Revista Colombiana de Estadística
Junio 2014, volumen 37, no. 1, pp. 25 a 34

A New Extension of the Exponential Distribution

Una nueva extensión de la distribución exponencial

Yolanda M. Gómez[1,a], Heleno Bolfarine[1,b], Héctor W. Gómez[2,c]

[1]Instituto de Matematica e Estatística, Universidade de São Paulo, São Paulo, Brazil

[2]Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile

Abstract

The present paper considers an extension of the exponential distribution based on mixtures of positive distributions. We study the main properties of this new distribution, with special emphasis on its moments, moment generator function and some characteristics related to reliability studies. We also discuss parameter estimation considering the maximum likelihood and moments approach. An application reveals that the model proposed can be very useful in fitting real data. A final discussion concludes the paper.

***Key words*:** Exponential distribution, Mixtures of distribution, Likelihood.

Resumen

En el presente paper se considera una extensión de la distribución exponencial basada en mezclas de distribuciones positivas. Estudiamos las principales propiedades de esta nueva distribución, con especial énfasis en sus momentos, función generadora de momentos, y algunas características relacionadas a estudios de confiabilidad. También se analiza la estimación de parámetros a través de los métodos de momentos y de máxima verosimilitud. Una aplicación muestra que el modelo propuesto puede ser muy útil para ajustar datos reales. Una discusión final concluye el artículo.

***Palabras clave*:** distribución exponencial, mezcla de distribuciones, verosimilitud.

[a]Professor. E-mail: ymgomez@ime.usp.br
[b]Professor. E-mail: hbolfar@ime.usp.br
[c]Professor. E-mail: hector.gomez@uantof.cl

1. Introduction

In lifetime data analysis it is usually the case that models with monotone risk functions are preferred as is the case of the gamma distribution. For some models there are no closed form risk functions (such as the Gamma model) and numerical integration might be required for its computation. In recent statistical literature modified extensions of the exponential distributions have been proposed to contour such difficulties. For example, Gupta & Kundu (1999) and Gupta & Kundu (2001) introduced an extension of the exponential distribution typically called the generalized exponential (GE) distribution. Therefore, it is said that the random variable X follows the GE distribution if its density function is given by

$$g_1(x;\alpha,\beta)=\alpha\beta e^{-\alpha x}(1-e^{-\alpha x})^{\beta-1}$$

where $x>0$, $\alpha>0$ and $\beta>0$. We use the notation $X\sim GE(\alpha,\beta)$ for a random variable with such distribution.

More recently, Nadarajah & Haghighi (2011) introduced another extension of the exponential model, so that a random variable X follows the Nadarajah and Haghighi's exponential distribution (NHE) if its density function is given by

$$g_2(x;\alpha,\beta)=\alpha\beta(1+\alpha x)^{\beta-1}e^{\{1-(1+\alpha x)^{\beta}\}}$$

where $x>0$, $\alpha>0$ and $\beta>0$. We use the notation $X\sim NHE(\alpha,\beta)$.

Both distributions have the exponential distribution (E) with scale parameter α, as a special case when $\beta=1$, that is,

$$g_1(x;\alpha,\beta=1)=g_2(x;\alpha,\beta=1)=\alpha e^{-\alpha x}$$

where $x>0$, $\alpha>0$ with the notation $X\sim E(\alpha)$. Other extensions of the exponential model in the survival analysis context are considered in the Marshall & Olkin's (2007) book.

The main object of this paper is to present yet another extension for the exponential distribution that can be used as an alternative to the ones mentioned above. We discuss some properties for this new distribution like its moments and moment generating function which can be used for parameter estimation as starting values for computing maximum likelihood estimators.

The paper is organized as follows. Section 2 delivers the density and distribution functions, moments, moment generating function, asymmetry and kurtosis coefficients and hazard function. Section 3 is devoted to parameter estimation based on maximum likelihood and moments approach. It is recommended that the moment estimators are used to initialize the maximum likelihood approach. In Section 4 an application to a real data set is presented and comparisons between the proposed model and other extensions of the exponential distribution are reported. The main conclusion is that the new model can perform well in applied situations.

2. Density and Properties

A random variable X is distributed according to the extended exponential distribution (EE) with parameters α and β if its density function is given by

$$f(x;\alpha,\beta)=\frac{\alpha^2(1+\beta x)e^{-\alpha x}}{\alpha+\beta} \tag{1}$$

where $x>0$, $\alpha>0$ and $\beta\geq 0$. We use the notation $X\sim EE(\alpha,\beta)$.

