical Society Series B - Statistical Methodology, 41, 113–147.

Bernardo, J.M., 2005. Reference analysis. In Dey, D.K., Rao, C.R., (eds.). Handbook of Statistics. Elsevier, Amsterdam, pp. 17–90.

Brooks, S.P., 2002. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde (2002). Journal of The Royal Statistical Society Series B - Statistical Methodology, 64, 616–618.

Cancho, V.G., Louzada-Neto, F., Barriga, G.D.C., 2011. The Poisson-exponential lifetime model. Computational Statistics and Data Analysis, 55, 677–686.

Carlin, B.P., Louis, T.A., 2001. Bayes and Empirical Bayes Methods for Data Analysis. Second edition. Chapman & Hall/CRC, Boca Raton.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician, 49, 327–335.

Cowles, M.K., Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association, 91, 883–904.

Cox, D.R., Oakes, D., 1984. Analysis of Survival Data. Chapman & Hall, London.

Crowder, M.J., Kimber, A.C., Smith, R.L., Sweeting, T.J., 1991. Statistical Analysis of Reliability Data. Chapman & Hall, London.

Datta, G.S., Ghosh, M., 1996. On the invariance of noninformative priors. The Annals of Statistics, 24, 141–159.

Davison, A., Louzada-Neto, F., 2000. Inference for the poly-Weibull model. The Statistician, 49, 189–196.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Second edition. Chapman & Hall/CRC, Boca Raton, FL.

Gelfand, A.E., 1996. Model determination using sampling-based methods. In Gilks, W.R., Richardson, S., Spiegelhater, D.J., (eds.). Markov Chain Monte Carlo in Practice. Chapman & Hall, London.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, 85, 398–409.

Goetghebeur, E., Ryan, L., 1995. A modified log rank test for competing risks with missing failure type. Biometrika, 77, 207–211.

Gupta, R., Kundu, D., 1999. Generalized exponential distributions. Australian and New Zealand Journal of Statistics, 41, 173–188.

Jaynes, E.T., 1976. Confidence intervals vs Bayesian intervals. In Harper, W.L., Hooker, C.A., (eds.). Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science. D. Reidel Publishing Company Dordrecht-Holland, pp. 175–257.

Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences, 186, 453–461.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association, 90, 773–795.

Kus, C., 2007. A new lifetime distribution. Computational Statistics and Data Analysis, 51, 4497–4509.

Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data. Second edition. Wiley, New York.

Lindley, D.V., 1958. Fiducial distributions and Bayes theorem. Journal of The Royal Statistical Society Series B - Statistical Methodology, 20, 102–107.

Louzada-Neto, F., 1999. Poly-hazard regression models for lifetime data. Biometrics, 55, 1121–1125.

Louzada-Neto, F., Tomazella, V.L.D., Andrade, M.G., 2007. Bayesian modeling for recurrent lifetime data with a nonhomogeneous Poisson process with a frailty term with a gamma or inverse gaussian distribution. J. Applied Statistical Science, 15, 413–424.

Lu, K., Tsiatis, A.A., 2001. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. Biometrics, 57, 1191–1197.

Lu, K., Tsiatis, A.A., 2005. Comparison between two partial likelihood approaches for the competing risks model with missing cause of failure. Lifetime Data Analysis, 11, 29–40.

Marin, J.-M., Robert, C.P., 2010. Importance sampling methods for Bayesian discrimination between embedded models. In Chen, M.-H., Dey, D.K., Müller, P., Sun, D., Ye, K., (eds.). Frontiers of Statistical Decision Making and Bayesian Analysis. Chapter 14, pp. 513–553.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at `http://www.R-project.org`.

Reiser, B., Guttman, I., Lin, D.K.J., Guess, M., Usher, J.S., 1995. Bayesian inference for masked system lifetime data. Applied of Statistics, 44, 79–90.

Robert, C.P., Wraith, D., 2009. Computational methods for Bayesian model choice. Bayesian Inference and Maximum Entropy Methods in Science and Engeneering: The 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conference Proceedings. pp. 251–262.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of The Royal Statistical Society Series B - Statistical Methodology, 64, 583–639.