The Complementary Exponential-Geometric Distribution for Lifetime Data

Mari Roman\textsuperscript{a}, Francisco Louzada-Neto\textsuperscript{a} and Vicente G. Cancho\textsuperscript{b}
\textsuperscript{a}Universidade Federal de S\~ao Carlos \textsuperscript{b}Universidade de S\~ao Paulo

Abstract

In this paper we proposed a new two-parameters lifetime distribution with increasing failure rate, the complementary exponential geometric distribution, which is complementary to the exponential geometric model proposed by Adamidis & Loukas (1998). The new distribution arises on a latent complementary risks scenarios, where 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 prove 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 misspecification simulation study performed in order to assess the extent of misspecification errors when testing the exponential geometric distribution against its complementary one in presence of censoring data. The methodology is illustrated on four real data set, where we also made a comparison between both modelling approach.

Keywords: Complementary Risks, Exponential Distribution, Geometric Distribution, Survival Analysis, censured data, Exponential-Geometric Distribution.

1 Introduction

The exponential distribution is widely used for modeling many problems in lifetime testing and reliability studies. However, the exponential distribution does not provide a reasonable parametric fit for some practical applications where the underlying failure rates are nonconstant, presenting monotone shapes. In recent years, in order to overcome such problem, new classes of models were introduced grounded in its simple elegant and close form, and interested readers can refer to Gupta & 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 & Cribari-Neto (2009), which generalizes the distribution proposed by Kus (2007) by including a power parameter in his distribution.

We focus on Adamidis & 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 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.

In this paper, following Adamidis & Loukas (1998), we propose a new distribution family based on a complementary risk problem (Basu & J., 1982) in presence of latent risks, in the sense that there is no information about which factor was responsible for the component failure but only the maximum lifetime value among all risks is observed. The distribution is a counterpart of the EG distribution and then, hereafter shall be called complementary exponential-geometric (CEG) distribution.

The paper is organized as follows. In Section 2, we introduce the new CEG distribution and present some its properties. We also show that the failure rate function can be increasing. Furthermore, we derive the expressions for the $r$-th 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 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 usual distribution comparison criterion. In Section 4 we fit the EG and CEG distribution to four dataset. Some final comments in Section 5 conclude the paper.

2 The model

Let $Y$ a nonnegative random variable denoting the lifetime of a component in some population. The random variable $Y$ is said to have a CEG distribution with parameters $\lambda > 0$ and $0 < \theta < 1$ if its probability density function (pdf) is given by,

$$f(y) = \frac{\lambda \theta \exp\{-\lambda y\}}{(\exp\{-\lambda y\}(1 - \theta) + \theta)^2}. \quad (1)$$

The parameter $\lambda$ controls the scale of the distribution, while $\theta$ controls the shape of the distribution. The Figure 1 (left panel) shows the CEG probability density function for $\theta = 0.001, 0.01, 0.2, 0.7, 0.99$. The function is decreasing if $\theta > 1/2$ and unimodal for $\theta < 1/2$.

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

$$S(y) = \frac{\exp\{-\lambda y\}}{\exp\{-\lambda y\}(1 - \theta) + \theta}. \quad (2)$$
Figure 1: Left panel: Probability density function of the CEG distribution. Right panel: Failure rate function of the CEG distribution. We fixed $\lambda = 1$.

where, $y > 0$, $\theta \in (0, 1)$, $\lambda > 0$.

From (2), the failure rate function, according to the relationship $h(y) = -\frac{d}{dy} \ln(S(y))$, is given by,

$$h(y) = \frac{\theta \lambda}{\exp\{-\lambda y\}(1 - \theta) + \theta}.$$  \hspace{1cm} (3)

The initial and long-term hazard function values are finite and given by $h(0) = \lambda \theta$ and $h(\infty) = \lambda$. The failure rate (3) is increasing throughout the parametrical space, as shown in the Figure 2 (right panel), which shows some failure rate function shapes for $\theta = 0.001, 0.01, 0.2, 0.7, 0.99$.

