ORIGINAL RESEARCH PAPER

Exponential-Poisson distribution: estimation and applications to rainfall and aircraft data with zero occurrence

Francisco Louzada[a], Pedro Luiz Ramos[a*] and Paulo Henrique Ferreira[b]

[a]Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, Brazil

[b]Institute of Mathematics and Statistics, Federal University of Bahia, Salvador, Brazil

ARTICLE HISTORY
Compiled March 7, 2021

ABSTRACT

In this study, different frequentist estimation procedures for the parameters of the exponential-Poisson distribution are considered, such as the maximum likelihood, method of moments, ordinary and weighted least-squares, percentile, maximum product of spacings, Cramr-von Mises and Anderson-Darling maximum goodness-of-fit estimators. We compare them using extensive numerical simulations, which show that using a nested expectation-maximization algorithm in the maximum likelihood estimators with bootstrap bias correction does not require numerical procedures to solve nonlinear equations and returns accurate parameter estimates. Finally, our proposed methodology is fully illustrated using two real data sets (rainfall and aircraft data) with the occurrence of zero values.

KEYWORDS
Kus distribution; maximum likelihood estimation; Anderson-Darling estimators; meteorological applications; zero occurrence.

1. Introduction

Proposed by Kus (2007), a random variable X has an exponential-Poisson (EP) distribution[1], denoted by $\mathrm{EP}(\lambda, \beta)$, if its probability density function is given by

$$f(x|\lambda,\beta) = \frac{\lambda\beta e^{-\lambda-\beta x+\lambda e^{-\beta x}}}{1-e^{-\lambda}}, \tag{1}$$

for all $x > 0$, where $\lambda > 0$ and $\beta > 0$ are, respectively, the shape and scale parameters. Note that as λ tends to zero, the EP converges to the exponential distribution with parameter β.

Macera et al. (2015) discussed a new model for recurrent event data characterized by a fully parametric baseline rate function, which is based on the EP distri-

*Corresponding author. Email: pedrolramos@usp.br.

[1]The distribution is obtained by mixing exponential and zero-truncated Poisson distributions; see Kus (2007) for details.

bution. Barreto-Souza and Cribari-Neto (2009) and Preda et al. (2011) introduced three-parameter generalizations of the EP distribution. Barreto-Souza and Silva (2015) pointed out that the EP distribution is a good alternative to the gamma distribution for modelling lifetime, reliability and time intervals of successive natural disasters, and proposed a likelihood ratio test based on Cox's statistic to discriminate between these distributions.

Common lifetime distributions, such as gamma, Weibull and lognormal just to list a few, do not allow the occurrence of zero values. Despite the fact that it has been proposed for situations where $x > 0$, the EP distribution (1) can take value on 0, i.e.

$$f_X(0|\lambda, \beta) = \frac{\lambda\beta}{1 - e^{-\lambda}},$$

where $f_X(0|\lambda, \beta) > 0$ for all $\lambda > 0$ and $\beta > 0$. In environmental studies, the occurrence of zero values is common during the study of precipitation, especially in dry periods. Such characteristic is also observed in reliability studies, where instantaneous failures (inliers) may occur due to inferior quality or problems during the construction of the components. This allows the EP distribution to become a good alternative to be used in problems with the occurrence of zero value.

Different inferential procedures for the parameters of the EP distribution have been discussed earlier. Kus (2007) and Karlis (2009) discussed the maximum likelihood estimation (MLE) method via the use of the expectation-maximization (EM) and nested EM algorithms, respectively. Lupu and Lupu (2010) investigated the mixture model of two EP distributions. Considering a Bayesian approach, Yan et al. (2012) derived the Bayes estimators of the parameters in EP model under general entropy, LINEX and scaled squared loss function based on type-II censoring. Furthermore, Xu et al. (2016) studied the Bayes estimators under symmetric and asymmetric loss functions based on general non-information prior distribution. However, the literature offers several classical methods that can be used for estimating the unknown parameters for parametric distributions. Besides, in many cases the MLE method does not perform well for small samples. Therefore, it is of our main interest to compare the MLE method with other estimation procedures, such as the method of moments, least-squares, weighted least-squares, percentile, maximum product of spacings, Cramér-von Mises and Anderson-Darling maximum goodness-of-fit estimators. Similar studies for other distributions have been carried out (Gupta and Kundu 2001; Mazucheli et al. 2013; Teimouri et al. 2013; Louzada et al. 2016; Dey et al. 2017; Rodrigues et al. 2018). In addition, bias correction approaches can be considered for the MLEs (see Efron and Tibshirani (1994) for more details).

In this paper, we compared the different estimation procedures for the EP distribution parameters. For the different procedures, convergence problems in the numerical methods to solve the nonlinear equation system are observed. The nested EM algorithm is the only method that achieves the estimates without numerical instability. However, the obtained estimates are biased for small and moderate sample sizes. Therefore, we considered the bootstrap resampling method, which can be used to reduce bias. A numerical simulation is performed to examine the effect of the estimation and bias correction approaches in the parameter estimates. Additionally, we present some simple closed-form expressions to be used as initial values in the iterative methods to increase the convergence speed. Finally, our proposed methodology is fully illustrated on two new real data sets.

The remainder of this study is organized as follows. In Section 2, we present some properties of the EP distribution. In Section 3, we discuss the estimation methods considered in this paper. In Section 4, a simulation study is presented in order to identify the most efficient estimation procedure. In Section 5, we apply our proposed methodology to rainfall and aircraft data sets, both with zero occurrence. Some final comments are presented in Section 6.

2. Some Properties of the EP Distribution

Statistical analysis in the presence of competing risks is a modeling concept that aims to account for situations where the risks are latent, in the sense that there is no information about which component was responsible for the object failure. Such issue/problem arises naturally in several areas, like public health, actuarial science, biomedical studies, demography and industrial reliability. In the classical latent competing risks scenarios, the lifetime associated with a particular risk is not observable, but rather we observe only the maximum lifetime value among all risks. For instance, in reliability, we observe only the maximum component lifetime of a parallel system.

Let T_i $(i = 1, 2, \ldots)$ denote the time-to-event due to the j-th competing risk and N be a random variable with a zero-truncated Poisson (ZTP) distribution indexed by a parameter λ, hereafter ZTP(λ), given by

$$P(N = n) = \frac{\lambda^n}{n!(e^{\lambda} - 1)}, \quad \text{for } n \in \mathbb{N}^* \text{ and } \lambda > 0.$$

Now, let $X = \min\{T_i\}_{i=1}^{N}$, where the T_i's are independent of N and assumed to be independent and identically distributed according to an exponential distribution with parameter β. The conditional cumulative distribution function (c.d.f.) of X is given by

$$F(x|N) = 1 - P(X > x|N) = 1 - [P\left(T_1 > x\right)]^N = 1 - \left(e^{-\beta x}\right)^N.$$

Thus, the unconditional c.d.f. of X is

$$F(x|\lambda, \beta) = \sum_{n=1}^{\infty} \frac{\lambda^n}{n!(e^{\lambda} - 1)} - \sum_{n=1}^{\infty} \frac{\lambda^n \left(e^{-\beta x}\right)^n}{n!(e^{\lambda} - 1)} = \frac{e^{\lambda e^{-\beta x}} - e^{\lambda}}{1 - e^{\lambda}}. \tag{2}$$

If $X \sim$ EP(λ, β), then the mean and variance of X are given, respectively, by

$$E(X|\lambda, \beta) = \frac{\lambda}{\beta(e^{\lambda} - 1)} F_{2,2}([1, 1], [2, 2], \lambda) \quad \text{and} \tag{3}$$

$$Var(X|\lambda, \beta) = \frac{\lambda}{\beta^2(e^{\lambda} - 1)} \left[2F_{3,3}([1, 1, 1], [2, 2, 2], \lambda) - \frac{\lambda}{(e^{\lambda} - 1)} F_{2,2}^2([1, 1], [2, 2], \lambda)\right],$$

where $F_{p,q}(\boldsymbol{a}, \boldsymbol{b}, \lambda)$ is the generalized hypergeometric function, with $\boldsymbol{a} = [a_1, a_2, \ldots, a_p]$, p is the number of operands of $\boldsymbol{a}$, $\boldsymbol{b} = [b_1, b_2, \ldots, b_q]$ and q is the number of operands of $\boldsymbol{b}$.

For $r \in \mathbb{N}$, the raw moments of X about the origin are

$$E(X^r|\lambda,\beta) = \frac{\lambda\Gamma(r+1)}{\beta^r(e^\lambda - 1)} F_{r+1,r+1}([1,1,\ldots,1],[2,2,\ldots,2],\lambda),$$

where $\Gamma(\cdot)$ is the gamma function.

