The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart

Francisco Louzada\textsuperscript{a,*}, Mari Roman\textsuperscript{a}, Vicente G. Cancho\textsuperscript{b}

\textsuperscript{a} Department of Statistics, Universidade de São Carlos, Brazil
\textsuperscript{b} Department of Applied Mathematics and Statistics, Universidade de São Paulo, Brazil

Abstract

In this paper, we proposed a new two-parameter lifetime distribution with increasing failure rate, the complementary exponential geometric distribution, which is complementary to the exponential geometric model proposed by Adamidis and Loukas (1998). The new distribution arises on a latent complementary risks scenario, in which the lifetime associated with a particular risk is not observable; rather, we observe only the maximum lifetime value among all risks. The properties of the proposed distribution are discussed, including a formal proof of its probability density function and explicit algebraic formulas for its reliability and failure rate functions, moments, including the mean and variance, variation coefficient, and modal value. The parameter estimation is based on the usual maximum likelihood approach. We report the results of a misspecifications simulation study performed in order to assess the extent of misspecification errors when testing the exponential geometric distribution against our complementary one in the presence of different sample size and censoring percentage. The methodology is illustrated on four real datasets; we also make a comparison between both modeling approaches.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Complementary risk (CR) problems (Basu and Klein, 1982) arise in several areas, such as public health, actuarial science, biomedical studies, demography, and industrial reliability. In classical CR scenarios, the lifetime associated with a particular risk is not observable; rather, we observe only the maximum lifetime value among all risks. 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. The CR dual is the so-called competing risk scenario, in which the lifetime associated with a particular risk is not observable; rather, we observe only the minimum lifetime value among all risks. 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). A difficulty arises if the risks are latent in the sense that there is no information about which factor was responsible for the component failure (or individual death), which can often be observed in field data. We call these latent CR data. On many occasions this information is not available, or it is impossible that the true cause of failure can be specified by an expert. In reliability, 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 cannot be applied when all observations have an unknown cause of failure. Lu and Tsiatis (2001) presents a multiple imputation method for estimating regression coefficients for risk modeling with missing cause of failure. A comparison of two partial likelihood approaches for risk modeling with missing cause of failure is presented in Lu and Tsiatis (2005).

In this context, in this paper, we propose a new distribution family conceived inside a latent CR scenario, in which there is no information about which factor was responsible for the component failure (or individual death), and only the maximum lifetime value among all risks is observed.

Our distribution is based on a generalization of the exponential distribution, which is a widely used lifetime distribution for modeling many problems in lifetime testing and reliability studies. In recent years, several new classes of models have been introduced grounded in its simple, elegant and closed form, such as Gupta and Kundu (1999), which proposed a generalized exponential distribution, which can accommodate data with increasing and decreasing failure rates, Kus (2007), which proposed another modification of the exponential distribution with decreasing failure rate, and Barreto-Souza and Cribari-Neto (2009), which generalizes the distribution proposed by Kus (2007) by including a power parameter in the distribution.

We focus on Adamidis and Loukas (1998), which proposed a variation of the exponential distribution, the exponential geometric (EG) distribution, with decreasing hazard function. Its genesis is based on a competing risk problem in the presence of latent risks (Louzada-Neto, 1999), in the sense that there is no information about which factor was responsible for the component failure, and only the minimum lifetime value among all risks is observed. Our distribution, however, is a counterpart of the EG distribution and thus, hereafter, it shall be called the complementary exponential geometric (CEG) distribution.

The paper is organized as follows. In Section 2, we introduce the new CEG distribution and present some of its properties. Furthermore, we derive the expressions for the probability density function and the survival function, and the rth raw moments of the CEG distribution, including the mean and variance, variation coefficient, and modal value. Also, in this section, we present the inferential procedure. In Section 3, we discuss the relationship between the EG and CEG distributions based on the failure rate function, and report the results of a misspecification study performed in order to verify if we can distinguish between the EG and CEG distributions in the light of the data based on some typical distribution comparison criterion. In Section 4, we fit the EG and CEG distribution to four real datasets. Some final comments in Section 5 conclude the paper.