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

$$Q(u) = F^{-1}(u) = \ln((1 - u)^2 + \frac{u(1-u)}{\theta}) \lambda,$$  \hspace{1cm} (4)

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

2.1 Genesis

In the classical complementary risks scenarios (Basu & J., 1982) 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. Complementary risks problems arise in
several areas and full statistical procedures and extensive literature are available with this and

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, which can be often observed in field data. We
call these latent Complementary risks data. On many occasions this information is not
available or it is impossible that the true cause of failure is 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 & 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 & 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 & Tsiatis (2005).

Our model can be 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,

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

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

Let’s also consider $t_i, i = 1, 2, 3, \ldots$ realizations of a random variable denoting the failure
times, ie, the time-to-event due to the $j$-th CR and $T_i$ has an exponential distribution of
probability index by $\lambda$, given by,

$$f(t_i; \lambda) = \lambda \exp\{-\lambda t_i\}.$$  

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

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

The following result shows that the random variable $Y$ have probability density function
given by (1).

**Proposition 2.1** If the random variable $Y$ is defined as 7, then, considering (5) and (6), $Y$ is
distributed according to a CEG distribution, with probability density function given by (1).
The conditional density function of (7) given $M = m$ is given by

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

Them, the marginal probability density function of $Y$ is given by

$$f(y) = \sum_{m=1}^{\infty} m\lambda e^{-\lambda y} \left[1 - e^{-\lambda y}\right]^{m-1} \times \theta(1 - \theta)^{m-1}$$

$$= \frac{\lambda \theta e^{-\lambda y}}{(1 - e^{-\lambda y})(1 - \theta)} \sum_{m=1}^{\infty} m \left[\left(1 - e^{-\lambda y}\right)(1 - \theta)\right]^{m-1}$$

$$= \frac{\lambda \theta e^{-\lambda y}}{(1 - e^{-\lambda y})(1 - \theta)} \sum_{m=1}^{\infty} \left[\left(1 - e^{-\lambda y}\right)(1 - \theta)\right]^{m-1}$$

$$= \frac{\lambda \theta e^{-\lambda y}}{(1 - e^{-\lambda y})(1 - \theta)} \left[1 - (1 - e^{-\lambda y})(1 - \theta)\right]^2$$

This completes the proof.

The CEG distribution parameters have a direct interpretation in terms of complementary risks. The $\theta$ represents the rate of the number of complementary risks while $\lambda$ denotes the failure rate.

In comparison with the model formulated by Adamidis & Loukas (1998), we follow the opposite way, while they took the component lifetime as $Y = \min(t_1, t_2, \ldots, t_M)$, we consider $Y = \max(t_1, t_2, \ldots, t_M)$, some differences between both models shall be presented later.

### 2.2 Some Properties

Some of the most important features and characteristics of a distribution can be studied through its moments, such mean, variance and variation coefficient. A general expression for $r$-th ordinary moment $\mu'_r = E(Y^r)$ of the CEG distribution is obtained analytically, as it follows. Moment-generating of the $Y$ variable, with density function given by (1) can be obtained analytically, if we consider the expression, given in Abramowitz & Stegun (1972), p.558, equation (15.3.1),

$$F(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c - b)} \int_0^1 t^{b-1}(1 - t)^{c-b-1} \left(1 - tz\right)^a dt. \quad (8)$$

For any real number $t$, let $\Phi_Y(t)$ be the characteristic function of $Y$, that is, $\Phi_Y(t) = E[e^{itY}]$, where $i$ denotes the imaginary unit. With the preceding notations, we state the following.
Proposition 2.2 For the random variable $Y$ with CEG distribution, we have that, its characteristic function is given by

$$\Phi_Y(t) = \frac{\Lambda}{\theta(\lambda - it)} F(2, 1 - \frac{it}{\lambda}, 2 - \frac{it}{\lambda}, -\beta)$$  \hspace{1cm} (9)

