



# Ristić-Balakrishnan extended exponential distribution

Frank Gomes-Silva<sup>1\*</sup>, Thiago Alexandro Nascimento de Andrade<sup>2</sup> and Marcelo Bourguignon<sup>3</sup>

<sup>1</sup>Departamento de Informática e Estatística, Universidade Federal Rural de Pernambuco, Rua Manuel de Medeiros, s/n., 52171-900, Recife, Pernambuco, Brazil. <sup>2</sup>Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil. <sup>3</sup>Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil. \*Author for correspondence. E-mail: franksinatrags@gmail.com

**ABSTRACT.** In this paper, we introduce and study a new generalization of the extended exponential distribution, called the Ristić-Balakrishnan extended exponential distribution. The new model adds one parameter in the baseline model and its failure rate function can accommodate both inverted bathtub and bathtub shapes. Important distributions are obtained as a special case of our model, such as exponential and Lindley distributions. The main purpose is to define a new flexible distribution with great power adjustment to survival data sets. For this reason, we provide a comprehensive mathematical treatment of the new model. Furthermore, we use a real data set that proves empirically the power of adjustment of the new distribution compared to other competitive models in the literature.

**Keywords:** generalized distribution; statistical properties; quantile function; maximum likelihood estimation; model fit.

## Distribuição Ristić-Balakrishnan exponencial estendida

**RESUMO.** Neste artigo introduzimos e estudamos uma nova generalização da distribuição exponencial estendida, denominada Ristić-Balakrishnan exponencial estendida. O novo modelo adiciona um parâmetro a distribuição-base e sua função taxa de falha pode acomodar as formas de banheira e banheira invertida. Importantes distribuições são obtidas como caso especial de nosso modelo, tais como a Lindley e a exponencial. O principal objetivo é definir uma nova distribuição flexível com alto poder de ajuste a dados de sobrevivência. Por este motivo, estabelecemos um tratamento matemático abrangente ao novo modelo. Além disso, utilizamos um conjunto de dados reais que comprovam empiricamente o poder de ajuste da nova distribuição em comparação a outros modelos competitivos na literatura.

**Palavras-chave:** distribuição generalizada; propriedades estatísticas; função quantílica; estimação por máxima verossimilhança; ajuste de modelos.

## Introduction

It is hardly necessary to emphasize that a probabilistic model is commonly employed to attack practical situations in which a deterministic model is not feasible. This definition, albeit implicitly, had already been part of common sense since the Renaissance era, in which the notion of probability was unconsciously employed to propose solutions in games of chance (Bernstein, 1996). In fact, this intrinsic sense of probability lies at the heart of scientific methodology. Here it is worth quoting the classic book 'The logic of scientific discovery' by Sir Karl Popper:

The most important application of the theory of probability is to what we may call chance-like or random events, or occurrences. These seem to be characterized by a peculiar kind of incalculability which makes one disposed to believe after many unsuccessful attempts that all known rational methods of prediction must fail in their case. We

have, as it were, the feeling that not a scientist but only a prophet could predict them. And yet, it is just this incalculability that makes us conclude that the calculus of probability can be applied to these events (Popper, 1959, p. 167).

Taking a leap forward in time, we see that probabilistic models still arouse the fascination of applied scholars and researchers. This interest materializes in the great amount of works that are dedicated to the proposal of new distributions. In particular, those dealing with distribution generators. Our research presented below is related to the generalization of probabilistic models through generators of distributions. In the generator approach, we refer to the following papers: Marshall and Olkin (1997) for the 'Marshall-Olkin' class; Eugene, Lee, and Famoye (2002) for the 'beta' class; Zografos and Balakrishnan (2009) for the 'Gamma' class; Cordeiro and Castro (2011) for the 'Kumaraswamy' class and Cordeiro, Ortega, and

Cunha (2013) defined the ‘exponentiated generalized’ class of distributions.

Recently, Gómez, Bolfarine, and Gómez (2014) introduced a new *extended exponential* (EE for short) distribution. For  $x > 0$ , its cumulative density function (cdf) and probability density function (pdf) are given by Equation 1 and 2:

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

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

where:

$\alpha > 0$  and  $\beta \geq 0$ . Several mathematical properties of the EE distribution, including expectation, variance, moment generating function (mgf), asymmetry and kurtosis coefficients, among others, were studied by Gómez et al. (2014). In particular, they proved that the density of the EE model is a mixture of the exponential and gamma densities.

We believe that the addition of parameters to the EE model may generate new distributions with great adjustment capability and, for this reason, we propose a generalization of it. On the other hand, Ristić and Balakrishnan (2012) defined the ‘Ristić-Balakrishnan’ –G (RB –G for short) family for  $x \in \mathbb{R}$  and  $a > 0$  having, respectively, pdf and cdf given by Equation 3 and 4:

$$f(x; a) = \frac{1}{\Gamma(a)} \{-\log[G(x; \xi)]\}^{a-1} g(x; \xi) \quad (3)$$

$$F(x; a) = 1 - \frac{\gamma(a, -\log[G(x; \xi)])}{\Gamma(a)} \quad (4)$$

where:  $g(x, \xi) = dG(x, \xi)$ , with  $\xi$  a parametric vector,  $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$  is the gamma function and  $\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt$  denotes the lower incomplete gamma function. The main motivation for this family is that, for  $a = n \in \mathbb{N}$ , Equation 3 is the pdf of the  $n$ th lower record value of a sequence independent and identically distributed variables from a population with density  $g(x, \xi)$ .

In this paper we propose a new lifetime model called ‘Ristić-Balakrishnan extended exponential’ (RBEE) distribution by taking Equation 1 in 4. As we will see later, the proposed model is quite flexible and its failure rate function can accommodate both inverted bathtub and bathtub

shapes, which are important for reliability, life time, biological and medical sciences, among others. In addition, the new density may be expressed as a mixture of ‘Erlang’ densities. Thus, many properties can be derived using this simple representation. As will also be clear later, many important distributions are obtained as a special case of our model. Finally, we prove the new model is very superior in terms of adjustment to real data, when compared to the base model and other important models well established in the literature.