The survival and hazard functions of the EP distribution are given, respectively, by

$$\bar{F}(x|\lambda,\beta) = \frac{1-e^{\lambda e^{-\beta x}}}{1-e^\lambda} \quad \text{and} \quad h(x|\lambda,\beta) = \frac{\lambda\beta e^{-\lambda-\beta x+\lambda e^{-\beta x}}(1-e^\lambda)}{(1-e^{\lambda e^{-\beta x}})(1-e^{-\lambda})}.$$

The hazard function is decreasing for all $\lambda > 0$ and $\beta > 0$. Figure 1 presents the density and hazard functions of the EP distribution considering different values of λ and β.

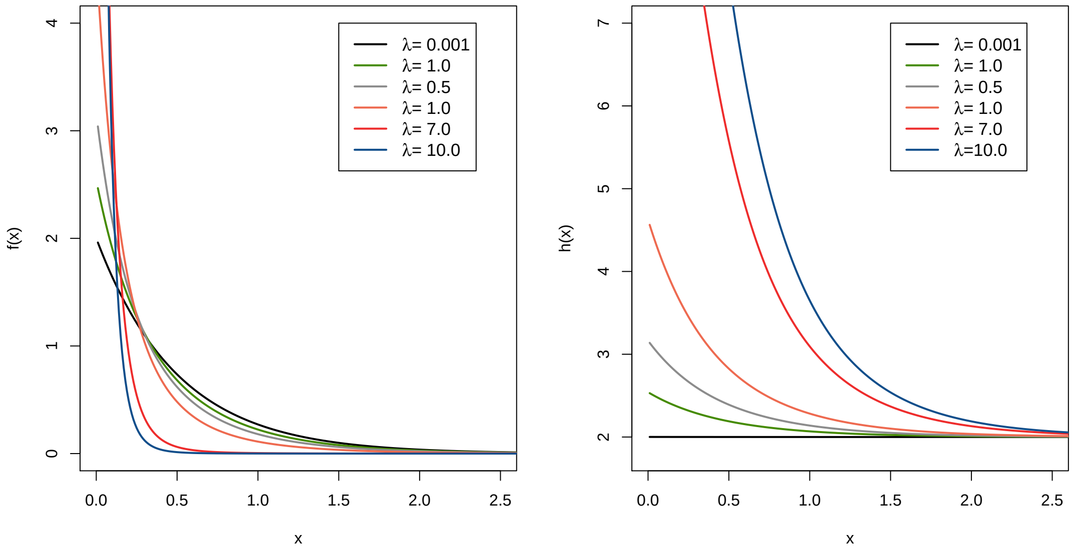

Figure 1. Left panel: probability density function of the EP distribution. Right panel: hazard function of the EP distribution.

3. Classical Inference

In this section we discuss different frequentist estimation methods for the parameters λ and β of the EP distribution.

3.1. *Maximum Likelihood Estimators*

The maximum likelihood estimation method is a standard procedure in statistical inference. It has many desirable properties, including consistency, asymptotic efficiency and invariance.

Let $X_1, \cdots, X_n$ be a random sample of size n from EP(λ,β). Then, the likelihood function of (1) is given by

$$L\left(\lambda,\beta|\boldsymbol{x}\right)=\left(\frac{\lambda\beta}{1-e^{-\lambda}}\right)^{n}\exp\left\{-n\lambda-\beta\sum_{i=1}^{n}x_{i}+\lambda\sum_{i=1}^{n}e^{-\beta x_{i}}\right\}. \tag{4}$$

The log-likelihood function of (4) is given by

$$\ell\left(\lambda,\beta|\boldsymbol{x}\right)=n\log\left(\lambda\beta\right)-n\log\left(1-e^{-\lambda}\right)-n\lambda-\beta\sum_{i=1}^{n}x_{i}+\lambda\sum_{i=1}^{n}e^{-\beta x_{i}}. \tag{5}$$

The maximum likelihood estimators (MLEs) of λ and β, i.e. $\hat{\lambda}_{\text{MLE}}$ and $\hat{\beta}_{\text{MLE}}$, are obtained by maximizing the log-likelihood function (5) or solving the following two nonlinear equations:

$$\frac{\partial\ell\left(\lambda,\beta|\boldsymbol{x}\right)}{\partial\lambda}=\frac{n}{\lambda}+\frac{n}{1-e^{\lambda}}-n+\sum_{i=1}^{n}e^{-\beta x_{i}}=0, \tag{6}$$

$$\frac{\partial\ell\left(\lambda,\beta|\boldsymbol{x}\right)}{\partial\beta}=\frac{n}{\beta}-\sum_{i=1}^{n}x_{i}-\lambda\sum_{i=1}^{n}x_{i}e^{-\beta x_{i}}=0. \tag{7}$$

The solution of (7) is given by

$$\hat{\lambda}_{\text{MLE}}=\left(\frac{1}{\hat{\beta}_{\text{MLE}}}-\bar{x}\right)\left(\frac{1}{n}\sum_{i=1}^{n}x_{i}e^{-\hat{\beta}_{\text{MLE}}x_{i}}\right)^{-1}, \tag{8}$$

where $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_{i}$ is the sample mean. However, to obtain the solution of (6), we have to consider numerical methods, such as Newton-Raphson (NR). The existence and uniqueness of the MLE have been discussed by Kus (2007) under some conditions of the parameters. E.g., for $\lambda\in(0,e^{2})$, the root $\hat{\beta}_{\text{MLE}}$ of (7) is unique and lies in the interval $\left([\bar{x}(\lambda+1)]^{-1},\bar{x}^{-1}\right)$.

The MLEs are asymptotically normally distributed with a joint bivariate normal distribution given by

$$\left(\hat{\lambda}_{\text{MLE}},\hat{\beta}_{\text{MLE}}\right)\sim N_{2}\left[\left(\lambda,\beta\right),I^{-1}\left(\lambda,\beta\right)\right]\qquad\text{as}\qquad n\to\infty,$$

where $I(\lambda,\beta)$ is the Fisher information matrix given by

$$I(\lambda,\beta)=\begin{bmatrix}\dfrac{n\left[1+e^{2\lambda}-\lambda^{2}e^{\lambda}-2e^{\lambda}\right]}{\lambda^{2}(1-e^{-\lambda})^{2}} & \dfrac{n\lambda e^{-\lambda}}{4\beta\left(1-e^{-\lambda}\right)}F_{2,2}\left([2,2],[3,3],\lambda\right)\\ \dfrac{n\lambda e^{-\lambda}}{4\beta\left(1-e^{-\lambda}\right)}F_{2,2}\left([2,2],[3,3],\lambda\right) & I_{22}\left(\lambda,\beta\right)\end{bmatrix}, \tag{9}$$

with $I_{22}\left(\lambda,\beta\right)=\dfrac{n}{\beta^{2}}-\dfrac{n\lambda^{2}e^{-\lambda}}{4\beta^{2}\left(1-e^{-\lambda}\right)}F_{3,3}\left([2,2,2],[3,3,3],\lambda\right)$.

3.2. *Moments Estimators*

Introduced by Karl Pearson in 1894, the method of moments is one of the oldest procedures used for estimating parameters in statistical models. Considering that

$$E\left(X|\lambda,\beta\right)=\frac{\lambda F_{2,2}\left([1,1],[2,2],\lambda\right)}{\beta\left(e^{\lambda}-1\right)}$$

and

$$Var\left(X|\lambda,\beta\right)=\frac{\lambda\left[2(e^{\lambda}-1)F_{3,3}\left([1,1,1],[2,2,2],\lambda\right)-\lambda F_{2,2}^{2}\left([1,1],[2,2],\lambda\right)\right]}{\beta^{2}\left(e^{\lambda}-1\right)^{2}},$$

the population coefficient of variation is given by

$$CV\left(X|\lambda,\beta\right)=\sqrt{\frac{2\left(e^{\lambda}-1\right)F_{3,3}\left([1,1,1],[2,2,2],\lambda\right)}{\lambda F_{2,2}^{2}\left([1,1],[2,2],\lambda\right)}-1}.$$

Note that CV does not depend on the scale parameter β.