2. The CEG model

The CEG model is derived as follows. Let M be a random variable denoting the number of failure causes, \( m = 1, 2, \ldots \), and considering \( M \) with geometrical distribution of probability given by

\[
P(M = m) = \theta (1 - \theta)^{m-1},
\]

(1)

where \( 0 < \theta < 1 \) and \( M = 1, 2, \ldots \).

Let us also consider \( t_i, i = 1, 2, 3, \ldots \), realizations of a random variable denoting the failure times, i.e., the time-to-event due to the \( i \)th CR, with \( T_i \) having an exponential distribution with probability index \( \lambda \), given by

\[
f(t_i; \lambda) = \lambda \exp(-\lambda t_i).
\]

(2)

In the latent complementary risk scenario, the number of causes \( M \) and the lifetime \( t_i \) associated with a particular cause are not observable (latent variables), and only the maximum lifetime \( Y \) among all causes is usually observed. So, we only observe the random variable given by

\[
Y = \max(t_1, t_2, \ldots, t_M).
\]

(3)

The following result derives the distribution of \( Y \).

**Proposition 2.1.** If the random variable \( Y \) is defined as in (3), then, considering (1) and (2), \( Y \) is distributed according to a CEG distribution, with probability density function (pdf) given by

\[
f(y) = \frac{\lambda \theta e^{-\lambda y}}{(e^{-\lambda y}(1 - \theta) + \theta)^2}.
\]

(4)

**Proof.** The conditional density function of (3) given \( M = m \) is given by

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

Then, the marginal probability density function of \( Y \) is given by
Fig. 1. Left panel: Probability density function of the CEG distribution. Right panel: Failure rate function of the CEG distribution. We fixed $\lambda = 1$.

\[
\begin{align*}
    f(y) &= \sum_{m=1}^{\infty} m\lambda e^{-\lambda y}(1 - e^{-\lambda y})^{m-1} \times \theta(1 - \theta)^{m-1} \\
    &= \frac{\lambda \theta e^{-\lambda y}}{(1 - e^{-\lambda y})(1 - \theta)} \sum_{m=1}^{\infty} m[(1 - e^{-\lambda y})(1 - \theta)]^m \\
    &= \frac{\lambda \theta e^{-\lambda y}}{(1 - e^{-\lambda y})(1 - \theta)} [1 - (1 - e^{-\lambda y})(1 - \theta)^2] \\
    &= \frac{\lambda \theta e^{-\lambda y}}{[e^{-\lambda y}(1 - \theta) + \theta]^2}.
\end{align*}
\]

This completes the proof. □

The parameter $\lambda$ controls the scale of the distribution, while $\theta$ controls the shape of the distribution. Fig. 1 (left panel) shows the CEG probability density function for $\theta = 0.001, 0.01, 0.2, 0.5, 0.7, \text{and} 0.99$. The function is decreasing if $\theta \geq 1/2$ and unimodal for $\theta < 1/2$. Our approach is complementary to that of Adamidis and Loukas (1998) in the sense that they consider the distribution $\min(t_1, t_2, \ldots, t_M)$ while we deal with $\max(t_1, t_2, \ldots, t_M)$.

The survival function of a CEG distributed random variable is given by

\[
S(y) = \frac{e^{-\lambda y}}{e^{-\lambda y}(1 - \theta) + \theta},
\]

where $y > 0$, $\theta \in (0, 1)$, and $\lambda > 0$. Fig. 2 (right panel) shows some survival function shapes for $\theta = 0.001, 0.01, 0.2, 0.7, \text{and} 0.99$.