## Material and methods

### The RBEE distribution

Let  $X$  be a random variable with support on the positive real line having the RBEE distribution, say  $X \sim RBEE(a, \alpha, \beta)$ . The cdf of  $X$  is defined by inserting Equation 1 in Equation 4, according Equation 5:

$$F(x) = F(x; a, \alpha, \beta) = \frac{1}{\Gamma(a)} \{\Gamma(a) - \gamma[a, \rho(x)]\} \quad (5)$$

where:

$$\rho(x) = \log(\alpha + \beta) - \log[\alpha + \beta - (\alpha + \beta + \alpha\beta x)e^{-\alpha x}].$$

The density of  $X$ , for  $x > 0$ , can be reduced to Equation 6:

$$f(x) = f(x; a, \alpha, \beta) = \frac{\alpha^2(1 + \beta x)e^{-\alpha x}}{(\alpha + \beta)\Gamma(a)} \rho^{a-1}(x). \quad (6)$$

We write  $F(x) = F(x; a, \alpha, \beta)$  in order to eliminate the dependence on the model parameters. Clearly, the EE model is a special case of Equation 5 when  $a = 1$ . The exponential and Lindley distributions arise as special cases when  $\beta = 0$  and  $\beta = 1$ , respectively, in addition to  $a = 1$ . If  $\beta = 0$ ,  $\beta = 1$  and  $a \neq 1$ , we obtain the RB-exponential and RB-Lindley respectively.

Some plots of the pdf Equation 6 are displayed in Figure 1. These plots reveal that the RBEE pdf is quite flexible and can take various forms reinforcing the importance of the proposed model.

The survival function is Equation 7:

$$S(x) = \frac{\gamma[a, \rho(x)]}{\Gamma(a)}. \quad (7)$$

The hazard rate function (hrf) and reversed hazard rate function (rhrrf) of  $X$  are given by Equation 8 and 9:

$$h(x) = \frac{\alpha^2(1+\beta x)e^{-\alpha x}\rho^{a-1}(x)}{(\alpha+\beta)\gamma[a, \rho(x)]} \quad (8)$$

$$\tau(x) = \frac{\alpha^2(1+\beta x)e^{-\alpha x}\rho^{a-1}(x)}{(\alpha+\beta)\{\Gamma(a) - \gamma[a, \rho(x)]\}} \quad (9)$$

respectively. Some plots of the hrf Equation 8 are displayed in Figure 2. Non-monotone forms such as bathtub and inverted bathtub are particularly important because of its great practical applicability. For example, the time of human life is one phenomena that the bathtub form is applicable.

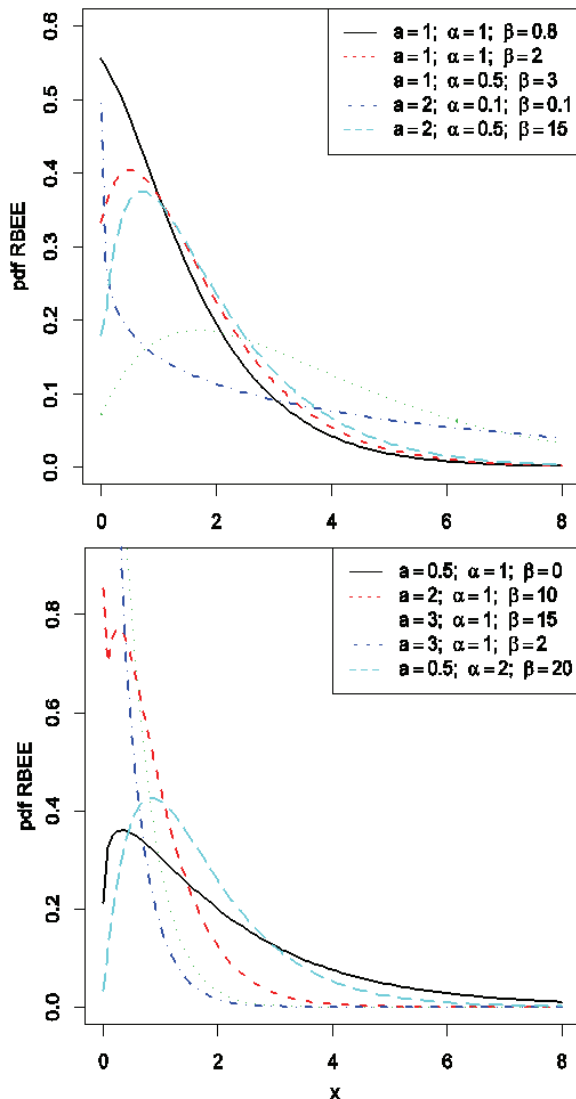


Figure 1. Plots of the R BEE density function for some parameter values.

#### Asymptotic and shapes of the RBEE

For a detailed mathematical approach for the RBEE model, we investigate the shapes of its pdf and hrf using their first and second derivatives.

The first derivatives of  $\log\{f(x)\}$  and  $\log\{h(x)\}$  for the RBEE model are given by Equation 10 and 11:

$$\frac{d \log\{f(x)\}}{dx} = -\alpha + \frac{\beta}{1+\beta x} - \frac{(a-1)\alpha^2(1+\beta x)e^{-\alpha x}}{\varphi(x)\rho(x)} \quad (10)$$

$$\frac{d \log\{h(x)\}}{dx} = -\alpha + \frac{\beta}{1+\beta x} - \alpha^2(1+\beta x)e^{-\alpha x} \left[ \frac{a-1}{\varphi(x)\rho(x)} - \frac{\rho^{a-1}(x)}{(\alpha+\beta)\gamma[a, \rho(x)]} \right] \quad (11)$$

with  $\varphi(x) = \alpha + \beta - (\alpha + \beta + \alpha\beta x)e^{-\alpha x}$ .