The moments estimator (ME) for λ, i.e. $\hat{\lambda}_{\mathrm{ME}}$, can be obtained by solving the non-linear equation

$$\sqrt{\frac{2(e^{\lambda}-1)F_{3,3}([1,1,1],[2,2,2],\lambda)}{\lambda F_{2,2}^{2}([1,1],[2,2],\lambda)}-1}-\frac{S}{\bar{X}}=0,$$

where $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ and $S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}$ are the sample mean and standard deviation, respectively.

Further, the ME of β, i.e. $\hat{\beta}_{\mathrm{ME}}$, can be obtained as

$$\hat{\beta}_{\mathrm{ME}}=\frac{\hat{\lambda}_{\mathrm{ME}}}{\bar{X}(e^{\hat{\lambda}_{\mathrm{ME}}}-1)}F_{2,2}\left([1,1],[2,2],\hat{\lambda}_{\mathrm{ME}}\right).$$

3.3. *Least-Squares Estimators*

The ordinary least-squares estimators (OLSEs) and weighted least-squares estimators (WLSEs) were originally proposed by Swain et al. (1988) to estimate the parameters of a Beta distribution.

Let $X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ be the order statistics of a random sample of size n from EP(λ,β). The OLSEs of λ and β, denoted by $\hat{\lambda}_{\mathrm{OLSE}}$ and $\hat{\beta}_{\mathrm{OLSE}}$, respectively, are obtained by minimizing

$$S\left(\lambda,\beta\right)=\sum_{i=1}^{n}\left[F\left(x_{(i)}|\lambda,\beta\right)-\frac{i}{n+1}\right]^{2}$$

with respect to λ and β, where $F\left(\cdot|\lambda,\beta\right)$ is given by (2). Equivalently, they can be obtained by solving

$$\sum_{i=1}^{n}\left[F\left(x_{(i)}|\lambda,\beta\right)-\frac{i}{n+1}\right]\Psi_1\left(x_{(i)}|\lambda,\beta\right)=0,$$
$$\sum_{i=1}^{n}\left[F\left(x_{(i)}|\lambda,\beta\right)-\frac{i}{n+1}\right]\Psi_2\left(x_{(i)}|\lambda,\beta\right)=0,$$

where

$$\Psi_1\left(x_{(i)}|\lambda,\beta\right)=\frac{e^{\lambda}\left(-e^{\lambda}+e^{\lambda e^{-\beta x_{(i)}}}\right)+\left(1-e^{\lambda}\right)\left(-e^{\lambda}+e^{-\beta x_{(i)}+\lambda e^{-\beta x_{(i)}}}\right)}{\left(e^{\lambda}-1\right)^2} \tag{10}$$

and

$$\Psi_2\left(x_{(i)}|\lambda,\beta\right)=\frac{x_{(i)}\lambda e^{-\beta x_{(i)}+\lambda e^{-\beta x_{(i)}}}}{e^{\lambda}-1}. \tag{11}$$

The WLSEs of λ and β, i.e. $\hat{\lambda}_{\text{WLSE}}$ and $\hat{\beta}_{\text{WLSE}}$, can be obtained by minimizing

$$W(\lambda,\beta)=\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F\left(x_{(i)}|\lambda,\beta\right)-\frac{i}{n+1}\right]^2.$$

These estimators can also be obtained by solving

$$\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F\left(x_{(i)}|\lambda,\beta\right)-\frac{i}{n+1}\right]\Psi_1\left(x_{(i)}|\lambda,\beta\right)=0,$$
$$\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F\left(x_{(i)}|\lambda,\beta\right)-\frac{i}{n+1}\right]\Psi_2\left(x_{(i)}|\lambda,\beta\right)=0.$$

3.4. *Percentile Estimators*

Kao (1958; 1959) proposed a statistical method to estimate the parameters of probability distributions by comparing the sample points with the theoretical ones. This method has been used in distributions that have the quantile function in a closed-form expression, such as the Weibull and generalized exponential distributions.

For $p\in(0,1)$, the quantile function of the EP distribution is given by

$$X_p=Q(p|\lambda,\beta)=-\frac{\log\left(\lambda^{-1}\log\left(p(1-e^{\lambda})+e^{\lambda}\right)\right)}{\beta}.$$

Hence, the percentile estimators (PEs) for λ and β, say $\hat{\lambda}_{\text{PE}}$ and $\hat{\beta}_{\text{PE}}$, can be obtained by minimizing

$$\sum_{i=1}^{n}\left(X_{(i)}+\frac{\log\left(\lambda^{-1}\log\left(\hat{p}_i\left(1-e^{\lambda}\right)+e^{\lambda}\right)\right)}{\beta}\right)^2$$

with respect to λ and β, where $\hat{p}_i$ denotes some estimator of $F(x_{(i)}|\lambda,\beta)$. The PEs can also be obtained by solving the following nonlinear equations:

$$\frac{1}{\beta}\sum_{i=1}^{n}\left[X_{(i)}+\frac{\log\left(\lambda^{-1}\log\left(\hat{p}_i\left(1-e^{\lambda}\right)+e^{\lambda}\right)\right)}{\beta}\right]\frac{e^{\lambda}(\hat{p}_i-1)}{\lambda\left(e^{\lambda}\left(\hat{p}_i-1\right)-\hat{p}_i\right)\log\left(\hat{p}_i\left(1-e^{\lambda}\right)+e^{\lambda}\right)}=0,$$

$$\frac{1}{\beta^2}\sum_{i=1}^{n}\left[X_{(i)}+\frac{\log\left(\lambda^{-1}\log\left(\hat{p}_i\left(1-e^{\lambda}\right)+e^{\lambda}\right)\right)}{\beta}\right]\log\left(\lambda^{-1}\log\left(\hat{p}_i\left(1-e^{\lambda}\right)+e^{\lambda}\right)\right)=0.$$

In this paper, we consider $\hat{p}_i=\frac{i}{n+1}$, but several other estimators of $F(x_{(i)}|\lambda,\beta)$ could be used instead; see, e.g., Mann et al. (1974).

3.5. *Maximum Product of Spacings Estimators*

Cheng and Amin (1983) introduced a powerful alternative to maximum likelihood method for estimating the unknown parameters of continuous univariate distributions, named maximum product of spacings method. Such method can also be obtained as an approximation to the Kullback-Leibler measure of information; see Ranneby (1984). Let

$$D_i(\lambda,\beta)=F\left(x_{(i)}|\lambda,\beta\right)-F\left(x_{(i-1)}|\lambda,\beta\right),\quad i=1,2,\ldots,n+1,$$

be the uniform spacings of a random sample from the EP distribution, with $F(x_{(0)}|\lambda,\beta)=0$ and $F(x_{(n+1)}|\lambda,\beta)=1$. Thus, $\sum_{i=1}^{n+1}D_i(\lambda,\beta)=1$.

The maximum product of spacings estimators (MPSEs) of the parameters λ and β, denoted by $\hat{\lambda}_{\text{MPSE}}$ and $\hat{\beta}_{\text{MPSE}}$, respectively, are obtained by maximizing the geometric mean of the spacings

$$G(\lambda,\beta)=\left[\prod_{i=1}^{n+1}D_i(\lambda,\beta)\right]^{\frac{1}{n+1}}$$

with respect to λ and β, or equivalently, by maximizing the function

$$H(\lambda,\beta)=\frac{1}{n+1}\sum_{i=1}^{n+1}\log D_i(\lambda,\beta).$$

The MPSEs can also be obtained by solving the nonlinear equations:

$$\frac{\partial H(\lambda,\beta)}{\partial\lambda}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(\lambda,\beta)}\left[\Psi_1(x_{(i)}|\lambda,\beta)-\Psi_1(x_{(i-1)}|\lambda,\beta)\right]=0,$$

$$\frac{\partial H(\lambda,\beta)}{\partial\beta}=\frac{1}{n+1}\sum_{i=1}^{n+1}\frac{1}{D_i(\lambda,\beta)}\left[\Psi_2(x_{(i)}|\lambda,\beta)-\Psi_2(x_{(i-1)}|\lambda,\beta)\right]=0,$$

where $\Psi_1(\cdot|\lambda,\beta)$ and $\Psi_2(\cdot|\lambda,\beta)$ are given by (10) and (11), respectively.

Observe that for $x_{(i+k)} = x_{(i+k-1)} = \cdots = x_{(i)}$, we have $D_{i+k}(\lambda, \beta) = D_{i+k-1}(\lambda, \beta) = \cdots = D_i(\lambda, \beta) = 0$. Hence, the MPSEs are sensitive to closely-spaced observations, especially ties (i.e. multiple observations with the same value). In this case (ties occurrence), $D_i(\lambda, \beta)$ should be replaced by the corresponding likelihood $L\left(\lambda, \beta | x_{(i)}\right) = f(x_{(i)}|\lambda, \beta)$, since $x_{(i)} = x_{(i-1)}$.