We can simulate a CEG distributed variable considering the inverse transformation of the cumulated function given by

\[
Q(u) = F^{-1}(u) = -\lambda^{-1} \ln \left( \frac{\theta(1 - u)}{\theta(1 - u) + u} \right),
\]

where $u$ has a uniform $U(0, 1)$ distribution and $F(y) = 1 - S(y)$ is the distribution function of $Y$.

2.1. Some properties

Some of the most important features and characteristics of a distribution can be studied through its moments, such as the mean, variance, and variation coefficient. A general expression for the $r$th ordinary moment $\mu'_r = E(Y^r)$ of the CEG distribution is obtained analytically, as follows.

Moment-generating of the $Y$ variable, with density function given by (4), can be obtained analytically, if we consider the expression, given in Abramowitz and Stegun (1972, p. 558, equation (15.3.1)):

\[
\Psi(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c - b)} \int_0^1 t^{b-1}(1 - t)^{c-b-1} \frac{e^{-zt}}{(1 - tz)^a} dt.
\]

For any real number $r$, let $\Phi_Y(t)$ be the characteristic function of $Y$; that is, $\Phi_Y(t) = E[e^{itY}]$, where $|rmi$ denotes the imaginary unit. With the preceding notation, we state the following.
Fig. 2. Failure rate function of the CEG distribution (left panel) and the EG distribution (right panel) for the same fixed $\theta$ values. We fixed $\lambda = 1$.

**Proposition 2.2.** For a random variable $Y$ with CEG distribution, we have that its characteristic function is given by

$$\Phi_Y(t) = \frac{\lambda}{\theta(\lambda - it)} \Psi \left(2, 1 - \frac{it}{\lambda}, 2 - \frac{it}{\lambda}, -\beta\right),$$

where $\beta = \frac{1-\theta}{\theta}$ and $i = \sqrt{-1}$.

**Proof.** From (4),

$$\Phi_Y(t) = \int_0^\infty e^{ity} f(y) dy = \int_0^\infty e^{ity} \frac{\theta e^{-\lambda y}}{[e^{-\lambda y} (1-\theta) + \theta]^2} dy = \frac{1}{\theta} \int_0^\infty e^{ity} \frac{\lambda e^{-\lambda y}}{[1 + e^{-\lambda y} (1-\theta)]^2} dy = \frac{1}{\theta} \int_0^1 u^{-\frac{t}{\lambda}} \left(1 + \frac{1-\theta}{\theta} u\right)^2 du,$$

where $u = e^{-\lambda y}$.

Comparing the last integral with (7), we obtain $b = 1 - \frac{t}{\lambda}, c = 2 - \frac{t}{\lambda}, -z = \beta = \frac{1-\theta}{\theta},$ and $a = 2$. Since $\Gamma(s+1) = s\Gamma(s)$, making the appropriate substitutions, the proof is completed. \(\square\)

Jodrá (2008) presents the following result:

$$\frac{1}{\lambda} \Psi \left(2, 1 - \frac{it}{\lambda}, 2 - \frac{it}{\lambda}, -\beta\right) = \sum_{k=0}^{\infty} \frac{(k+1)(-\beta)^k}{\lambda(k+1) - it},$$

where $-\infty < t < \infty, i = \sqrt{-1}$, and $1 - \beta < 1$.

Considering (9), the characteristic function (8) can be rewritten as

$$\Phi_Y(t) = \frac{\lambda}{\theta} \sum_{k=0}^{\infty} \frac{(k+1)(-\beta)^k}{\lambda(k+1) - it},$$

and this enables us to obtain the following result.