Figures 1 and 2 depict the behavior of the distribution for some parameter values.

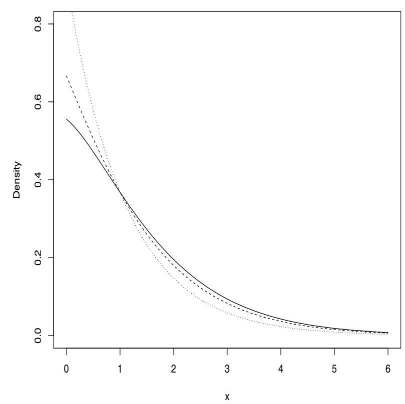

FIGURE 1: Plots of the $EE(1,0.8)$ (solid line), $EE(1,0.5)$ (dashed line), $EE(1,0.1)$ (dotted line).

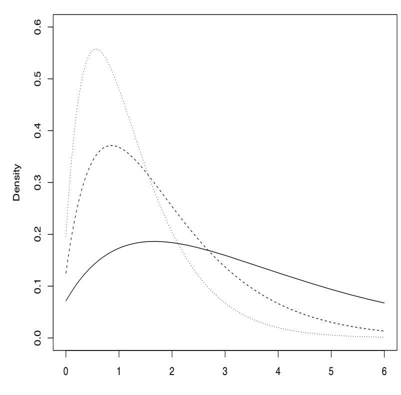

FIGURE 2: Plots of the $EE(0.5,3)$ (solid line), $EE(1,7)$ (dashed line), $EE(1.5,10)$ (dotted line).

2.1. Properties

Let $X \sim EE(\alpha, \beta)$, $Y \sim E(\alpha)$ and $Z \sim Gamma(2, \alpha)$. Then, the distribution function for the random variable X is given by

$$F_X(x) = \frac{\alpha + \beta - (\beta + \alpha + \alpha\beta x)e^{-\alpha x}}{\alpha + \beta} \tag{2}$$

The expectation and variance are given by

$$E(X) = \frac{\alpha + 2\beta}{\alpha(\alpha + \beta)}$$

$$Var(X) = \frac{\alpha^3 + 5\alpha^2\beta + 6\alpha\beta^2 + 2\beta^3}{\alpha^5 + 3\alpha^4\beta + 3\alpha^3\beta^2 + \alpha^2\beta^3}$$

The moment generation function can also be obtained in closed form and is given by

$$M_X(t) = \frac{\alpha^2(\alpha + \beta - t)}{(\alpha + \beta)(t - \alpha)^2} \tag{3}$$

It also follows that its density can be obtained as a mixture of two positive ones, namely,

$$f_X(x; \alpha, \beta) = \frac{\alpha}{\alpha + \beta} f_Y(x; \alpha) + \frac{\beta}{\alpha + \beta} f_Z(x; \alpha) \tag{4}$$

Using the representation as a mixture of two positive densities, we can provide a general representation for the distribution moments, namely,

$$E(X^r) = \frac{\alpha}{\alpha + \beta} E(Y^r) + \frac{\beta}{\alpha + \beta} E(Z^r) = \frac{r\Gamma(r)}{\alpha^r(1 + \beta)} \left[\alpha + (1 + r)\beta\right], \quad r = 1, 2, ..., \tag{5}$$

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

Using the moments above for the EE model, we can compute asymmetry ($\sqrt{\beta_1}$) and kurtosis (β_2) coefficients, which are given by

$$\sqrt{\beta_1} = \frac{2(\alpha + 2\beta)^3 - 12\beta^2(\alpha + \beta)}{(\alpha^2 + 4\alpha\beta + 2\beta^2)^{3/2}} \tag{6}$$

$$\beta_2 = \frac{3(\alpha + 2\beta)^2(3\alpha^2 + 12\alpha\beta + 8\beta^2) - 72\beta^2(\alpha + \beta)^2}{(\alpha^2 + 4\alpha\beta + 2\beta^2)^2} \tag{7}$$

Lemma 1. *Note that as $\beta \to 0$, then $\sqrt{\beta_1} \to 2$ and $\beta_2 \to 9$ which correspond to the asymmetry and kurtosis respectively for the exponential model. General coefficients of asymmetry and kurtosis are such that $\sqrt{2} < \sqrt{\beta_1} \leq 2$ and $6 < \beta_2 \leq 9$, respectively, as shown in Figures 3 and 4.*