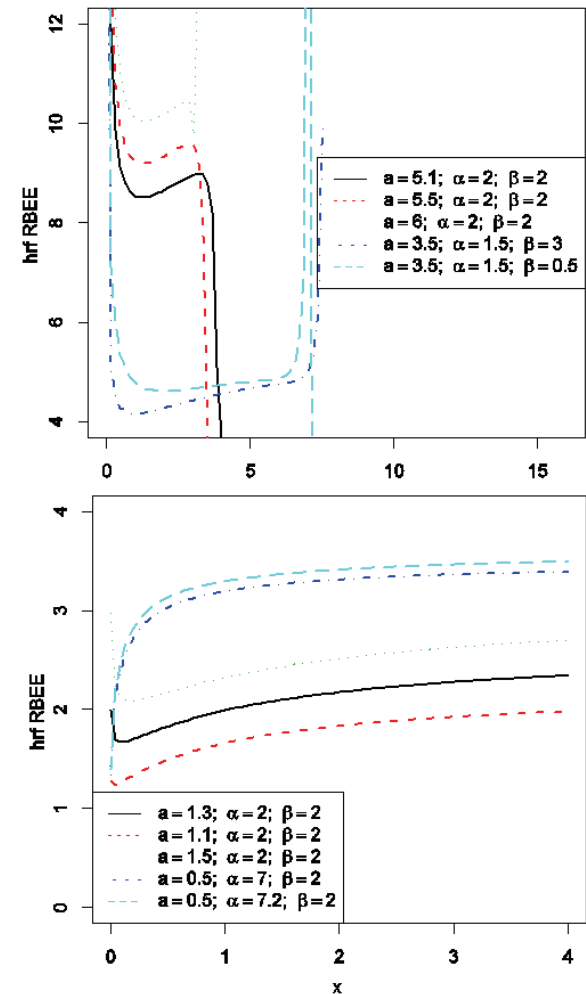


Figure 2. Plots of the R BEE hazard function for some parameter values.

Hence, the critical values of  $f(x)$  and  $h(x)$  are the roots of the Equation 12 and 13:

$$\frac{(a-1)\alpha^2(1+\beta x)e^{-\alpha x}}{\varphi(x)\rho(x)} = -\alpha + \frac{\beta}{1+\beta x} \quad (12)$$

$$\alpha^2(1+\beta x)e^{-\alpha x} \left[ \frac{a-1}{\varphi(x)\rho(x)} - \frac{\rho^{a-1}(x)}{(\alpha+\beta)\gamma[a, \rho(x)]} \right] = -\alpha + \frac{\beta}{1+\beta x} \quad (13)$$

respectively. The values  $x_0$  and  $x'_0$  which solves the Equations 12 and 13 above can be a maximum,

minimum or inflection point. To check this, we evaluate the signs of the second derivatives of  $\log \{f(x)\}$  and  $\log \{h(x)\}$ , respectively, at  $x = x_0$  and  $x = x'_0$ . We have Equation 14 and 15:

$$\frac{d^2 \log \{f(x)\}}{dx^2} = -\frac{\beta^2}{(1+\beta x)^2} + \left\{ \alpha - \beta + \alpha \beta x \alpha^2 (1+\beta x)^2 [\varphi^{-1}(x) - \rho^{-1}(x)] \right\} \frac{(a-1)\alpha^2 e^{-ax}}{\varphi(x)\rho(x)} \quad (14)$$

$$\frac{d^2 \log \{h(x)\}}{dx^2} = -\frac{\beta^2}{(1+\beta x)^2} + \left\{ \alpha - \beta + \alpha \beta x + [1 - \rho^{-1}(x)] \frac{\alpha^2 (1+\beta x)^2}{\varphi(x)} \right\} + \frac{\alpha^2 e^{-ax} \rho^{-1}(x)}{(\alpha + \beta) \gamma(a, \rho(x))} \quad (15)$$

$$\times \left\{ (a-1)(\alpha - \beta + \alpha \beta x) - \alpha^2 (1+\beta x)^2 e^{-ax} \rho^{-1}(x) \left[ \frac{(a-1)\varphi^{-1}(x) + \rho^{-1}(x)}{(\alpha + \beta) \gamma(a, \rho(x))} \right] \right\}.$$

It is common to obtain numerical solutions with high accuracy through optimization routines in most mathematical and statistical platforms.

### Quantile function

For many practical purposes, it is important to make explicit the quantile function (qf) of  $X$ . The RBEE qf, say  $q(u)$  can be obtained by inverting Equation 5 (for  $0 < u < 1$ ) as Equation 16:

$$q(u) = -\frac{1}{\beta} \frac{1}{\alpha} \left[ 1 + W \left\{ -\frac{1}{\beta} (\alpha + \beta) (1 - \exp[-Q^{-1}(a, u)]) \exp[-(\alpha + \beta)/\beta] \right\} \right] \quad (16)$$

where:

$z = Q^{-1}$  is the inverse function of  $Q(a, z) = 1 - \frac{\gamma(a, z)}{\Gamma(a)}$  and  $W(\cdot)$  denotes the Lambert

W-function. In a recent paper, Nadarajah, Bakouch, and Tahmasbi (2011) used the Lambert W-function to derive the qf of the exponentiated Lindley distribution. For any complex  $t$ , the Lambert W-function is defined as the inverse of the function  $g(t) = te^t$ . For more details, see <http://mathworld.wolfram.com/LambertW-Function.html>. An

implementation in R software is available through the 'LambertW' package. See <http://cran.r-project.org/web/packages/LambertW/LambertW.pdf>. In the 'Mathematica platform', the 'LambertW' is available through the function 'ProductLog[z]', which gives the principal solution for  $W$  in  $z = we^w$ . By using the Lagrange inversion theorem, we can write an expansion for the qf of  $X$  as follows Equation 17:

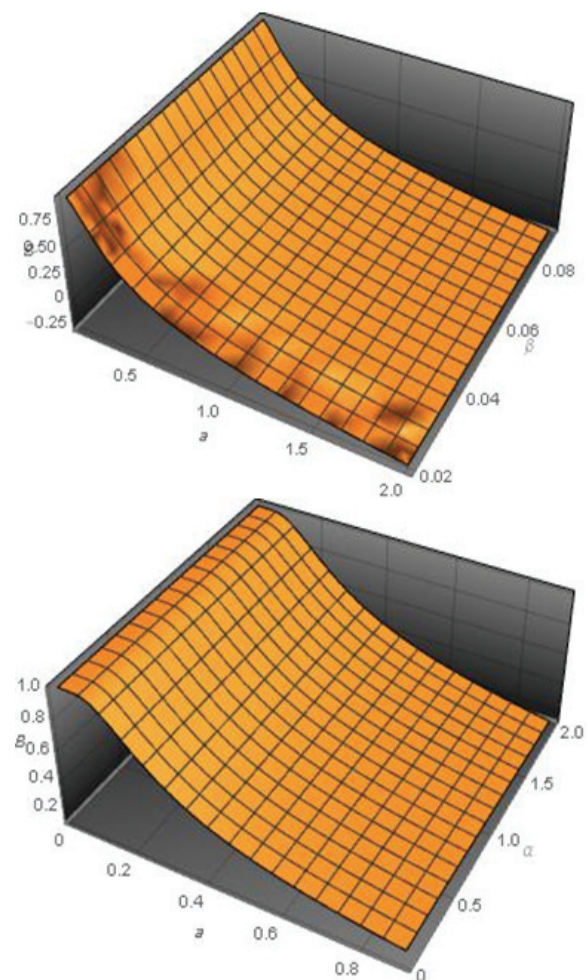
$$q(u) = -\frac{1}{\beta} \frac{1}{\alpha} \left[ 1 + \sum_{k=1}^{\infty} \frac{(-1)^k (-k)^{k-1}}{\beta^k k!} (\alpha + \beta)^k (1 - \exp[-Q^{-1}(a, u)])^k \exp \left[ -\frac{k(\alpha + \beta)}{\beta} \right] \right]. \quad (17)$$

Note that the above equation can be easily implemented in computational platforms that have numerical elementary routines.

The applications of qf are diverse and include: calculation of the moments, estimation of

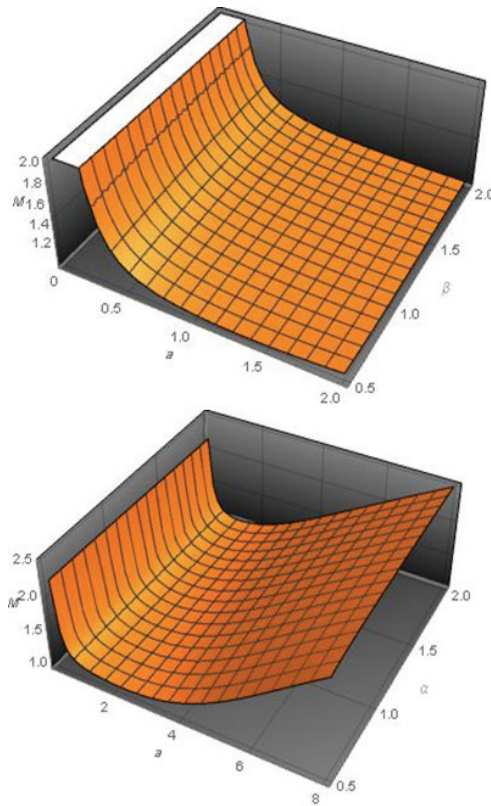
parameters, simulations, calculation of asymmetry and kurtosis measurements, among others. For illustration, we use the qf of  $X$  to determine the Bowley skewness (Kenney & Keeping, 1962) (B) and Moors kurtosis (Moors, 1988) (M). The Bowley skewness is based on quartiles  $B = [Q(3/4) - 2Q(1/2) + Q(1/4)] / [Q(3/4) - Q(1/4)]$ , whereas the Moors kurtosis is based on octiles  $M = [Q(7/8) - Q(5/8) - Q(3/8) + Q(1/8)] / [Q(6/8) - Q(2/8)]$ . These two measures are less sensitive to outliers and they exist even for distributions without moments.

In Figure 3 and 4, we present 3D plots of B and M measures for several parameters values. These plots are obtained using the 'Wolfram Mathematica' software. Based on these plots, it is possible to conclude that changes in the additional parameter  $a$  have a considerable impact on the skewness and kurtosis of the RBEE distribution, thus showing its greater flexibility.



**Figure 3.** Plots of the Bowley skewness for the RBEE distribution for some parameter values.





**Figure 4.** Plots of the Moors kurtosis for the R BEE distribution for some parameter values.

### Properties

#### A useful representation

We provide useful linear representations for Equation 5 and 6 based on the exponentiated class of distributions. Mathematical properties of the exponentiated distributions have been published by many authors in the 90s and more recently. See, for example, Gupta and Kundu (1999) for exponentiated exponential, Nadarajah et al. (2011) for exponentiated Lindley, Sarhan and Kundu (2009) for exponentiated linear failure rate and, more recently, Lemonte (2013) for the exponentiated Nadarajah-Haghighi distributions. For an arbitrary baseline cdf  $G(x)$  a random variable  $Y_a$  has the exp-G class with power parameter  $a > 0$  say  $Y_a \sim \text{exp-G}(a)$ , if its cdf and pdf are given by  $H_a(x) = G(x)^a$  and  $h_a(x) = ag(x)G(x)^{a-1}$  respectively. For a comprehensive discussion about the exponentiated class, see a recent paper by Tahir and Nadarajah (2015).