**Proposition 2.3.** If the random variable $Y$ has CEG distribution and $r \in N$, then

$$E(y^r) = -\frac{r!}{\lambda^r (1-\theta)^r} L(-\beta, r),$$

where $\beta = \frac{1-\theta}{\theta}$, and $L(-\beta, r) = \sum_{k=1}^{\infty} \frac{(-\beta)^k}{k^r}$ is a polylogarithm function.
Proof. Let $t$ be a real number, and let $\lambda > 0$. From (10), the $r$th derivative of $\Phi_Y$ is given by

$$
\Phi_Y^{(r)}(t) = \frac{\lambda^r (r+1)!}{\theta^r} \sum_{k=0}^{\infty} \frac{(k+1)(-\beta)^k}{(\lambda(k+1)-it)^{r+1}},
$$

with $r = 1, 2, \ldots$.

Setting $t = 0$ in (12), and considering that $E(Y^r) = \Phi_Y^{(r)}(0)/r!$, we have

$$
E[Y^r] = -\frac{\Gamma(r+1)}{(1-\theta)\lambda^r} \sum_{k=1}^{\infty} (-\beta)^k / k^r.
$$

From Jodrá (2008), since $L(-\beta, r) = \sum_{k=1}^{\infty} (-\beta)^k / k^r$, we obtain

$$
E[Y^r] = -\frac{\Gamma(r+1)}{(1-\theta)\lambda^r} L(-\beta, r),
$$

for $\beta \in (0, \infty)$ and $r = 1, 2, \ldots$. □

**Proposition 2.4.** The random variable $Y$ with pdf given by (4) has mean and variance given, respectively, by

$$
E(Y) = -\frac{\ln(\theta)}{\lambda (1-\theta)} \quad \text{and} \quad \text{Var}(Y) = -\frac{1}{\lambda^2 (1-\theta)} \left( 2L(-\beta, 2) + \ln(\theta)^2 \right).
$$

**Proof.** We have $L(-\beta, 1) = -\ln(1 + \beta)$ (Adamidis and Loukas (1998)). Using this result and Proposition 2.3, we easily conclude the proof. □

We have highlighted the fact that, in (13), the variance is a function of the mean. In cases where the mean assumes large values, the variance may not be representative; thus the variation coefficient, $VC$, is an alternative to evaluate the data in this case. From (13), it is given by

$$
VC = -\frac{1-\theta}{\ln(\theta)} \sqrt{-\frac{2L(-\beta, 2)}{1-\theta} - \left( \frac{\ln(\theta)}{1-\theta} \right)^2}.
$$

**Proposition 2.5.** The modal value for the variable $Y$ with pdf given by (4) is given by

$$
\tilde{Y} = \frac{1}{\lambda} \ln \frac{1-\theta}{\theta}.
$$

**Proof.** This proof is directly obtained by solving, considering (4), the equation $df(y)/dy = 0$. □

2.2. Inference

Assuming that the lifetimes are independently distributed and are independent from the censoring mechanism considered, random right censoring, the maximum likelihood estimates (MLEs) of the parameters are obtained by direct maximization of the log-likelihood function given by

$$
\ell(\lambda, \theta) = \log(\lambda \theta) \sum_{i=1}^{n} c_i - \lambda \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} (c_i + 1) \log(e^{-\lambda y_i} (1-\theta) + \theta),
$$

where $c_i$ is a censoring indicator, which is equal to 0 or 1, respectively, if the data is censored or observed. The advantage of this procedure is that it runs immediately using existing statistical packages. We have considered the optim routine of the R package (R Development Core Team, 2008). Large-sample inference for the parameters are based on the MLEs and their estimated standard errors.

For comparison of nested models, which is the case when comparing the CEG distribution with the E distribution, we can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain the likelihood ratio statistics (LRS). For testing $H_0 : \theta = 1$ we consider the LRS, $w_n = 2(l_{CEG} - l_E)$, where $l_E$ and $l_{CEG}$ are the log-likelihoods for the model under the restricted hypothesis $H_0$ and under the unrestricted hypothesis $H_1$ under a sample of size $n$, respectively. Taking into account that the test is performed in the boundary of the parameter space, following Maller and Zhou (1995), the LRS, $w_n$, is assumed to be asymptotically distributed as a symmetric mixture of a chi-squared distribution with one degree of freedom and a point-mass at zero. Then, $\lim_{n \to \infty} P(w_n \leq c) = 1/2 + 1/2P(\chi^2_1 \leq c)$, where $P(\chi^2_1 \leq c)$ denotes a random variable with a chi-squared distribution with one degree of freedom. Large positive values of $w_n$ give favorable evidence to the full model.
3. On the relationship between the EG and CEG distributions

In this section, we discuss some relationship between the EG distribution (Adamidis and Loukas, 1998) and the CEG distribution proposed here.