Cheng and Amin (1983) derived some desirable properties of the MPSEs, such as asymptotic efficiency and invariance. They also showed that the consistency of MPSEs holds under much more general conditions than for MLEs. Thus, under mild conditions, the MPSEs are asymptotically normally distributed with a joint bivariate normal distribution given by

$$\left(\hat{\lambda}_{\text{MPSE}}, \hat{\beta}_{\text{MPSE}}\right) \sim N_2\left[(\lambda, \beta), I^{-1}(\lambda, \beta)\right] \quad \text{as} \quad n \to \infty,$$

where $I^{-1}(\lambda, \beta)$ is the inverse of the Fisher information matrix given in (9).

3.6. *Minimum Distance Estimators*

In this subsection we present three minimum distance estimators (also called maximum goodness-of-fit estimators) for λ and β. This class of estimators is based on the difference between the estimate of the c.d.f. and the empirical distribution function; see Luceño (2006).

3.6.1. Cramér-von Mises Estimators

The Cramér-von Mises estimators (CMEs) for the parameters λ and β, i.e. $\hat{\lambda}_{\text{CME}}$ and $\hat{\beta}_{\text{CME}}$, are obtained by minimizing, with respect to λ and β, the function

$$C(\lambda, \beta) = \frac{1}{12n} + \sum_{i=1}^{n}\left(F\left(x_{(i)}|\lambda, \beta\right) - \frac{2i-1}{2n}\right)^2.$$

These estimators can also be obtained by solving the nonlinear equations:

$$\begin{aligned}
\sum_{i=1}^{n}\left(F\left(x_{(i)}|\lambda, \beta\right) - \frac{2i-1}{2n}\right)\Psi_1\left(x_{(i)}|\lambda, \beta\right) &= 0, \\
\sum_{i=1}^{n}\left(F\left(x_{(i)}|\lambda, \beta\right) - \frac{2i-1}{2n}\right)\Psi_2\left(x_{(i)}|\lambda, \beta\right) &= 0,
\end{aligned}$$

where $\Psi_1(\cdot|\lambda, \beta)$ and $\Psi_2(\cdot|\lambda, \beta)$ are given by (10) and (11), respectively.

3.6.2. Anderson-Darling and Right-tail Anderson-Darling Estimators

The Anderson-Darling estimators (ADEs) for λ and β, i.e. $\hat{\lambda}_{\text{ADE}}$ and $\hat{\beta}_{\text{ADE}}$, are obtained by minimizing, with respect to λ and β, the function

$$A(\lambda, \beta) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left\{\log F\left(x_{(i)}|\lambda, \beta\right) + \log \bar{F}\left(x_{(n+1-i)}|\lambda, \beta\right)\right\}.$$

These estimators can also be obtained by solving the nonlinear equations:

$$\begin{aligned}
\sum_{i=1}^{n}(2i-1)\left[\frac{\Psi_1\left(x_{(i)}|\lambda,\beta\right)}{F\left(x_{(i)}|\lambda,\beta\right)}-\frac{\Psi_1\left(x_{(n+1-i)}|\lambda,\beta\right)}{\bar{F}\left(x_{(n+1-i)}|\lambda,\beta\right)}\right] &= 0,\\
\sum_{i=1}^{n}(2i-1)\left[\frac{\Psi_2\left(x_{(i)}|\lambda,\beta\right)}{F\left(x_{(i)}|\lambda,\beta\right)}-\frac{\Psi_2\left(x_{(n+1-i)}|\lambda,\beta\right)}{\bar{F}\left(x_{(n+1-i)}|\lambda,\beta\right)}\right] &= 0,
\end{aligned}$$

where $\Psi_1(\cdot|\lambda,\beta)$ and $\Psi_2(\cdot|\lambda,\beta)$ are given by (10) and (11), respectively.

On the other hand, the right-tail Anderson-Darling estimators (RTADEs) for λ and β (i.e. $\hat{\lambda}_{\mathrm{RTADE}}$ and $\hat{\beta}_{\mathrm{RTADE}}$) are obtained by minimizing

$$R(\lambda,\beta)=\frac{n}{2}-2\sum_{i=1}^{n}F\left(x_{(i)}|\lambda,\beta\right)-\frac{1}{n}\sum_{i=1}^{n}(2i-1)\log\bar{F}\left(x_{(n+1-i)}|\lambda,\beta\right)$$

with respect to λ and β. These estimators can also be obtained by solving the following nonlinear equations:

$$\begin{aligned}
-2\sum_{i=1}^{n}\Psi_1\left(x_{(i)}|\lambda,\beta\right)+\frac{1}{n}\sum_{i=1}^{n}(2i-1)\frac{\Psi_1\left(x_{(n+1-i)}|\lambda,\beta\right)}{\bar{F}\left(x_{(n+1-i)}|\lambda,\beta\right)} &= 0,\\
-2\sum_{i=1}^{n}\Psi_2\left(x_{(i)}|\lambda,\beta\right)+\frac{1}{n}\sum_{i=1}^{n}(2i-1)\frac{\Psi_2\left(x_{(n+1-i)}|\lambda,\beta\right)}{\bar{F}\left(x_{(n+1-i)}|\lambda,\beta\right)} &= 0,
\end{aligned}$$

where $\Psi_1(\cdot|\lambda,\beta)$ and $\Psi_2(\cdot|\lambda,\beta)$ are given by (10) and (11), respectively.

3.7. *Nested EM Algorithm with Correction Approach*

The EM algorithm has been widely used to obtain the MLEs for many distributions. For the EP distribution, Kus (2007) presented such approach to finding the desirable estimates. Although it is useful for finding the MLEs, the EM algorithm still uses an NR step during the M-step of the algorithm, which is undesirable.

Karlis (2009) overcomes this problem by proposing a nested EM algorithm, which does not depend on numerical methods to achieve the parameter estimates of the EP distribution. The main idea is to consider another EM algorithm to solve the M-step of the first EM algorithm. In this case, the second EM algorithm is used to estimate the expected number of zero observations n_0. Although in this procedure we have to start the algorithm with three initial values λ,β and n_0, such values will always lead to the estimates of the parameters. Additionally, in Section 5 we will present useful initial values to decrease the number of steps up to the convergence of the proposed algorithm. By setting the initial values for the parameters, the nested EM algorithm is given as follows.

(1) Compute $w_i=1+\lambda e^{-\beta x_i}$ (E-step);

(2) Compute $n_0^{new}=(n+n_0)e^{-\lambda}$ (E-step);

(3) Update $\beta^{new}=\dfrac{1}{\sum_{i=1}^{n}w_i x_i}$ (M-step);

(4) Update $\lambda^{new} = \dfrac{\sum_{i=1}^{n} w_i}{n + n_0^{new}}$ (M-step);

(5) Repeat the steps (1) - (4) m times (e.g., $m = 500$).

The most interesting aspect of this algorithm is that its computational cost is very small in comparison with the other procedures. However, the MLEs that are achieved by the nested EM algorithm have a significant bias, especially for small and moderate sample sizes. This problem can be overcome by considering a bias correction approach using bootstrap (see Ramos et al., 2018). The bootstrap bias-corrected MLEs using the nested EM algorithm (BNEM) is given by subtraction of the estimated bias from the original MLEs obtained through the procedure above.

In order to perform the bootstrap, let $\mathbf{x} = (x_1, \ldots, x_n)$ be a sample with n observations randomly selected from the random variable X with the distribution function given in (2). The pseudo-samples $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ are obtained from the original sample $\mathbf{x}$ through resampling with replacement. If we have B bootstrap samples $(\mathbf{x}^{*(1)}, \mathbf{x}^{*(2)}, \ldots, \mathbf{x}^{*(B)})$ that are generated independently from the original sample $\mathbf{x}$ and their respective bootstrap estimates $(\hat{\theta}^{*(1)}, \hat{\theta}^{*(2)}, \ldots, \hat{\theta}^{*(B)})$ are calculated using the nested EM algorithm, then the bootstrap expectations are approximated by

$$\hat{\theta}^{*(\cdot)} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*(b)}. \tag{12}$$

From (12), the bootstrap bias estimate is given by $\hat{B}_F(\hat{\theta}, \theta) = \hat{\theta}^{*(\cdot)} - \hat{\theta}$, where $\hat{\theta}$ is the MLE obtained from the nested EM algorithm. The bias-corrected estimator obtained through the bootstrap resampling method is given by

$$\hat{\theta}^{B} = \hat{\theta} - \hat{B}_F(\hat{\theta}, \theta) = 2\hat{\theta} - \hat{\theta}^{*(\cdot)}.$$

Here, we have $\hat{\theta}^B$ denoted by $\hat{\boldsymbol{\theta}}_{\text{BNEM}} = (\hat{\lambda}_{\text{BNEM}}, \hat{\beta}_{\text{BNEM}})$.

