An Extended Exponentiated-G-Negative Binomial Family with Threshold Effect *

Josemar Rodrigues[a], Gauss M. Cordeiro[b]
and Jorge Bazan[a]

[a] *Universidade de São Paulo*,
Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brazil
[b] *Universidade Federal de Pernambuco, Recife-PE, Brazil*

Abstract

In this paper, we formulate a very flexible family of distributions which unifies most recent lifetime distributions. The main idea is to obtain a cumulative distribution function in the unit interval to transform the baseline distribution with an activation mechanism characterized by a latent threshold variable. The new family has a strong biological interpretation from the competitive risks point of view and an elegant solution thorough the Box-Cox transformation. Several structural properties of the new model are investigated. A Bayesian analysis using Markov Chain Monte Carlo is developed to illustrate with a numerical example the usefulness of the proposed family.

Keywords: Bayesian inference; Box-Cox transformation; competitive risks; contagious models; first hitting time; mathematical properties; stochastic processes.

1 Introduction

The negative binomial distribution has been used extensively in survival analysis and other areas for many years. This distribution has an interesting link with the Box-Cox transformation which has been explored to unify long-term cure rate models (Yin & Ibrahim, 2005; Castro *et al.*, 2010). The main purpose of this paper is to use a general

*Corresponding author: Josemar Rodrigues. Universidade de Saão Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos-SP, Brazil. E-mail: josemar@icmc.usp

version of the negative binomial distribution with an latent random threshold to unify most of the recent lifetime distributions, and to explore its connection with the Box-Cox transformation. The major contribution of this procedure is to change the exponentiated baseline distribution with an activation mechanism based on this threshold variable with an strong biological meaning.

The rest of the paper is organized as follows. In Section 2, the basic assumptions and a latent stochastic process are formulated. In Section 3, the *extended exponentiated-G-negative binomial* ("EEGNB" for short) family is defined with a specific activation mechanism based on the latent stochastic process and the latent threshold variable. In Section 4, some remarks and specific models are discussed. Section 5 is concerned with identifiability problems. In Section 6, we demonstrate that the new density family is a linear combination of exponentiated densities based on the baseline distribution. Section 7 is devoted to some structural properties of the EEGNB distributions. In Section 8, an elegant link with the Box-Cox transformation is presented. An application to real data from the Bayesian viewpoint is discussed in Section 9. Concluding remarks are addressed in Section 10.

2 Assumptions and a latent stochastic process

In order to introduce a new family of distributions the following assumptions will be considered:

- The number N of the competing causes (risk factors) that can produce an event of interest follows the negative binomial distribution with parameters $\theta > 0$ and $\beta \geq -1$ such that $\beta > -\frac{1}{\theta}$. The probability mass function (pmf) of N is

  $$p_n = \frac{\Gamma(\beta^{-1}+n)}{\Gamma(\beta^{-1})n!}\left(\frac{\beta\theta}{1+\beta\theta}\right)^n (1+\beta\theta)^{-1/\beta},\ n = 0, 1, 2, \ldots, \tag{1}$$

  so that $E(N) = \theta$ and $\text{var}(N) = \theta(1+\beta\theta)$.

  The probability generating function (pgf) of N is given by

  $$A(s) = \sum_{n=0}^{\infty} p_n s^n = \{1+\beta\theta(1-s)\}^{-1/\beta},\ \text{for } 0 \leq s \leq 1. \tag{2}$$

  For $s = 0$, the pgf (2) is given by $A(0) = p_0 = (1+\beta\theta)^{-1/\beta}$. This quantity is called the *survival fraction* and it plays an important role in long-term survival analysis. For more details, we suggest to read Rodrigues *et al.* (2008). If $-1/\theta \leq \beta < 0$, we obtain under-dispersion from the Poisson model, and for $\beta > 0$ there is over-dispersion. Therefore, the parameter β is known as dispersion parameter and it unifies most of the recent models in survival analysis.

- Conditional on $N = n$, let the independent latent random variable $X_j,\ j = 1, \ldots, n$, denote the time for the jth risk factor to become responsible (critical factor) for the occurrence of the event of interest. In long-term survival analysis it means the time to initiated cells (risk factors) to be a damaged cell (critical factor) which could be responsible for the death of the patient or detectable tumor in a competing risks approach. Consider that these random variables have probability density function (pdf) $g(x)$ and cumulative distribution function (cdf) $G(x)$, and let $G(x)^\alpha$ be the corresponding exponentiated G (exp-G) distribution function with power $\alpha > 0$.

- Given $N = n$ and a fixed time x, let $Z_1 \ldots, Z_n$ be independent random variables, independently of N, following a Bernoulli distribution with success probability $p(x)$ indicating the presence of the jth critical cause at time x. In long-term survival analysis, it could be a malignant or unrepaired cells.