By using results presented in Cordeiro and Bourguignon (2016) we can be expressed the pdf  $f(x)$  as Equation 18:

$$f(x) = \sum_{j=0}^{\infty} w_j h_{j+1}(x) \quad (18)$$

With

$$w_j = \frac{(-1)^j}{(j+1)\Gamma(a)} \left[ \binom{a-1}{j} + (a-1) \sum_{i=0}^{\infty} d_i(a-1) \binom{a+i}{j} \right] \quad (19)$$

where:

quantities  $d_i(a-1)$  (for  $i \geq 0$ ) determined by  $d_0(c) = c/2$ ,  $d_1(c) = c(3c+5)/24$ ,  $d_2(c) = c(c^2+5c+6)/48$ ,  $d_3(c) = c(15c^3+150c^2+485c+502)/5760$ , etc.

Note that, by integrating Equation 18, we can express  $F(x)$  as Equation 20:

$$F(x) = \sum_{j=0}^{\infty} w_j H_{j+1}(x) \quad (20)$$

where:

$H_{j+1}(x)$  denotes the exp-EE cumulative distribution with power parameter  $j+1$ . Here,  $h_{j+1}(x)$  is the exp-EE density function with power parameter  $j+1$ , and is given by (for  $j \geq 0$ )

$$h_{j+1}(x) = \sum_{k=0}^j \sum_{\ell=0}^k \sum_{i=0}^{\ell+1} \frac{(j+1)(-1)^k \alpha^{\ell+2} \beta^{i+k-\ell}}{(\alpha+\beta)^{k+1}} \binom{j}{k} \binom{k}{\ell} \binom{\ell+1}{i} x^i e^{-\alpha(k+1)x}.$$

By interchanging  $\sum_{\ell=0}^k \sum_{i=0}^{\ell+1}$  by  $\sum_{i=0}^{k+1} \sum_{\ell=\delta_i}^k$  in the last

equation, where Equation 21:

$$\delta_i = \begin{cases} 0, & \text{if } i = 0, 1 \\ i-1, & \text{if } i \geq 2, \end{cases} \quad (21)$$

and, after a simple algebraic manipulation, we obtain Equation 22 and 23:

$$h_{j+1}(x) = \sum_{k=0}^j \sum_{i=0}^{k+1} p_{k,i}^{(j+1)} \pi(x; i+1, (k+1)\alpha) \quad (22)$$

$$p_{k,i}^{(j+1)} = \sum_{\ell=\delta_i}^k \frac{(j+1)(-1)^k \alpha^{\ell+1} \beta^{i+k-\ell} i!}{(k+1)^{i+1} (\alpha+\beta)^{k+1}} \binom{j}{k} \binom{k}{\ell} \binom{\ell+1}{i} \quad (23)$$

where:

$\pi(x; i+1, (k+1)\alpha)$  denotes the pdf of the Erlang distribution with parameters  $i+1$  (for  $i \geq 0$ ) and  $(k+1)\alpha$ . If  $Z$  is an Erlang random variable with parameters  $s(=1, 2, 3, \dots)$  and  $\lambda > 0$ , its pdf is given

by  $\pi(z; s, \lambda) = \lambda^s z^{s-1} e^{-\lambda z} / (s-1)!$ . Changing  $\sum_{j=0}^{\infty} \sum_{k=0}^j$  by

$\sum_{k=0}^{\infty} \sum_{j=k}^{\infty}$ , the density function reduces to Equation 24:

$$f(x) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} v_{k,i} \pi(x; i+1, (k+1)\alpha) \quad (24)$$

where  $v_{k,i} = \sum_{j=k}^{\infty} w_{j+1} p_{k,i}^{(j+1)}$ . Equation 24 is the main result of this section.

### Moments, incomplete moments and generating function

Then, the  $n$ th moment of  $X$  and its incomplete moments, respectively, are given by Equation 25 and 26:

$$E(X^n) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \frac{(n+i)! v_{k,i}}{i! [(k+1)\alpha]^n} \quad (25)$$

$$T_n(z) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \frac{\{\Gamma(n+i+1) - \Gamma[n+i+1, z(k+1)\alpha]\} v_{k,i}}{i! [(k+1)\alpha]^n} \quad (26)$$

where  $\Gamma(a, z) = \int_z^{\infty} t^{a-1} e^{-t} dt$  denotes the upper incomplete gamma function.

The moment generating function (mgf) of  $X$  can be determined from Equation 24, according Equation 27:

$$M(t) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} v_{k,i} \int_0^{\infty} e^{tx} \pi(x; i+1, (k+1)\alpha) dx. \quad (27)$$

Then, for all  $t < (k+1)\alpha$ , we have Equation 28:

$$M(t) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \left[ 1 - \frac{t}{(k+1)\alpha} \right]^{-(i+1)} v_{k,i}. \quad (28)$$

### Order statistics

By using results presented in Cordeiro and Bourguignon (2016) the density function  $f_{i:n}(x)$  of the  $i$ th order statistic, say  $X_{i:n}$ , for  $i=1, \dots, n$ , from a random sample  $X_1, \dots, X_n$  having the RBEE distribution, can be expressed as Equation 29:

$$f_{i:n}(x) = \sum_{j=0}^{n-i} \sum_{r,k=0}^{\infty} m_{j,r,k} h_{j+i+r+k}(x) \quad (29)$$

with

$$m_{j,r,k} = \frac{(-1)^j n!}{(i-1)!(n-i-j)!j!} \frac{(r+1)w_r f_{j+i-1,k}}{(j+i+r+k)} \quad (30)$$

where  $w_r$  is defined by Equation 19, and

$$f_{j+i-1,k} = (kd_0)^{-1} \sum_{m=1}^k [m(j+1) - k] w_m f_{j+i-1,k-m}.$$

### Reliability

In reliability the stress-strength model describes the life of a component which has a random strength  $X_1$  that is subjected to a random stress  $X_2$ . The component fails at the instant that the stress applied to it exceeds the strength, and the component will function satisfactorily whenever