4. Numerical Analysis

A Monte Carlo simulation study was carried out to compare the efficiency of the different frequentist estimation methods for the parameters of the EP distribution. The following approach was adopted.

(1) Generate N samples of size n from the EP(λ, β) distribution;
(2) For each generated sample, obtain the estimates of λ and β, i.e. $\hat{\lambda}$ and $\hat{\beta}$, via the PEs, MLEs, MEs, OLSEs, WLSEs, BNEM, MPSEs, CMEs, ADEs and RTADEs;
(3) Considering $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2\right) = \left(\hat{\lambda}, \hat{\beta}\right)$ and $\boldsymbol{\theta} = (\theta_1, \theta_2) = (\lambda, \beta)$, compute the Bias and Mean Squared Error (MSE) of $\hat{\theta}_j$, for $j = 1, 2$, which are given, respectively, by $\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\theta}_j^{(k)} - \theta_j\right)$ and $\frac{1}{N}\sum_{k=1}^{N}\left(\hat{\theta}_j^{(k)} - \theta_j\right)^2$. Here, $\hat{\theta}_j^{(k)}$ denotes the estimate of θ_j obtained from sample k, for $k = 1, 2, \cdots, N$.

By this approach, it is expected that the most efficient estimation method will have both Bias and MSE closer to zero.

Our simulations were performed using the R software (2014). The maximization method used is the NR (Henningsen and Toomet, 2011). The pseudo-random samples were generated using the seed 2017. The chosen values of the simulation parameters were: $N = 10,000$ and $n = \{20, 25, 30, 35, \cdots, 160\}$. Due to lack of space, we will present the results only for $\boldsymbol{\theta} = \{(1.5, 0.02), (2, 0.5)\}$. However, the following results are similar for other choices of $\boldsymbol{\theta}$.

Figure 2 presents the proportion of error related to each estimation procedure, i.e. the frequency of times that each estimation method did not converge to an estimate of the parameters. We observe from this figure that the PEs, OLSEs, WLSEs, MEs and CMEs failed in finding the numerical solution for a significant number of samples (convergence errors). It is worth mentioning that the initial values used were the true values. However, in real applications, such values are hardly known. Therefore, we considered the same analysis, but assuming that initial values came from an uniform distribution on the interval $(0, 4)$. Figure 3 shows the proportion of error related to the estimation procedures using random initial guess. Notice that, in general, the methods that used numerical techniques were not very efficient in finding the estimates, while the method using the nested EM algorithm was able to find the solution in all cases, i.e. for all sample sizes. Since the OLSEs, WLSEs, MEs, CMEs and the PEs returned a high percentage of errors, these methods were removed in order to avoid the inclusion of bias in the simulation study. Hereafter, we consider the MLEs, MPSEs, ADEs, RTADEs and the BNEM.

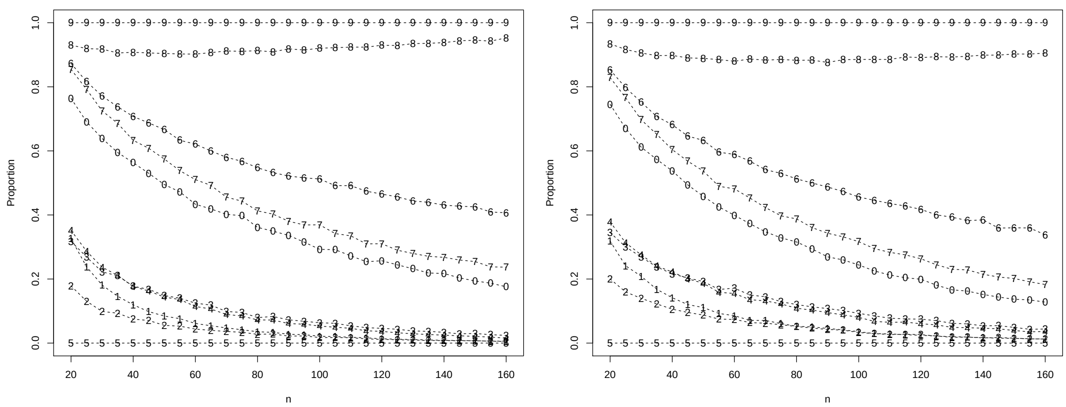

Figure 2. Proportion of failures/errors considering $\boldsymbol{\theta} = (1.5, 0.02)$ (left panel) and $\boldsymbol{\theta} = (2, 0.5)$ (right panel) under different sample sizes n and the following estimation procedures: 0 - PEs, 1 - MLEs, 2 - MPSEs, 3 - ADEs, 4 - RTADEs, 5 - BNEM, 6 - OLSEs, 7 - WLSEs, 8 - MEs, 9 - CMEs ($N = 10,000$).

Figures 4 and 5 show the Bias and MSE of the observed estimates of λ and β obtained using the MLEs, MPSEs, ADEs, RTADEs and the BNEM, with $N = 10,000$ simulated samples, and different values of n and $\boldsymbol{\theta}$: $\boldsymbol{\theta} = (1.5, 0.02)$ (Figure 4) and $\boldsymbol{\theta} = (2, 0.5)$ (Figure 5). From these figures, we can see that both Bias and MSE of all estimators decrease as n increases, i.e. the estimators are asymptotically unbiased and consistent. The bias-corrected MLEs obtained using the nested EM algorithm returned better results when compared with the other estimation procedures. The advantage of this approach is that the asymptotic properties of the MLEs can be used in the BNEM. Therefore, we can easily build confidence intervals for the proposed estimates. In summary, we recommend the use of the BNEM for all practical purposes.

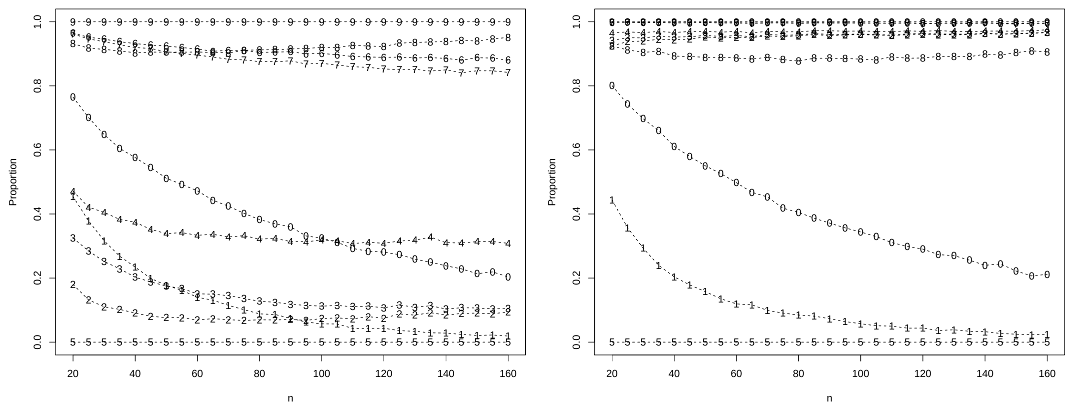

Figure 3. Proportion of failures/errors considering $\boldsymbol{\theta} = (1.5, 0.02)$ (left panel) and $\boldsymbol{\theta} = (2, 0.5)$ (right panel) under different sample sizes n and using random initial values in the following estimation procedures: 0 - PEs, 1 - MLEs, 2 - MPSEs, 3 - ADEs, 4 - RTADEs, 5 - BNEM, 6 - OLSEs, 7 - WLSEs, 8 - MEs, 9 - CMEs ($N = 10,000$).