From (5), the failure rate function of a CEG distributed random variable, according to the relationship \( h(y) = -\frac{d}{dy} \ln(S(y)) \), is given by

\[
h(y) = \frac{\theta \lambda}{\exp(-\lambda y)(1 - \theta) + \theta}.
\]  
(17)

The failure rate function of the EG distribution is given by (Adamidis and Loukas, 1998)

\[
h(y) = \frac{\lambda}{1 - \theta e^{-\lambda y}}.
\]  
(18)

**Proposition 3.1.** The failure rate functions, (17) and (18), converge to \( \lambda \) for \( y \to \infty \), but, for \( y \to 0 \), the hazard function (17) converges to \( \lambda \theta \) while (18) converges to \( \frac{\lambda}{1-\theta} \).

**Proof.** (a) For \( y \to \infty \), for the EG model, \( \lim_{y \to \infty} h(y) = \lim_{y \to \infty} \frac{\lambda}{1 - \theta e^{-\lambda y}} = \lambda \), while for the CEG distribution, \( \lim_{y \to \infty} h(y) = \lim_{y \to \infty} \frac{\theta \lambda}{\exp(-\lambda y)(1 - \theta) + \theta} = \lambda \), concluding the proof of the convergence to \( \lambda \).

(b) For \( y \to 0 \), for the EG model, \( \lim_{y \to 0} h(y) = \lim_{y \to 0} \frac{\lambda}{1 - \theta e^{-\lambda y}} = \frac{\lambda}{1-\theta} \), while for the CEG distribution, \( \lim_{y \to 0} h(y) = \lim_{y \to 0} \frac{\theta \lambda}{\exp(-\lambda y)(1 - \theta) + \theta} = \lambda \theta \), concluding the proof. \( \square \)

Fig. 2 shows the behavior of the failure rate functions of both distributions for \( \theta = 0.001, 0.01, 0.2, 0.5, 0.7, \) and 0.99. The CEG failure rate function (17) increases with \( y \) (left panel), while the EG failure rate function (18) decreases (right panel), but both converge to \( \lambda \) for \( y \to \infty \), corroborating Proposition 3.1. The behavior of the EG and CEG failure rate functions perfectly justifies the word “complementary” in the name of our model.

A misspecification study was performed in order to verify if we can distinguish between the EG and CEG distributions in the light of a dataset based on a typical comparison criterion for non-nested distributions. The EG and CEG distributions are separated family of hypothesis (Cox, 1961). Then, for instance, we consider here the Akaike information criterion (AIC), \(-2\ell(\hat{\theta}, \hat{\lambda}) + 2q\), where \( q \) is the number of parameters in the distribution. The preferred distribution is the one with the smaller AIC value. However, since the EG and the CEG distributions have the same number of parameters, the two criteria, the AIC and the Bayesian information criterion (BIC), identify the same distribution.

We generate 1000 samples of the CEG distribution (4) by considering the inverse transformation of the cumulated density function (6). We consider different sample sizes \( (n = 10, 20, 30, 50, \) and 100) and different censoring percentages \( (p = 0.1, 0.2, 0.35, \) and 0.50). The same procedure was performed for the EG distribution. We fixed \( \lambda = 0.5 \) for both distributions and \( \theta = 0.75 \) and \( \theta = 0.25 \) for the EG and CEG distributions, respectively. Both distributions were fitted to each sample and their AIC was calculated. Table 1 shows the percentage of time that the distribution, which originated the sample, was the best fitted distribution according to the AIC. We discovered that it is usually possible to discriminate between the distributions even for small samples in the presence of heavy censoring.