$X_1 > X_2$ . Hence,  $R = P(X_2 < X_1)$  is a measure of component reliability. When  $X_1$  and  $X_2$  have independent RBEE( $a_1, \alpha, \beta$ ) and RBEE( $a_2, \alpha, \beta$ ) models the reliability is defined by  $R = \int_0^{\infty} f_1(x) F_2(x) dx$ . The pdf of  $X_1$  and cdf of  $X_2$  are expressed from Equations 18 and 19 as Equation 31:

$$f_1 = \sum_{j=0}^{\infty} w_{j,i}(a_1) h_{j+1}(x) \quad \text{and} \quad F_2 = \sum_{k=0}^{\infty} w_{k,n}(a_2) H_{j+1}(x) \quad (31)$$

where (for  $s = j, k; p = i, n$  and  $q = 1, 2$ )

$$w_{s,p}(a_q) = \frac{(-1)^s}{(s+1)\Gamma(a_q)} \left[ \binom{a_q-1}{s} + (a_q-1) \sum_{p=0}^{\infty} d_p (a_q-1) \binom{a_q+p}{s} \right] \quad (32)$$

Thus, we have Equation 33:

$$R = \sum_{j,k=0}^{\infty} I_{j,k} \int_0^{\infty} h_{j+1}(x) H_{k+1}(x) dx \quad (33)$$

where  $I_{j,k} = w_{j,i}(a_1) w_{k,n}(a_2)$ .

Hence, after a simple algebraic manipulation, the reliability of the RBEE distribution is given by Equation 34:

$$R = \sum_{j,k=0}^{\infty} \sum_{l=0}^{j+k+1} \sum_{r=0}^l \sum_{m=0}^{r+1} \frac{(-1)^l (j+1) \alpha^{-m+1} \beta^{l+m-r} m!}{(l+1)^{m+1} (\alpha+\beta)^{l+1}} \binom{j+k+1}{l} \binom{l}{r} \binom{r+1}{m} I_{j,k}. \quad (34)$$

### Entropy

The Rényi entropy is defined (for  $\delta > 0$  and  $\delta \neq 0$ ), according Equation 35:

$$J_R(\delta) = \frac{1}{1-\delta} \log \left\{ \int_0^{\infty} f^{\delta}(x) dx \right\}. \quad (35)$$

Let  $\bar{G}(x) = 1 - G(x)$  be the baseline survival function. Following similar idea given in Nadarajah, Cordeiro, and Ortega (2015) (Section 10), we have Equation 36:

$$\{-\log[1 - \bar{G}(x)]\}^{\delta a - \delta} = (\delta a - \delta) \sum_{k=0}^{\infty} \binom{k - \delta a + \delta}{k} \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k}{j} p_{j,k}}{[\delta(a-1) - j]} \bar{G}(x)^{[\delta(a-1) + k]} \quad (36)$$

where:

The constants  $p_{j,k}$  can be calculated recursively by Equation 37:

$$p_{j,k} = k^{-1} \sum_{m=1}^k [k - m(j+1)] c_m p_{j,k-m} \quad (37)$$

for  $k=1, 2, \dots, p_{j,0}=1$  and  $c_k=(-1)^{k+1}(k+1)^{-1}$ . By using Equation 36 and generalized binomial expansion we obtain the Rényi entropy for the family as Equation 38 and 39:

$$J_R(\delta) = \frac{1}{1-\delta} \log \left[ \frac{\delta(a-1)}{\Gamma^\delta(a)} \right] + \frac{1}{1-\delta} \log \left\{ \sum_{i,k=0}^{\infty} \binom{k-\delta(a-1)}{k} \binom{k+\delta(a-1)}{i} \sum_{j=0}^k \frac{(-1)^{i+j+k} \binom{k}{j} p_{j,k}}{[\delta(a-1)-j]} I_i \right\} \quad (38)$$

where,

$$I_i = \int_0^{\infty} G^i(x) g^\delta(x) dx \quad (39)$$

comes from the baseline distribution. Based on the cdf Equation 1 and pdf Equation 2, we can express Equation 39 as Equation 40:

$$I_i = \frac{(-1)^i \alpha^{2\delta+i}}{(\alpha+\beta)^{i+\delta}} e^{\frac{\alpha(i+\delta)}{\beta}} \beta^{-1+i} E \left[ -\delta, \frac{\alpha(i+\delta)}{\beta} \right] \quad (40)$$

where  $E[n, z] = \int_1^{\infty} t^{-n} e^{-zt} dt$  is 'exponential integral function'.

### Estimation and inference

The maximum likelihood method is the one that stands out most among the estimation methods admitting good asymptotic properties. The maximum likelihood estimators (MLEs) can be used when constructing confidence intervals and regions and also in test statistics. Let  $x_1, \dots, x_n$  be a random sample of size  $n$  from the RBEE  $(a, \alpha, \beta)$  model. The log-likelihood function for the vector of parameters  $\Theta = (\alpha, \alpha, \beta)^T$  can be expressed from Equation 41:

$$\ell(\Theta) = n \log \left[ \frac{\alpha^2}{(\alpha+\beta)\Gamma(a)} \right] - \alpha \sum_{i=1}^n x_i + \sum_{i=1}^n \log(1+\beta x_i) + (a-1) \sum_{i=1}^n \log[\rho(x_i)] \quad (41)$$

where:

$$\rho(x_i) = \log(\alpha+\beta) - \log[\alpha+\beta - (\alpha+\beta+\alpha\beta x_i)e^{-\alpha x_i}].$$

The components of the score vector  $U(\Theta)$  are given by Equation 42, 43 and 44:

$$U_a(\Theta) = -n\psi(a) + \sum_{i=1}^n \phi(x_i) \quad (42)$$

$$U_\alpha(\Theta) = \frac{n(\alpha+2\beta)}{\alpha(\alpha+\beta)} - \sum_{i=1}^n x_i + (a-1) \sum_{i=1}^n \frac{1}{\rho(x_i)} \left\{ \frac{1}{\alpha+\beta} - \frac{1-(1-\alpha x_i - \alpha\beta x_i^2)e^{-\alpha x_i}}{\phi(x_i)} \right\} \quad (43)$$

$$U_\beta(\Theta) = -\frac{n}{\alpha+\beta} + \sum_{i=1}^n \frac{x_i}{1+\beta x_i} + (a-1) \sum_{i=1}^n \frac{1}{\rho(x_i)} \left\{ \frac{1}{\alpha+\beta} - \frac{1-(1+\alpha x_i)e^{-\alpha x_i}}{\phi(x_i)} \right\} \quad (44)$$

where

$\psi(\cdot)$  is the digamma function and  $\phi(x_i) = \alpha+\beta - (\alpha+\beta+\alpha\beta x_i)e^{-\alpha x_i}$ .