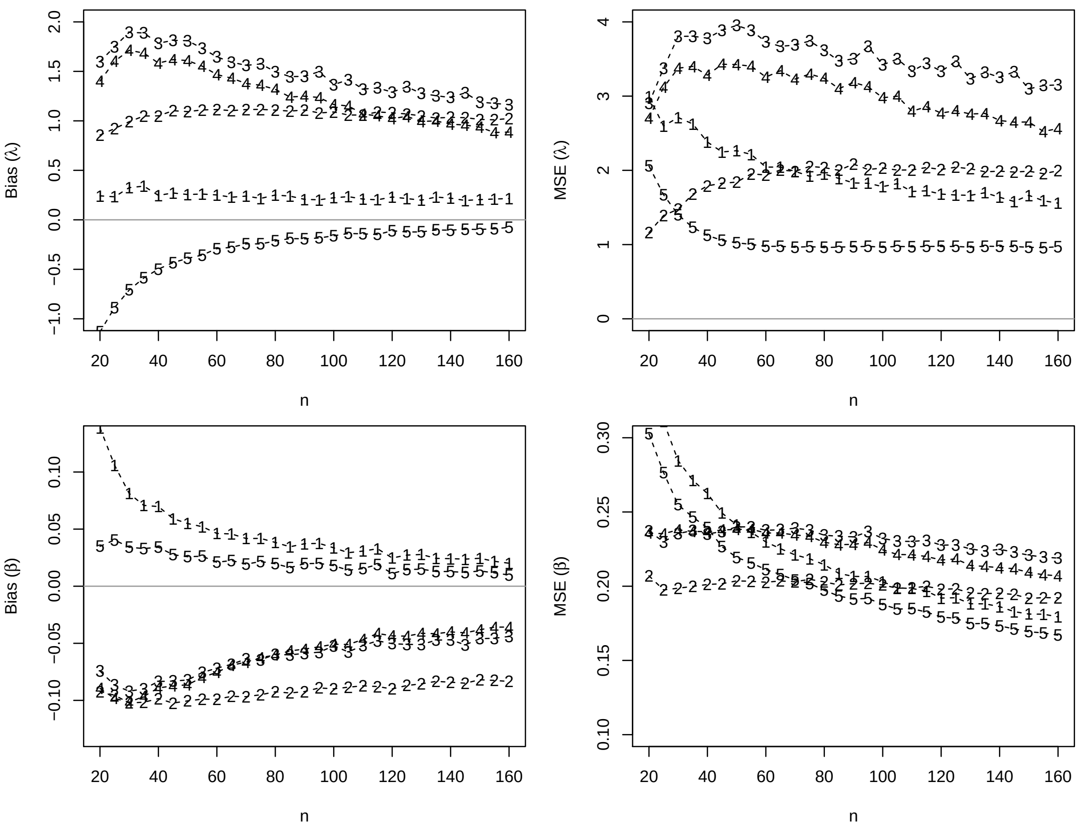

Figure 4. Bias and MSE of different estimators of $\lambda = 1.5$ and $\beta = 0.02$, for each sample size n and using the following methods: 1 - MLEs, 2 - MPSEs, 3 - ADEs, 4 - RTADEs, 5 - BNEM ($N = 10,000$).

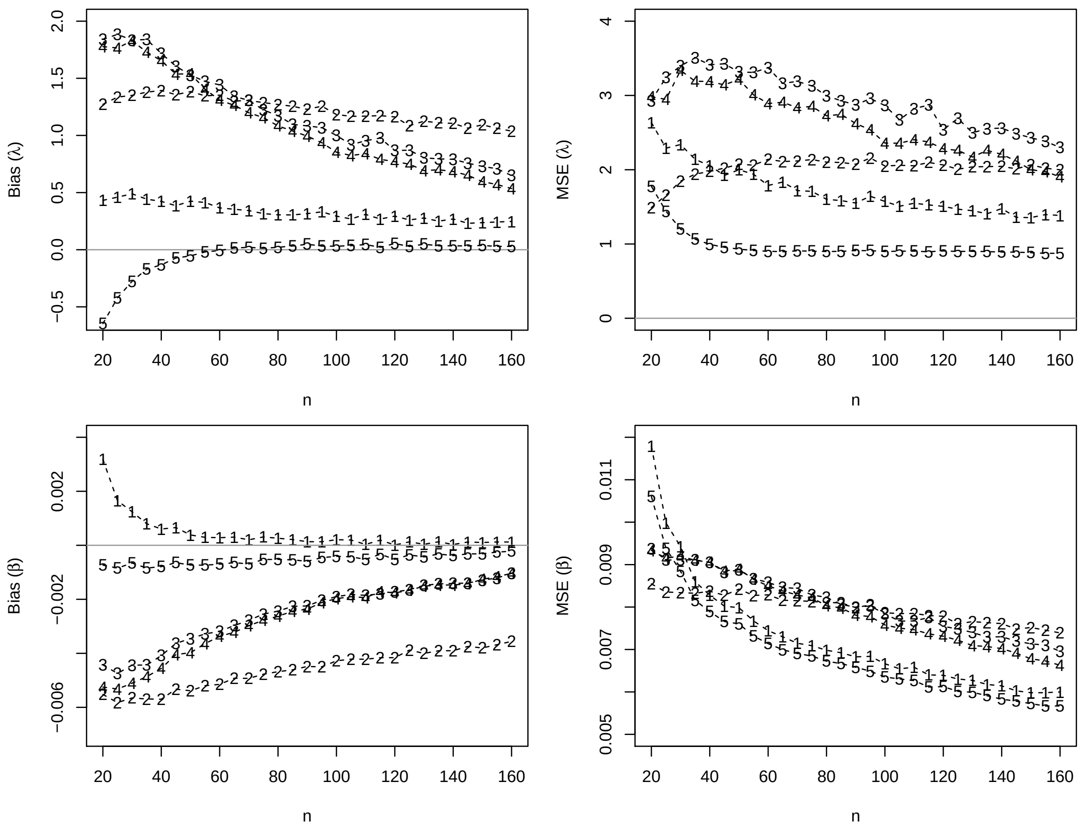

Figure 5. Bias and MSE of different estimators of $\lambda = 2$ and $\beta = 0.5$, for each sample size n and using the following methods: 1 - MLEs, 2 - MPSEs, 3 - ADEs, 4 - RTADEs, 5 - BNEM ($N = 10,000$).

5. Applications

In this section, we illustrate the proposed methodology on rainfall data (Section 5.1), as well as on aircraft data (Section 5.2), both with zero occurrence.

An important discussion in real applications, specially using iterative methods, is the selection of good initial values to start the algorithm. A common approach is to consider some estimator that has closed-form expression. However, for the EP distribution, we were not able to find any. In order to overcome this problem, we first consider the MLEs under the assumption that $\lambda \to 0^+$; in this case, we can compute the estimate of β by the approximation of the exponential distribution. But the ratio $1/\bar{x}$ tends to overestimate β and will return $\lambda = 0$; in this case, we consider an approximation to $1/(1.5\bar{x})$. Substituting this result in (8), we can compute the initial value of λ after some algebraic manipulations, i.e.

$$\beta_0 = \frac{1}{1.5\bar{x}} \quad \text{and} \quad \lambda_0 = \frac{\bar{x}}{2}\left(\frac{1}{n}\sum_{i=1}^{n} x_i \exp\left\{-\frac{x_i}{1.5\bar{x}}\right\}\right)^{-1}, \tag{13}$$

in which β_0 and λ_0 should only be used as initial values to increase the convergence speed of the nested EM algorithm. In the case of n_0, a good initial value is the sample size n.

The results obtained using the EP distribution are now compared to the corresponding ones achieved with the usage of the Poisson-exponential (PE) distribution by Cancho et al. (2011). It is important to point out that the results considering the Weibull, gamma and lognormal distributions will not be presented, since it is not possible to compute the estimates of these distributions' parameters in the presence of zero values.

The goodness-of-fit of the models are checked using the Kolmogorov-Smirnov (KS) test, which is based on the KS statistic $D_n = \sup_x |F_n(x) - F(x|\boldsymbol{\theta})|$, where $\sup_x$ is the supremum of the set of distances, $F_n(x)$ is the empirical c.d.f. and $F(x|\boldsymbol{\theta})$ is the theoretical c.d.f. In this case, we test if the data come from $F(x|\boldsymbol{\theta})$ (null hypothesis) and, with significance level of 5%, such hypothesis is rejected if the p-value is smaller than 0.05. Note that the KS test should only be used to verify the goodness-of-fit, and not as a discrimination criterion. Therefore, different discrimination criteria were considered, such as the AIC (Akaike Information Criterion), AICc (Corrected Akaike Information Criterion), HQIC (Hannan-Quinn Information Criterion) and CAIC (Consistent Akaike Information Criterion), which are computed, respectively, by $\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}|x) + 2c$, $\text{AICc} = \text{AIC} + \frac{2c(c+1)}{(n-c-1)}$, $\text{HQIC} = -2\ell(\hat{\boldsymbol{\theta}}|x) + 2c\log(\log(n))$ and $\text{CAIC} = -2\ell(\hat{\boldsymbol{\theta}}|x) + c\log(n) + 1$, where c is the number of parameters to be fitted and $\hat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$. Given a set of candidate models for the data at hand, the preferred model is the one that provides the minimum values.