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

Proof 2.2

$$\Phi_Y(t) = \int_0^\infty e^{ity} f(y)dy$$

$$= \int_0^\infty e^{ity} \frac{\theta e^{-\lambda y}}{[e^{-\lambda y}(1 - \theta) + \theta]^2} dy$$

$$= \frac{1}{\theta} \int_0^\infty e^{ity} \frac{e^{-\lambda y} \lambda e^{-\lambda y}}{[1 + e^{-\lambda y}(1 - \theta)]^2} dy$$

$$= \frac{1}{\theta} \int_0^1 \frac{u^{-\frac{\beta}{\lambda}}}{(1 + \frac{\theta}{\lambda} u)^2} du$$

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

Comparing the last integral with (8), obtain, $b = 1 - \frac{it}{x}$, $c = 2 - \frac{it}{x}$, $c - b = 0$, $-z = \beta = \frac{1 - \theta}{\theta}$, $a = 2$. Since $\Gamma(s + 1) = s\Gamma(s)$, making the appropriate substitutions, the prof is completed.

Jodr´ a (2008) presents the following result,

$$\frac{1}{(\lambda - it)} F(2, 1 - \frac{it}{\lambda}, 2 - \frac{it}{\lambda}, -\beta) = \sum_{k=0}^{\infty} \frac{(k + 1)(-\beta)^k}{\lambda(k + 1) - it}$$  \hspace{1cm} (10)

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

Considering (10), the characteristic function (9) can be rewritten as,

$$\Phi_Y(t) = \frac{\Lambda}{\theta} \sum_{k=0}^{\infty} \frac{(k + 1)(-\beta)^k}{\lambda(k + 1) - it}$$  \hspace{1cm} (11)

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

$$E(y^r) = -\frac{r!}{\lambda^n(1 - \theta)} L(-\beta, r) ,$$  \hspace{1cm} (12)

where, $\beta = \frac{\theta - 1}{\theta}$, and $L(-\beta, r) = \sum_{k=1}^{\infty} \frac{(-\beta)^k}{k^r}$ is the generalization of Euler’s dilogarithm function of $\beta$.  

6
Proof 2.3 Let \( t \) be a real number, \( \lambda > 0 \). From (11), the \( r \)th derivative of \( \Phi_Y \) is given by,

\[
\Phi_Y^{(r)}(t) = \frac{\lambda \Gamma(r+1) i^r}{\theta} \sum_{k=0}^{\infty} \frac{(k+1)(-\beta)^k}{(\lambda k + 1 - it)^{r+1}}
\]  

(13)

\( r = 1, 2, \ldots \)

Setting \( t = 0 \) in 13 and considering that \( E(Y^r) = \Phi_Y^{(r)}(0)/i^r \), we have

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

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

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

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

\[\blacksquare\]

**Proposition 2.4** The random variable \( Y \) with pdf given by (1), has mean and variance given respectively by,

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

(14)

**Proof 2.4** We have, \( L(-\beta, 1) = \ln(1+\beta) \) (Adamidis & Loukas (1998)). Using this result and the proposition 2.3, we easily concludes the proof.

\[\blacksquare\]

We highlighted the fact that, in (14), the variance is 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 (14), it is given by,

\[
VC = -\frac{1-\theta}{\ln(\theta)} \sqrt{2L(\beta, 2) - \frac{(\ln(\theta))^2}{1-\theta}}.
\]  

(15)

**Proposition 2.5** The modal value for the variable \( Y \) with pdf given by (1) is given by,

\[
\hat{Y} = \frac{1}{\lambda} \ln \left( \frac{1-\theta}{\theta} \right).
\]  

(16)

**Proof 2.5** This proof is directly obtained by solving, from (1), the equation \( df(y)/dy = 0 \).

\[\blacksquare\]
2.3 Inference

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

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

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

In order to compare distributions we consider the \(-\ell(\hat{\psi}_g)\) values and the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are defined, respectively, by

\[-2 \log L(\hat{\psi}_g) + 2q\] and \[-2 \log L(\hat{\psi}_g) + q \log(n),\]

where \( \hat{\psi}_g = (\hat{\lambda}_g, \hat{\theta}_g) \) is the MLE vector under the distribution \( g \), \( q \) is the number of parameters estimated under the distribution \( g \) and \( n \) is the sample size. The best distribution corresponds to a lower \(-\ell(\hat{\psi}_g)\), AIC and BIC values.