The information matrix is given by  $J(\Theta) = \{-U_{rs}\}$  and its elements  $U_{rs}(\Theta) = \partial^2 \ell(\Theta) / \partial r \partial s$  for  $r, s \in \{a, \alpha, \beta\}$  can be obtained from the authors upon request.

## Results and discussion

### Application

We consider an uncensored data set from Murthy, Xie and Jiang (2004) (page 180) used in the industry, representing the failure time (in weeks) of 50 components put into use at time. The data are: 0.013, 0.065, 0.111, 0.111, 0.163, 0.309, 0.426, 0.535, 0.684, 0.747, 0.997, 1.284, 1.304, 1.647, 1.829, 2.336, 2.838, 3.269, 3.977, 3.981, 4.520, 4.789, 4.849, 5.202, 5.291, 5.349, 5.911, 6.018, 6.427, 6.456, 6.572, 7.023, 7.087, 7.291, 7.787, 8.596, 9.388, 10.261, 10.713, 11.658, 13.006, 13.388, 13.842, 17.152, 17.283, 19.418, 23.471, 24.777, 32.795, 48.105. Table 1 provides some descriptive statistics.

**Table 1.** Descriptive statistics for number of successive failure times of 50 components.

Statistics	n	Mean	Median	Variance	Minimum	Maximum
	50	7.821	5.320	84.76	0.013	48.100

For Murthy et al. (2004)'s data, we compared the RBEE model with the EE Gómez et al. (2014) and Lindley sub-models and other commonly used models in survival analysis, namely the log-logistic, Fréchet and Birnbaum-Saunders (BS) distributions. The densities of these models are given, for example, in the Wolfram alpha website (<https://www.wolframalpha.com>).

Table 2 gives the MLEs of the fitted models to the current data with their corresponding standard errors, in addition to the AIC, BIC and CAIC statistics. Table 3 lists the values of the  $A^*$  and  $W^*$  statistics. In general, it is considered that lower values of these criteria fit better the data.

**Table 2.** MLEs (and the corresponding standard errors in parentheses), AIC, BIC and CAIC statistics for number of successive failures for the air conditioning system.

Distributions	$\hat{\alpha}$	$\hat{\beta}$	$\hat{a}$	AIC	BIC	CAIC
RBEE	0.02792 (0.02486)	0.4257 (0.6573)	4.5200 (2.2804)	306.0	311.7	306.5
EE	0.1279 (0.0352)	2.338E-7 (0.0302)	1 (----	309.7	313.5	309.9
Lindley	0.2317 (0.590)	1 (----	1 (----	324.6	326.5	324.6
	$\hat{\alpha}$	$\hat{\beta}$				
Log-logistic	4.0938 (0.9218)	1.0834 (0.1304)		316.0	319.8	316.3
	$\hat{\sigma}$	$\hat{\lambda}$				
Fréchet	1.2802 (0.4028)	0.4791 (0.04541)		341.3	345.1	341.5
	$\hat{\alpha}$	$\hat{\beta}$				
BS	2.7621 (0.2973)	1.2576 (0.2721)		327.4	331.2	327.7

**Table 3.** Mean, Variance, Skewness and Kurtosis for the three main distributions.

Distributions	Mean	Variance	Skewness	kurtosis
R BEE	7.82154	82.7757	-0.23734	1.18290
EE( $a=1$ )	7.81862	61.1309	----	----
Lindley	7.81997	36.5953	0.17305	1.13473

Additionally, we took into consideration the Anderson-Darling ( $A^*$ ) and Cramér-von Mises ( $W^*$ ) statistics (Chen & Balakrishnan, 1995). Chen and Balakrishnan (1995) proposed a general approximate goodness-of-fit test for the hypothesis  $H_0: X_1, \dots, X_n$  with  $X_i$  following  $F(x; \theta)$ , where the form of  $F$  is known but the  $p$ -vector  $\theta$  is unknown. To obtain the statistics  $A^*$  and  $W^*$ , we can proceed as follows: (1) compute  $r_i = F(x_i; \theta)$ , where  $x_i$ 's are in ascending order, and then  $y_i = \Phi^{-1}(r_i)$ , where  $\Phi(\cdot)$  is the standard normal cumulative distribution; (2) compute  $u_i = \Phi\left(\frac{y_i - \bar{y}}{s}\right)$ , where  $\bar{y} = (1/n) \sum_{i=1}^n y_i$  and  $s^2 = (n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$ ; (3) calculate  $A^2 = -n - (1/n) \sum_{i=1}^n [(2i-1) \log(u_i) + (2n+1-2i) \log(1-u_i)]$  and  $W^2 = \sum_{i=1}^n [u_i - (2i-1)/(2n)]^2 + 1/(12n)$  and then  $A^* = A^2(1 + 0.75/n + 2.25/n^2)$  and  $W^* = W^2(1 + 0.5/n)$ . Table 3 lists the values of the  $A^*$  and  $W^*$  statistics. In general, it is considered that lower values of these criteria fit the data better.

Table 3 presents the mean, variance, asymmetry and kurtosis for the RBEE, EE and Lindley adjusted models. As we can see, the empirical and estimated means and variances do not differ considerably. This shows that the models are adequate to explain this data.

The figures in Table 2 and 4 reveals that the RBEE model has the lowest AIC, BIC, CAIC,  $A^*$  and  $W^*$  values among all fitted models. Thus, the proposed RBEE distribution is the best model to explain these data. Finally, Figure 5 displays the histogram of the data and the estimated pdf and cdf of the RBEE model. These plots reveal that the proposed model is quite suitable for these data.