5.1. *Environmental Data*

The study of precipitation, as well as the development of quantitative hydrological models in general, can be useful to support the adoption of new water resource management policies and assess water quality issues (Abbaspour et al. 2015; Beven 2011). However, as pointed out by Koutsoyiannis and Langousis (2011), the classical statistical models and methods may not be appropriate for precipitation, which exhibits some particular behaviours.

In this section, we consider a real data set (see Table 1) related to the total monthly precipitation (in millimeters, mm) during June in São Carlos city, located in the southeast region of Brazil, with 243,765 inhabitants. Including a period from 1960-2016, the data set was obtained from the Department of Water Resources and Power agency manager of water resources of the State of São Paulo.

Table 1. Data set related to total monthly precipitation (mm) during June (1960-2016) in São Carlos, Brazil.

45.6	0.0	82.0	0.0	18.9	71.2	0.0	201.8	0.0	35.5
55.7	113.1	2.5	18.9	123.5	0.7	62.6	41.7	29.6	0.0
90.1	78.4	42.7	82.8	0.0	6.0	0.1	21.9	36.5	22.0
4.2	4.0	0.9	62.2	25.5	18.3	18.0	4.6	33.4	5.1
6.8	0.0	7.8	37.8	7.4	13.7	3.1	20.2	36.2	24.8
28.7	126.6	61.9	5.5	23.1	109.8				

Figure 6 shows the fitted survival functions superimposed to the empirical survival function (Kaplan-Meier estimate), from which it can be observed that the EP distribution gives a better fit to the rainfall data.

The initial values obtained from (13) are: $\beta_0 = 0.0189$, $\lambda_0 = 1.6293$ and $n_0 = 56$. Table 2 displays the BNEM, standard errors (SE) and 95% bootstrap confidence intervals (CI) for λ and β. The computational time required to obtain the BNEM estimates is 3.040 seconds.

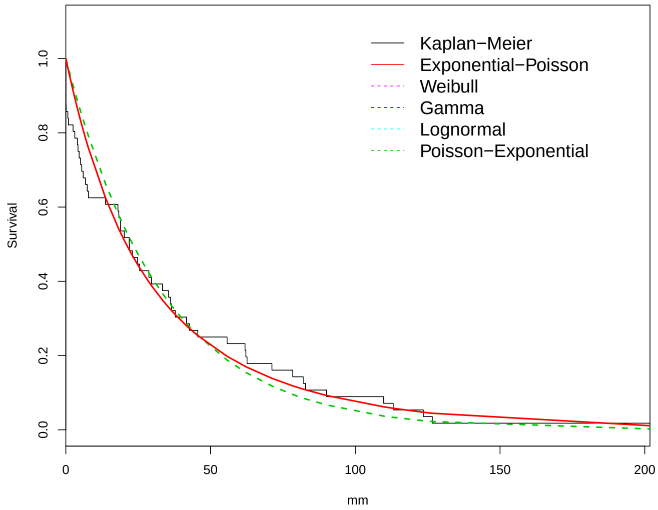

Figure 6. Fitted survival functions superimposed to the empirical survival function, considering the data set related to monthly total precipitation during June (1960-2016) in São Carlos, Brazil.

Table 2. MLEs using the nested EM algorithm, SE and 95% CI for the parameters of the EP distribution, considering the proposed data set.

Parameter	Estimate	SE	95% CI
λ	1.7202	0.0896	(1.5445 ; 1.8959)
β	0.0177	0.0122	(0.0028 ; 0.0232)

Table 3 presents the results from different model discrimination/selection criteria, such as the AIC, AICc, HQIC and CAIC, as well as the KS test, for the two considered probability distributions. From Table 3, we observe that the PE distribution has the KS test's p-value smaller than 0.10, so it is not a good candidate distribution for modeling the rainfall data.

Table 3. The AIC, AICc, HQIC and CAIC values, and the p-value from the KS goodness-of-fit test, for the fitted distributions, considering the data set related to total monthly precipitation during June (1960-2016) in São Carlos, Brazil.

	EP	PE
AIC	**511.6077**	514.9621
AICc	**511.8342**	515.1885
HQIC	**513.1782**	516.5325
CAIC	**517.6585**	521.0128
KS	0.1839	0.0613

According to the AIC, AICc, HQIC and CAIC values, the EP distribution provides a better fit to these data than the PE distribution, since the former has the minimum values in all criteria. Through the proposed methodology, the data related to the total monthly precipitation during June (1960-2016) in São Carlos, Brazil, can be well-described by the EP distribution considering the BNEM for λ and β. This distribution also allows us to obtain easily the Precipitation Return Level (PRL) from the closed-form expression

$$\hat{R}_p = -\frac{1}{\hat{\beta}} \log\left(\frac{1}{\hat{\lambda}} \log\left(1 - p(1 - e^{\hat{\lambda}})\right)\right),$$

where p is the return period. Table 4 shows the distribution return periods to predict the total monthly precipitation during June, 2017, in São Carlos, Brazil.

Table 4. PRL of the total monthly precipitation during June in São Carlos, Brazil, using the EP distribution.

1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
205.88	155.95	119.64	85.57	20.70	2.88	1.39	0.68	0.27

Using the expression (3) as predict value, we expect 35.27 mm of total monthly precipitation during June, 2017, in São Carlos city. Moreover, from Table 4 we may expect once every hundred years a total precipitation in this period of 205.88 mm.

The PRL of the total monthly precipitation can be used to plan maintenance policies in water streams. For instance, a recent flood in São Carlos led more than 100 local merchants to suffer great losses, as well as the destruction of public properties. Therefore, these results have important applications.

5.2. *Mechanical Components in an Aircraft*

In this section, we consider a real data set related to failure time of devices of an airline company. Such study is important in order to prevent customer dissatisfaction and customer attrition, and, consequently, to avoid customer loss. In this context, the choice of the distribution that better fits these data is fundamental for the company to reduce its costs. Table 5 presents the data related to failure time (in days) of 142 devices in an aircraft.

Table 5. Data set related to the failure time (in days) of 142 devices in an aircraft.

0	0	0	0	1	1	1	1	1	2	2	2	2	2
2	2	2	2	2	2	3	3	3	3	3	3	3	3
3	3	4	4	4	5	5	5	5	5	5	5	5	5
6	6	6	6	7	7	8	8	8	8	8	8	9	9
9	9	9	9	10	10	10	10	11	11	11	11	11	11
12	13	13	14	14	15	15	15	16	16	16	16	16	17
17	18	19	21	21	21	22	23	23	23	23	24	26	26
27	29	29	29	30	30	31	32	32	33	35	36	38	38
41	43	44	45	45	46	47	47	49	54	54	54	55	57
59	60	61	63	63	73	74	80	83	86	108	118	125	126
132	144												

The initial values obtained from (13) are: $\lambda_0 = 1.5824$, $\beta_0 = 0.0273$ and $n_0 = 142$. Table 6 displays the BNEM, SE and 95% CI for λ and β. The computational time required to obtain the BNEM estimates is 6.558 seconds. Table 7 presents the results of AIC, AICc, HQIC and CAIC criteria, as well as the KS test's p-value, for different probability distributions (EP and PE models).

Table 6. BNEM estimates, SE and 95% CI for the parameters of the EP distribution, considering the aircraft data.

Parameter	Estimate	SE	95% CI
λ	1.9553	0.0452	(1.8666 ; 2.0439)
β	0.0242	0.0194	(0.0024 ; 0.0289)

Table 7. The AIC, AICc, HQIC and CAIC values, and the p-value from the KS test, for the fitted distributions, considering the data set related to the failure time of 142 devices in an aircraft.

	EP	PE
AIC	**1186.084**	1194.868
AICc	**1186.171**	1194.954
HQIC	**1188.487**	1197.270
CAIC	**1193.996**	1202.780
KS	0.5641	0.0169

Finally, Figure 7 presents the survival function adjusted by different distributions and the Kaplan-Meier estimate.