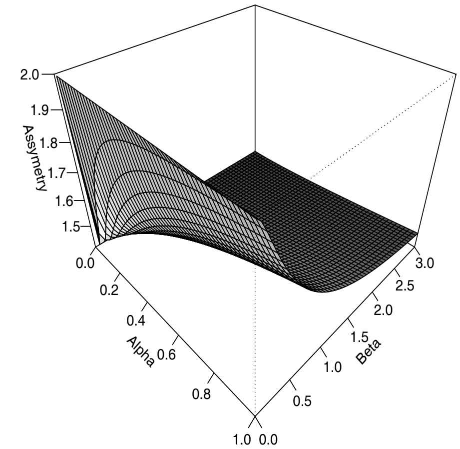

FIGURE 3: Graphs for asymmetry coefficient for the EE model.

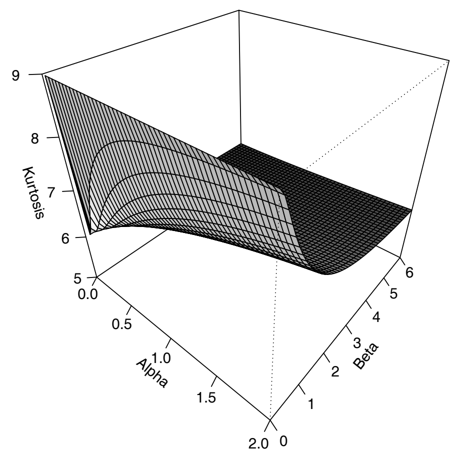

FIGURE 4: Graphs for the kurtosis coefficient for the EE model.

The hazard function for the random variable $X \sim EE(\alpha, \beta)$ is given by

$$h(x) = \frac{f(x;\alpha,\beta)}{1 - F_X(x)} = \frac{\alpha^2(1+\beta x)}{\beta + \alpha(1+\beta x)}$$

i) If $\beta = 0$, then $h(x) = \alpha$, is the hazard function for the exponential model $\forall x \in \mathbb{R}$.

ii) $\forall \beta$, $h(x)$ is monotonically increasing with $h(0) = \frac{\alpha^2}{\alpha+\beta}$.

iii) $\forall \beta$, $h(x) \to \alpha$, as $x \to \infty$.

iv) $h(x)$ is bounded, that is, $\frac{\alpha^2}{\alpha+\beta} < h(x) < \alpha$.

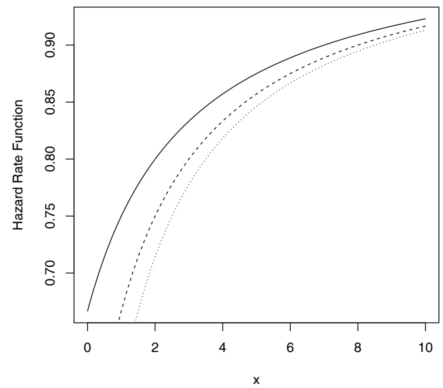

Figure 5: Plots for the hazard function for $\alpha = 1$ and $\beta = 0.5$ (solid line), $\beta = 1$ (dashed line), $\beta = 2$ (dotted line).

The Figure 5 illustrates the behavior of the hazard function for some parameter values.

3. Inferential Considerations

In this section, we consider inference for the EE using moments and the maximum likelihood approach.

3.1. Method of Moments

The moment estimators for the parameters α and β are obtained by solving

$$\begin{aligned} \frac{\alpha + 2\beta}{\alpha(\alpha+\beta)} &= \overline{x} \\ \frac{2\alpha + 6\beta}{\alpha^2(\alpha+\beta)} &= \overline{x}^2 \end{aligned} \tag{8}$$

From the first equation we obtain the moment estimators for $\beta(\widetilde{\beta})$ as a function of the moment estimator for $\alpha(\widetilde{\alpha})$.

$$\widetilde{\beta} = \frac{\widetilde{\alpha}(1 - \widetilde{\alpha}\overline{x})}{\widetilde{\alpha}\overline{x} - 2}, \qquad \widetilde{\alpha} \in \left(\frac{1}{\overline{x}}, \frac{2}{\overline{x}} \right) \tag{9}$$

using (9) and the second equation for the system given in (8) we obtain the moment estimator for α.

$$\widetilde{\alpha} = \frac{2\overline{x} \pm \sqrt{4\overline{x}^2 - 2\overline{x^2}}}{\overline{x^2}} \tag{10}$$

Therefore, $\widetilde{\alpha}$ from (10) replacing α in (9) we obtain $\widetilde{\beta}$. These estimators will be used as initial parts to get the maximum likelihood estimation in the next section.