4. Application

In this section, we compare the EG and CEG distribution fits on four datasets extracted from the literature, two with increasing failure rate function and two with decreasing failure rate function, two of them with censoring. The first dataset, hereafter T1, was extracted from Lawless (2003). The lifetimes are number of million revolutions before failure for each of 23 ball bearings on an endurance test of deep-groove ball bearings. The second dataset, T2, refers to the serum-reversal time (days) of 143 children contaminated with HIV by vertical transmission from the University Hospital of the Ribeirão Preto School of Medicine (Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto) from 1986 to 2001 Perdoná (2006). Serum-reversal can occur in children born from mothers infected with HIV. The third dataset, T3, consists of the number of successive failures for the air-conditioning system of each member in a fleet of 13 Boeing 720 jet airplanes. The pooled data with 214 observations was considered by Adamidis and Loukas (1998), who proposed the EG distribution discussed here. It was first analyzed by Proschan (1963) and discussed further by Daihya and Gurland (1972), Gleser (1989), Kus

<table>
<thead>
<tr>
<th>n/p</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.726/0.998</td>
<td>0.875/0.999</td>
<td>0.936/0.999</td>
<td>0.988/0.999</td>
<td>0.999/0.999</td>
</tr>
<tr>
<td>0.2</td>
<td>0.711/0.995</td>
<td>0.846/0.999</td>
<td>0.923/0.999</td>
<td>0.980/0.999</td>
<td>0.999/0.999</td>
</tr>
<tr>
<td>0.35</td>
<td>0.672/0.984</td>
<td>0.837/0.999</td>
<td>0.909/0.999</td>
<td>0.971/0.999</td>
<td>0.999/0.999</td>
</tr>
<tr>
<td>0.5</td>
<td>0.648/0.978</td>
<td>0.791/0.992</td>
<td>0.880/0.999</td>
<td>0.953/0.993</td>
<td>0.999/0.999</td>
</tr>
</tbody>
</table>
Table 2
The $-\ell(\psi_g)$ values and the AIC and BIC values for the EG and CEG fitted distributions.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>EG</th>
<th>CEG</th>
<th>AIC</th>
<th>BIC</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>121.43</td>
<td>114.35</td>
<td>246.87</td>
<td>232.70</td>
<td>234.97</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>860.87</td>
<td>805.47</td>
<td>1725.74</td>
<td>1614.95</td>
<td>1620.87</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>1175.92</td>
<td>1178.76</td>
<td>2355.85</td>
<td>2361.52</td>
<td>2368.25</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>219.31</td>
<td>223.92</td>
<td>442.62</td>
<td>451.84</td>
<td>455.22</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 3. Empirical scaled TTT transform for the data. Left panel: T1 and T2. Right panel: T3 and T4.](image)

(2007), and Barreto-Souza et al. (2010). The fourth dataset, T4, corresponds to the survival data on 40 advanced lung cancer patients, who underwent two chemotherapy treatments, but the covariates are not considered in this study. There was three censored observations, which correspond to patients who were still alive at the time the data were collected Prentice (1973).

First, in order to identify the shape of the lifetime data failure rate function, we shall consider a graphical method based on the Total Time on Test (TTT) plot (Aarset, 1987) (Fig. 3). In its empirical version, the TTT plot is given by $G(r/n) = \{(\sum_{i=1}^{r} Y_{i,n}) - (n - r) Y_{r,n})/\sum_{i=1}^{n} Y_{i,n}\}$, where $r = 1, \ldots, n$ and $Y_{i,n}$ represents the order statistics of the sample. It has been shown that the failure rate function is increasing (decreasing) if the TTT plot is concave (convex). Although the TTT plot is only a sufficient condition, not a necessary one, for indicating the failure rate function shape, it is used here as a crude indication of its shape. The left panel of Fig. 4 shows concave TTT plots for the T1 and T2 datasets, indicating increasing failure rate functions, which can be properly accommodated by a CEG distribution. The right panel of Fig. 4 shows convex TTT plots for the T3 and T4 datasets, indicating decreasing failure rate functions, which can be properly accommodated by an EG distribution.