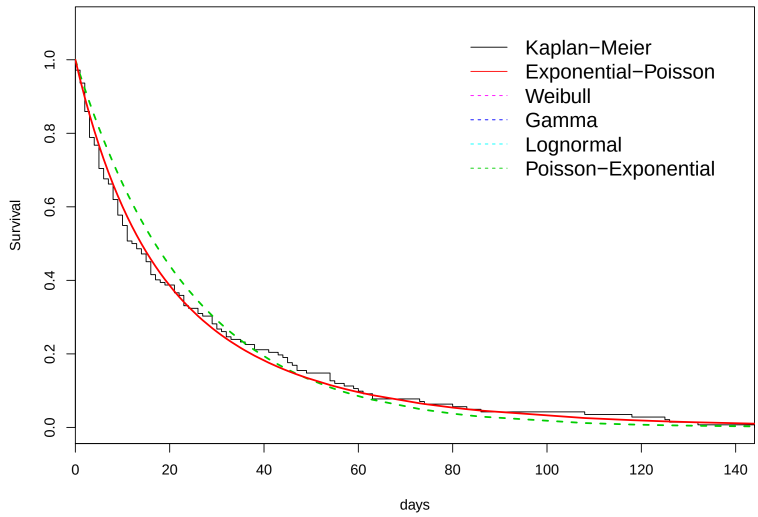

Figure 7. Survival function adjusted by different probability distributions and the Kaplan-Meier estimate, considering the data set related to the failure time of 142 devices in an aircraft.

Therefore, from the proposed methodology, the data related to the failure time of 142 devices in an aircraft can be well-described by the EP distribution. Using the adjusted parameters, we can easily propose an preventive approach for mechanical components (see Ramos et al. 2018). From the quantile function, we can obtain the number of days that are expected to have a certain percentage of failures. Table 8 presents different times of failure assuming different percentages.

Table 8. Days to perform preventive maintenance assuming different percentages of failures, based on the EP distribution.

5%	10%	15%	20%	25%	30%	50%
0.94	1.94	3.02	4.19	5.45	6.83	13.97

The results obtained from Table 8 shows that the preventive maintenance can be performed assuming different percentages of failures. In the case of the proposed data set, the airline assumed that 25% of failures would be the limit and, therefore, they will consider 5 days after the last failure to perform maintenance in the proposed component.

6. Final Comments

In this paper, we derived and compared, via an extensive simulation study, different frequentist estimation methods for the parameters of the EP distribution. The simulations showed that the corrected MLEs obtained from the nested EM algorithm give accurate estimates for both EP parameters, even for small sample sizes. Thus, the proposed estimator is the most efficient estimation procedure among the ones considered in this study, and should be used for all practical purposes. Moreover, from our simulations, we noticed that the estimation procedures that depend on numerical methods to find the solution of nonlinear equations failed in finding the parameter estimates for a significant number of samples. On the other hand, the nested EM algorithm estimators converge in all situations. In this case, we need good initial values to increase the convergence speed, therefore, we provided useful closed-form equations that can be used to initiate the nested EM algorithm.

Our research has also shown that the EP model can be used to describe data with the occurrence of zero values. Such characteristic is not observed in common models, like gamma, Weibull and lognormal distributions. In order to illustrate our proposed methodology, two real data sets were considered. The first is related to total monthly precipitation during June in São Carlos city, Brazil, demonstrating that the EP distribution is a simple alternative to be used in meteorological applications. The second is related to failure time of devices in an aircraft, whose estimation results revealed that such data can also be well-described by the considered distribution.

Disclosure statement

No potential conflict of interest was reported by the authors.

Acknowledgements

The authors are thankful to the Editorial Board and to the reviewers for their valuable comments and suggestions which led to this improved version. The research was partially supported by the Brazilian organizations, CNPq and FAPESP.

References

Abbaspour, K., E. Rouholahnejad, S. Vaghefi, R. Srinivasan, H. Yang, and B. Klove (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology 524*, 733–752.

Barreto-Souza, W. and F. Cribari-Neto (2009). A generalization of the exponential-Poisson distribution. *Statistics & Probability Letters 79*(24), 2493–2500.

Barreto-Souza, W. and R. B. Silva (2015). A likelihood ratio test to discriminate exponential-Poisson and gamma distributions. *Journal of Statistical Computation and Simulation 85*(4), 802–823.

Beven, K. (2011). *Rainfall-Runoff Modelling: The Primer.* John Wiley & Sons.

Cancho, V. G., F. Louzada-Neto, and G. D. Barriga (2011). The Poisson-exponential lifetime distribution. *Computational Statistics & Data Analysis 55*(1), 677–686.

Cheng, R. and N. Amin (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society, Series B (Methodological) 45*(3), 394–403.

Dey, S., D. Kumar, P. L. Ramos, and F. Louzada (2017). Exponentiated chen distribution: Properties and estimation. *Communications in Statistics - Simulation and Computation 46*(10), 8118–8139.

Efron, B. and R. J. Tibshirani (1994). *An Introduction to the Bootstrap.* CRC press.

Gupta, R. D. and D. Kundu (2001). Generalized exponential distribution: Different method of estimations. *Journal of Statistical Computation and Simulation 69*(4), 315–337.

Henningsen, A. and O. Toomet (2011). maxlik: A package for maximum likelihood estimation in r. *Computational Statistics 26*(3), 443–458.

Kao, J. H. (1958). Computer methods for estimating Weibull parameters in reliability studies. *IRE Transactions on Reliability and Quality Control 13*, 15–22.

Kao, J. H. (1959). A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. *Technometrics 1*(4), 389–407.

Karlis, D. (2009). A note on the exponential Poisson distribution: A nested EM algorithm. *Computational Statistics & Data Analysis 53*(1), 894–899.

Koutsoyiannis, D. and A. Langousis (2011). Chapter 27: Precipitation. In S. Wilderer and S. Uhlenbrook (Eds.), *Treatise on Water Science*, pp. 27–78. Oxford: Academic Press.

Kuş, C. (2007). A new lifetime distribution. *Computational Statistics & Data Analysis 51*(9), 4497–4509.

Louzada, F., P. L. Ramos, and G. S. Perdoná (2016). Different estimation procedures for the parameters of the extended exponential geometric distribution for medical data. *Computational and Mathematical Methods in Medicine 2016.*

Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis 51*(2), 904–917.

Lupu, E. C. and S. Lupu (2010). Mixture of two exponential-Poisson distribution. *Mircea cel Batran Naval Academy Scientific Bulletin 13.*

Macera, M. A., F. Louzada, V. G. Cancho, and C. J. Fontes (2015). The exponential-Poisson model for recurrent event data: An application to a set of data on malaria in Brazil. *Biometrical Journal 57*(2), 201–214.

Mann, N. R., R. E. Schafer, and N. D. Singpurwalla (1974). *Methods for Statistical Analysis of Reliability and Life Data.* John Wiley & Sons.

Mazucheli, J., F. Louzada, and M. Ghitany (2013). Comparison of estimation methods for the parameters of the weighted Lindley distribution. *Applied Mathematics and Computation 220*, 463–471.

Preda, V., E. Panaitescu, and R. Ciumara (2011). The modified exponential-Poisson distribution. *Proceedings of the Romanian Academy, Series A 12*(1), 22–29.

R Core Team (2014). *R: A Language and Environment for Statistical Computing. (Version 3.3.1).* Vienna, Austria: R Foundation for Statistical Computing.

Ramos, P. L., F. Louzada, T. K. Shimizu, and A. O. Luiz (2018). The inverse weighted lindley distribution: Properties, estimation and an application on a failure time data. *Communications in Statistics - Theory and Methods 99*, 1–20.

Ramos, P. L., D. C. Nascimento, C. Cocolo, M. J. Nicola, C. Alonso, L. G. Ribeiro, A. Ennes, and F. Louzada (2018). Reliability-centered maintenance: analyzing failure in harvest sugarcane machine using some generalizations of the weibull distribution. *Modelling and Simulation in Engineering 2018.*

Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics 11*, 93–112.

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

Swain, J. J., S. Venkatraman, and J. R. Wilson (1988). Least-squares estimation of distribution functions in Johnson's translation system. *Journal of Statistical Computation and Simulation 29*(4), 271–297.

Teimouri, M., S. M. Hoseini, and S. Nadarajah (2013). Comparison of estimation methods for the Weibull distribution. *Statistics 47*(1), 93–109.

Xu, B., Y. Guo, and N. Zhu (2016). The parameter Bayesian estimation of two-parameter exponential-Poisson distribution and its optimal property. *Journal of Interdisciplinary Mathematics 19*(4), 697–707.

Yan, W.-a., Y.-m. Shi, and Y. Liu (2012). Bayesian estination of exponential Poisson distribution under different loss function. *Fire Control & Command Control 2*, 124–126.