3 On the relationship between the EG and CEG distributions

In this section, we discuss some relationship between the EG distribution (Adamidis & Loukas, 1998) and CEG distribution proposed here. The failure rate function of the ED distribution is given by

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

Proposition 3.1 The failure rate functions, (3) and (18), of both distributions diverge for \( y \to 0 \) but converge to \( \lambda \) for \( y \to \infty \).

Proof 3.1 (a) 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, \( \lim_{y \to 0} h(y) = \lim_{y \to 0} \frac{\lambda \theta}{(1 - \theta) e^{-\lambda y} + \theta} = \frac{\lambda \theta}{(1 - \theta) \theta} = \lambda \theta \), concluding the proof of the initial divergence.

(b) 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, \( \lim_{y \to \infty} h(y) = \lim_{y \to \infty} \frac{\lambda \theta}{(1 - \theta) e^{-\lambda y} + \theta} = \frac{\lambda \theta}{\theta} = \lambda \), concluding the proof of the converge to \( \lambda \).
The Figure 2 shows the behavior of failure rate functions of both distributions for \( \theta = 0.50, 0.70, 0.85 \). The CEG failure rate function (3) increases while the EG failure rate function (18) decreases with \( y \), but both converge to \( \lambda \) for \( y \to \infty \) corrobarating with the Proposition 3.1.

A misspecification study was performed in order to verify if we can distinguish between the EG and CEG distributions in the light of a data set based on the criterions described in Section (2.3). We generate 1,000 samples of the CEG distribution (1) by considering the inverse transformation of cumulated density function (4). We consider different sample sizes \( n = 10, 20, 30, 50, 100, 200 \) and different censoring percentages \( p = 0.1, 0.2, 0.35, 0.50 \). 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 then, based on the AIC we decided for the best distribution. We however point out that since the EG and the CEG distributions considered here have the same number of parameters, the three criterions, \( \max \log L(\cdot) \), AIC and BIC, identify the same distribution. Table 1 shows the percentage of time that the distribution, which originated the sample, was the best fitted distribution according to the AIC. It is usually possible to discriminate between the distributions even for small samples in presence of heavy censoring.
Table 1: Percentage of times that the (EG / CEG) distribution, which originated the sample, was the best fitted distribution.

<table>
<thead>
<tr>
<th>n/p</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</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>
<td>0.999/0.999</td>
</tr>
<tr>
<td>0.20</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>
<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>
<td>0.999/0.999</td>
</tr>
<tr>
<td>0.50</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>0.995/0.999</td>
<td>0.999/0.999</td>
</tr>
</tbody>
</table>

4 Application

In this section, we compare the EG and CEG distribution fits on four data set extracted from the literature, two with increasing failure rate function and two with decreasing failure rate function, two of them with censuring. The first data set, hereafter \( T_1 \), was extracted from Lawless (2003). The lifetimes are number of million revolutions before failure for each one of the 23 ball bearing on an endurance test of deep groove ball bearings. The second set, \( T_2 \), refers to the recurrence times, in months, of a group of 20 patients with bladder cancer, with the patients underwent a laser surgical procedure, available in Colossino & Giolo (2006). The third data set, \( T_3 \), consists of the number of successive failure 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 & Loukas (1998), which proposed the EG distribution discussed here. It was first analyzed by Proschan (1963) and discussed further by Dahiya & Gurland (1972), Gleser (1989), Kus (2007) and Barreto-Souza et al. (2008). The fourth data set, \( T_4 \) correspond the lifetimes, in days, of 52 patients from the Stanford heart transplant program, which are composed by times to transplant and survival times after transplant, there are 18 data with censure, almost 35% of censure. The pooled data is present in Turnbull et al. (1974) and it was discussed further by Tierney & Kadane (1986).