**Table 4.** Goodness-of-fit tests.

Models	Statistics	
	$W^*$	$A^*$
RBEE	0.0539	0.2709
EE	0.0658	0.3295
Lindley	0.0657	0.3284
Log-logistic	0.2572	1.3816
Fréchet	0.6097	3.3138
BS	0.2794	1.5364

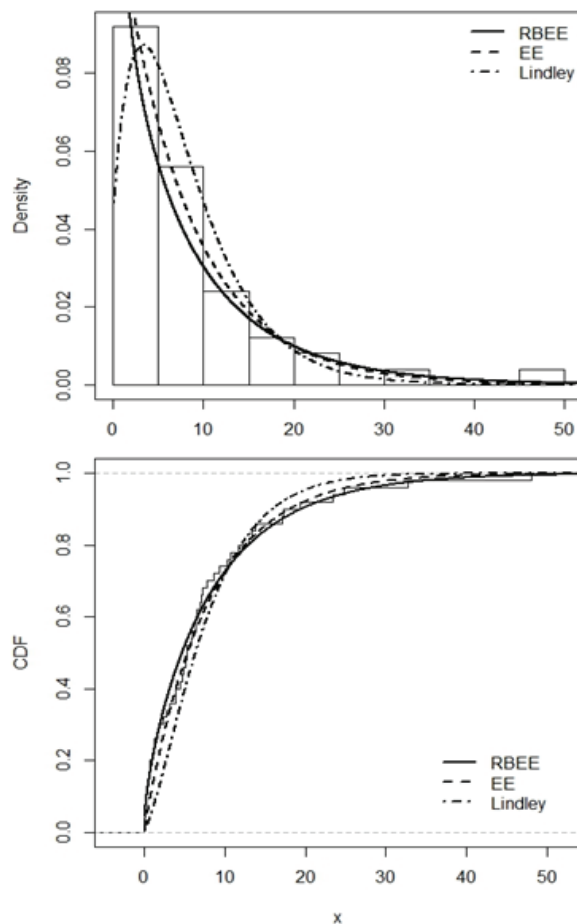
## Conclusion

In this article, we introduce and study a new model of lifetime, called the 'Ristić-Balakrishnan extended exponential' distribution. The proposed

model has three parameters and generalizes important distributions. We provide a comprehensive study of the mathematical and statistical properties of the new model. In addition, the practical utility of the new model was empirically demonstrated. We hope that the RBEE model can be useful for applied statisticians and other researches who refer to a model with few parameters but flexible to accommodate supported data in real positives.

## Acknowledgements

We thank two anonymous referees and the associated editor for their valuable suggestions, which certainly contributed to the improvement of this paper. Additionally, Thiago A. N. de Andrade is grateful the financial support from CAPES (Brazil).

**Figure 5.** Estimated pdf of the RBEE model; Estimated cdf of the RBEE model.

## References

- Bernstein, P. L. (1996). *Against the gods: the remarkable story of risk*. New York, NY: John Wiley & Sons.



- Chen, G., & Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, 27(2), 154-161.
- Cordeiro, G. M., & Bourguignon, M. (2016). New results on the Ristić-Balakrishnan family of distributions. *Communications in Statistics - Theory and Methods*, 45(23), 6969-6988. doi 10.1080/03610926.2014.972573
- Cordeiro, G. M., & Castro, M. (2011). A new family of generalized distribution. *Journal of Statistical Computation and Simulation*, 81(7), 883-893.
- Cordeiro, G. M., Ortega, E. M. M., & Cunha, D. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, 11(1), 1-27.
- Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics - Theory and Methods*, 31(4), 497-512. doi 10.1081/STA-120003130
- Gómez, Y. M., Bolfarine, H., & Gómez, H. W. (2014). A new extension of the exponential distribution. *Revista Colombiana de Estadística*, 37(1), 25-34.
- Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 41(2), 173-188.
- Kenney, J. F., & Keeping, E. S. (1962). *Mathematics of statistics* (3rd ed.). Trenton, NJ: Chapman & Hall.
- Lemonte, A. J. (2013). A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, 62, 149-170. doi 10.1016/j.csda.2013.01.011
- Marshall, A. N., & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrika*, 84(3), 641-652.
- Moors, J. J. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society. Series D*, 37(1), 25-32.
- Murthy, D., Xie, M., & Jiang, R. (2004). *Weibull models. Wiley series in probability and statistics*. Trenton, NJ: John Wiley and Sons.
- Nadarajah, S., Bakouch, H. S., & Tahmasbi, R. (2011). A generalized Lindley distribution. *Sankhya B*, 73(2), 331-359.
- Nadarajah, S., Cordeiro, G. M., & Ortega, E. M. M. (2015). The Zografos-Balakrishnan-G family of distributions: Mathematical properties and applications. *Communications in Statistics - Theory and Methods*, 44(1), 186-215. doi 10.1080/03610926.2012.740127
- Popper, K. (1959). *The logic of scientific discovery*. Viena: Taylor & Francis e-Library.
- Ristić, M. M., & Balakrishnan, N. (2012). The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82(8), 1191-1206. doi 10.1080/00949655.2011.574633
- Sarhan, A. M., & Kundu, D. (2009). Generalized linear failure rate distribution. *Communications in Statistics - Theory and Methods*, 38(5), 642-660. doi 10.1080/03610920802272414
- Tahir, M. H., & Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established G families. *Anais da Academia Brasileira de Ciências*, 87(2), 539-568. doi 10.1590/0001-3765201520140299
- Zografos, K., & Balakrishnan, N. (2009). On the families of beta-and gamma-generated generalized distribution and associated inference. *Statistical Methodological*, 6(4), 344-362. doi 10.1016/j.stamet.2008.12.003

Received on January 25, 2017.

Accepted on September 19, 2017.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.