However, we fitted both distributions for all datasets. Table 2 provides the $-\ell(\psi_g)$ values and the AIC and BIC values for both distributions. They provide evidence in favor of our CEG distribution for the T1 and T2 datasets and in favor of the EG distribution for the T3 and T4 datasets. These results are corroborated by the empirical Kaplan–Meier survival functions and the fitted survival functions via the EG and CEG distributions shown in Fig. 4.

The MLEs (and their corresponding standard errors in parentheses) of the CEG distribution parameters are given by $\hat{\lambda} = 0.0435$ (0.0096) and $\hat{\theta} = 0.0554$ (0.0449) for dataset T1, and $\hat{\lambda} = 0.0083$ (0.0007) and $\hat{\theta} = 0.0194$ (0.0076) for dataset T2. The MLEs (and their corresponding standard errors in parentheses) of the EG distribution parameters are given by $\hat{\lambda} = 0.0080$ (0.0013) and $\hat{\theta} = 0.4264$ (0.1432) for dataset T3, and $\hat{\lambda} = 0.0021$ (0.0016) and $\hat{\theta} = 0.8426$ (0.1267) for dataset T4.

For the sake of illustration, we fitted the exponential distribution to datasets T1 and T2, and compared the fittings to the CEG distribution fittings by considering the test procedure presented in Section 2.2. $w_n$ is equal to 14.178 and 110.788, respectively, for T1 and T2, much greater than $1/2 + 1/2P(\chi^2_1 \leq c) = 2.42$, leading to strong evidence in favor of the CEG distribution for both datasets.

Also, as pointed out by an anonymous referee, since a typical data analyst would use a traditional parametric model, such as the Weibull distribution, in order to fit the data, we fitted the Weibull distribution to datasets T1 and T2 and compared the fittings with the CEG distribution ones. For instance, we considered the AIC and BIC in order to compare their fittings, since the Weibull and CEG distributions are separated family of hypothesis, Cox (1961). The AIC and BIC values for the Weibull distribution for dataset T1 are 231.36 and 233.63, respectively, and they are 1630.52 and 1636.46, respectively, for
dataset $T_2$. These results lead to evidence in favor of the Weibull distribution for dataset $T_1$ and in favor of the CEG distribution for dataset $T_2$, showing the importance of the new distribution, which can be viewed as a competitor to traditional survival distributions.

5. Concluding remarks

In this paper, a new lifetime (CEG) distribution has been provided and discussed. The CEG distribution is complementary to the EG distribution proposed by Adamidis and Loukas (1998), accommodates increasing failure rate functions, and arises on latent complementary risk scenarios, where the lifetime associated with a particular risk is not observable, but only the maximum lifetime value among all risks. The properties of the proposed distribution have been discussed, including a formal proof of its probability density function and explicit algebraic formulas for its reliability and failure rate functions, moments, including the mean and variance, variation coefficient, and modal value. Maximum likelihood inference is implemented straightforwardly. From a misspecification simulation study performed in order to assess the extent of the misspecification errors when testing the EG distribution against the CEG one, we observed that it is usually possible to discriminate between both distributions even for small samples in the presence of heavy censoring. The practical importance of the new distribution and its counterpart was demonstrated in two applications: the CEG distribution provided the better fitting in comparison with the EG one.

Acknowledgements

The research of Francisco Louzada and Vicente G. Cancho is supported by the Brazilian organization CNPq.
References