Firstly, in order to identify the shape of a lifetime data failure rate function we shall consider a graphical method based on the TTT plot 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 \( r = 1, \ldots, n \) and \( Y_{i:n} \) represent 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 indicative of its shape. The left panel of Figure 4 shows concave TTT plots for the \( T_1 \) and \( T_2 \) data set, indicating increasing failure rate functions, which can be properly accommodated by a CEG distribution. The right panel of Figure 4 shows convex TTT plots for the \( T_3 \) and \( T_4 \) data set, indicating decreasing failure rate functions, which can be properly...
accommodated by a EG distribution.

We however, fitted both distribution for all data set. Table 4 provides the $-\ell(\hat{\psi}_g)$ values and the AIC and BIC criterion values for both distributions. They provide evidence in favour of our CEG distribution for the $T_1$ and $T_2$ dataset and in favour of the EG distribution for the $T_3$ and $T_4$ dataset. This results are corroborated by the empirical Kaplan-Meier survival functions and the fitted survival functions via the EG and CEG distributions shown in Figure 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 the dataset $T_1$, and $\hat{\lambda} = 0.1052(0.0358)$ and $\hat{\theta} = 0.2052(0.1758)$ for the dataset $T_2$. The MLEs (and their corresponding standard errors in parentheses) of the EG distribution parameters are given by $\hat{\lambda} = 0.0080(0.0013)$ and $\hat{\theta} = 0.4263(0.1432)$ for the dataset $T_3$, and $\hat{\lambda} = 1.06 \times 10^{-6}(0.0003)$ and $\hat{\theta} = 0.9997(0.0017)$ for the dataset $T_4$.

5 Concluding remarks

In this paper a new lifetime (CEG) distribution is provided and discussed. The CEG distribution is complementary to the EG distribution proposed by Adamidis & Loukas (1998), accommodates increasing failure rate functions, and arises on a latent complementary risks 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 are discussed, including a formal
Figure 4: Kaplan Meier curve with estimated survival function via EG and CEG distributions, for the four dataset. Top left panel: $T_1$; Top right panel: $T_3$. Bottom left panel: $T_2$; Bottom right panel: $T_4$. 
Table 2: The $-\ell(\hat{\psi}_g)$ values and the AIC and BIC criterion values for the EG and CEG fitted distributions

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$-\ell(\cdot)$</th>
<th>AIC</th>
<th>BIC</th>
<th>$-\ell(\cdot)$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>121.43</td>
<td>246.87</td>
<td>249.15</td>
<td>114.93</td>
<td>232.70</td>
<td>234.97</td>
</tr>
<tr>
<td>T2</td>
<td>68.27</td>
<td>140.55</td>
<td>142.54</td>
<td>66.62</td>
<td>137.23</td>
<td>139.22</td>
</tr>
<tr>
<td>T3</td>
<td>1175.92</td>
<td>2355.85</td>
<td>2362.57</td>
<td>1178.76</td>
<td>2361.52</td>
<td>2368.25</td>
</tr>
<tr>
<td>T4</td>
<td>241.17</td>
<td>486.34</td>
<td>490.25</td>
<td>248.21</td>
<td>500.42</td>
<td>504.32</td>
</tr>
</tbody>
</table>

prove 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 presence of heavy censoring. The practical importance of the new distribution and its counterpart was demonstrated in two applications where the CEG distribution provided the better fitting in comparison with the EG one. And also, in two other applications, where the EG distribution provided the better fitting in comparison with our CEG one.

Acknowledgments: The research of Francisco Louzada-Neto and Vicente G. Cancho are supported by the Brazilian organization CNPq.

References