3.2. Maximum Likelihood

Let $x_1, x_2, \dots, x_n$ a random sample from $X \sim EE(\alpha, \beta)$, so that we obtain the log-likelihood function

$$l(\alpha,\beta) = 2n\log(\alpha) - n\log(\alpha+\beta) - \alpha\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\log(1+\beta x_i) \tag{11}$$

Differentiating the log-likelihood function with respect to α and β, the following equations follow:

$$\frac{\partial l}{\partial \alpha} = \frac{2n}{\alpha} - \frac{n}{\alpha+\beta} - \sum_{i=1}^{n} x_i = 0 \tag{12}$$

$$\frac{\partial l}{\partial \beta} = -\frac{n}{\alpha+\beta} + \sum_{i=1}^{n} \frac{x_i}{1+\beta x_i} = 0 \tag{13}$$

From (12) we obtain

$$\widehat{\beta} = \frac{\widehat{\alpha}(1-\widehat{\alpha}\overline{x})}{\widehat{\alpha}\overline{x}-2}, \qquad \widehat{\alpha} \in \left(\frac{1}{\overline{x}}, \frac{2}{\overline{x}}\right) \tag{14}$$

and the maximum likelihood estimator for α is obtained by resolving numerically the following equation

$$\sum_{i=1}^{n} \frac{x_i}{1-(1-\overline{x}\widehat{\alpha})(\widehat{\alpha}x_i-1)} = \frac{n}{\widehat{\alpha}} \tag{15}$$

The estimator $\widehat{\alpha}$ is the solution to the equation (15), and replacing it in (14) we obtain $\widehat{\beta}$. This algorithm leads to the maximum likelihood estimators for α and β.

4. Real Data Illustration

We consider a data set of the life of fatigue fracture of Kevlar 373/epoxy that are subject to constant pressure at the 90% stress level until all had failed, so we have complete data with the exact times of failure. For previous studies with the data sets see Andrews & Herzberg (1985) and Barlow, Toland & Freeman (1984). These data are:

0.0251,0.0886,0.0891,0.2501,0.3113,0.3451,0.4763,0.5650,0.5671,0.6566,0.6748,0.6751, 0.6753,0.7696,0.8375,0.8391,0.8425,0.8645,0.8851,0.9113,0.9120,0.9836,1.0483,1.0596, 1.0773,1.1733,1.2570,1.2766,1.2985,1.3211,1.3503,1.3551,1.4595,1.4880,1.5728,1.5733, 1.7083,1.7263,1.7460,1.7630,1.7746,1.8275,1.8375,1.8503,1.8808,1.8878,1.8881,1.9316, 1.9558,2.0048,2.0408,2.0903,2.1093,2.1330,2.2100,2.2460,2.2878,2.3203,2.3470,2.3513, 2.4951,2.5260,2.9911,3.0256,3.2678,3.4045,3.4846,3.7433,3.7455,3.9143,4.8073,5.4005, 5.4435,5.5295,6.5541,9.0960.

Using results from Section 3.1, moment estimators were computed leading to the following values: $\widetilde{\alpha} = 0.889$ and $\widetilde{\beta} = 2.563$, which were used as initial estimates for the maximum likelihood approach.

Table 1 presents basic descriptive statistics for data set. We use the notation $\sqrt{b_1}$ and b_2 to represent sample asymmetry and kurtosis coefficients.

Table 1: Descriptive statistics for rupture time.

Data set	n	$\overline{X}$	S	$\sqrt{b_1}$	b_2
Kevlar	76	1.959	1.574	2.019	8.600

Table 2: Parameter estimates for GE, NHE and EE models for the stress-rupture life data set.

Parameter estimates	GE	NHE	EE
α	0.703	0.195	0.954
β	1.709	2.007	6.366
AIC	248.488	253.476	247.3

For comparing model fitting, Akaike (1974), namely

$$AIC = -2 * \hat{\ell}(\cdot) + 2 * k$$

where k is the number of parameters in the model under consideration. The AIC specifies that the model that best fits the data is the one with the smallest AIC value.

Table 2 shows parameter estimators for distributions GE, NHE and EE using maximum likelihood (MLE) approach and the corresponding Akaike information criterion (AIC). For these data, AIC shows a better fit for the EE model. Figure 6 reveals model fitting for the three models, and Figure 7 compares the distribution functions for the three models with the empirical distribution function.

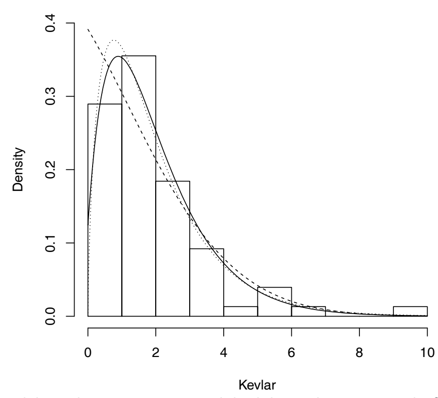

Figure 6: Models fitted by the maximum likelihood approach for the stress-rupture data set: $EE(\hat{\alpha}, \hat{\beta})$ (solid line), $NHE(\hat{\alpha}, \hat{\beta})$ (dashed line) and $GE(\hat{\alpha}, \hat{\beta})$ (dotted line)

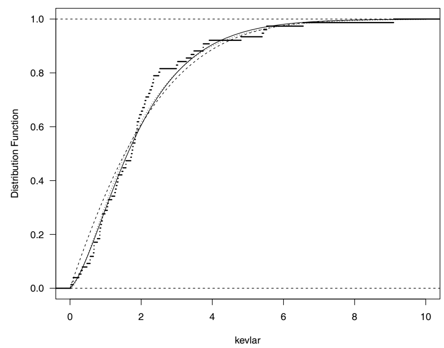

FIGURE 7: Empirical c.d.f. with estimated EE c.d.f. (solid line), estimated NHE c.d.f. (dashed line) and estimated GE c.d.f. (dotted line).

5. Concluding Remarks

This paper introduces a new model positive data. It is shown that the model can be represented as the mixture of two distributions. The scale-exponential distribution can be seen as a particular case of the new model. It is shown that the distribution function, hazard function and moment generating function can be obtained in closed form. Moment estimators are derived and maximum likelihood estimators can be computed using Newton-Raphson type algorithms. The moment estimators can be used as starting values for the maximum likelihood estimators. Asymmetry and kurtosis coefficients are derived and their ranges are plotted. It is illustrated the fact that the model proposed has more flexibility in terms of coefficients of asymmetry and kurtosis. A real data application has demonstrated that the model studied is quite useful for dealing with real data and behaves better in terms of fitting than other models proposed in the literature such as the GE and NHE models.

Acknowledgments

We acknowledge two referees for comments and suggestions that substantially improved the presentation. The research of Yolanda M. Gómez was supported by Becas-Chile of the Chilean government. The research of Heleno Bolfarine was supported by CNPq-Brasil.

[Recibido: enero de 2013 — Aceptado: diciembre de 2013]

References

Akaike, H. (1974), 'A new look at statistical model identification', *IEEE Transaction on Automatic Control* **AC-19**(6), 716–723.

Andrews, D. F. & Herzberg, A. M. (1985), *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, Springer Series in Statistics, New York.

Barlow, R. E., Toland, R. H. & Freeman, T. (1984), A Bayesian analysis of stress-rupture life of kevlar 49/epoxy spherical pressure vessels, *in* 'Proc. Conference on Applications of Statistics', Marcel Dekker, New York.

Gupta, R. D. & Kundu, D. (1999), 'Generalized exponential distributions', *Australian and New Zealand Journal of Statistics* **41**(2), 173–188.

Gupta, R. D. & Kundu, D. (2001), 'Exponentiated exponential family: An alternative to Gamma and Weibull distribution', *Biometrical Journal* **43**(1), 117–130.

Marshall, A. W. & Olkin, I. (2007), *Life Distributions: Structure of Nonparametric, Semiparametric and Parametric Families.*

Nadarajah, S. & Haghighi, F. (2011), 'An extension of the exponential distribution', *Statistics: A Journal of Theoretical and Applied Statistics* **45**(6), 543–558.