Now, we introduce the basic idea of our latent critical counting process to extend the exp-G with power parameter $\alpha > 0$. The stochastic counting process is defined as

$$N_x = \begin{cases} Z_1 + Z_2 + \cdots + Z_N, & \text{if } N > 0, \\ 0, & \text{if } N = 0, \end{cases} \tag{3}$$

where N_x is the number of critical factors which are available in the interval $(0, x)$. This counting process is quite attractive, not only because of the mathematical elegance, but also because of its quite intuitive interpretation. It is a thinning binomial process in the sense of counting the critical risk factors among the risk factors, and it is a fatal number when is great or equal to the threshold random variable R, i.e., the event of interest occurs when $N_x \geq R$. By defining the time to the event of interest occurs as a first hitting time (FHT) X, i.e., $X = \min\{x > 0 : N_x \geq R\}$, we are interested in finding a distribution function of X which is an extension of the exp-G with power $\alpha > 0$. This distribution function will be obtained by an activation mechanism characterized by the geometric probability distribution of the latent threshold random variable R as

$$P(R = r; \omega) = \omega\ (1 - \omega)^{r-1}\ ,\ r = 1, 2, \ldots,$$

for $0 < \omega \leq 1$. The parameter ω measures how much we believe in the first-activation mechanism $(R = 1)$ to change the baseline distribution $G(x)$. In order to take into account the effect of the first-activation mechanism on the baseline distribution $G(x)$, we assume that the probability of the jth cause or risk factor to be presented in the interval $(0, x)$ is given by $p(x) = G(x)$.

3 Extended Exponentiated-G-Negative Binomial (EEGNB) family

Now, under the above assumptions and motivated by Ferreira & Steel (2006), we formulate the EEGNB family of distributions with power $\alpha > 0$, by defining a cdf $P(s; \beta, \theta, \omega)$ such that

$$F(x) = [P(G(x)); \beta, \theta, \omega]^{\alpha} \text{ for } x > 0,$$

and $\lim_{\beta \longrightarrow -1} P(s; \beta, \theta, \omega) = s$ for $0 \leq s \leq 1$.

Assuming that N_x is independent of R, we consider the following improper cdf at time x

$$\begin{aligned} S_R(x) &= P(X > x) = P(N_x \leq R - 1) \\ &= P(R \geq N_x + 1) = \sum_{n=0}^{\infty} P(N_x = n) P(R \geq n + 1). \end{aligned} \quad (4)$$

Moreover, if R has a shifted geometric distribution, we obtain from (4) a simple expression for $S_R(x)$ as

$$S_R(x) = A_{N_x}(1 - \omega), \quad (5)$$

where $A_{N_x}(s)$ is the pgf of N_x (for $0 < s < 1$). From equations (2), (3) and (5) and a result in Feller's book (see p. 287 in Feller, 1968), we obtain

$$S_R(x) = A_N[1 - \omega G(x)] = [1 + \beta \rho G(x)]^{-\frac{1}{\beta}}, \quad (6)$$

where $\rho = \theta\omega > 0$. An important point in (6) is its similarity with the unified cure rate model formulated by Rodrigues *et al.* (2008), where ω is changing θ to ρ. Since $0 \leq \omega \leq 1$ and $E(N) = \theta$, it is also an intervention parameter.

Motivated by Rodrigues *et al.* (2011), we define the *extended-G-negative binomial* (EGNB) density function as a weighted density function given by

$$h(x) = \frac{g(x)\, W(x)}{E[W(X)]}, \quad (7)$$

where $W(x) = \omega \frac{dA_N(s)}{ds} \mid_{1 - \omega G(x)}$, and $H(x)$ is its corresponding cdf. Thus, using simple mathematical manipulation in the denominator of (7), the pdf and cdf of the EGNB model follow, under the above assumptions, as

$$h(x) = \frac{\omega g(x) \frac{dA_N(s)}{ds} \mid_{1-\omega G(x)}}{1 - A_N(1-\omega)} = g(x)p(G(x)), \tag{8}$$

and

$$H(x) = \frac{1 - A_N(1 - \omega G(x))}{1 - A_N(1 - w)} = P(G(x))), \tag{9}$$

for $0 < \omega \leq 1$, respectively. From equation (2), the cdf $P(s)$ becomes

$$P(s; \beta, \rho) = \frac{1 - A_N(h_\omega(s))}{1 - A_N(h_\omega(1))} = \frac{1 - (1 + \beta\rho s)^{-\frac{1}{\beta}}}{1 - (1 + \beta\rho)^{-\frac{1}{\beta}}}, \tag{10}$$

for $\beta > -1/\rho$ and $h_\omega(s) = 1 - \omega s$, $0 < s < 1$, where $p(s; \beta, \rho)$ is the corresponding density function of $P(s; \beta, \rho)$. The EGNB distribution function, $H(x)$, is obtained from the baseline cdf $G(x)$ by a composition of two specific stages which do not depend on $G(x)$ as follows:

- First-stage: It is characterized by the geometric distribution of R by means of $h_\omega(s) = 1 - \omega s, \ 0 \leq s \leq 1$,

- Second-stage: It is characterized by the negative-binomial distribution of N by means of $P(s; \beta, \rho) = \frac{1 - A_N(s)}{1 - A_N(h_\omega(1))}, \ 0 \leq s \leq 1$.

The EEGNB with power parameter $\alpha > 0$ for the FHT random variable X is given by

$$F(x) = [H(x)]^\alpha = [P(G(x); \beta, \rho)]^\alpha = [P(h_\omega(G(x)), \beta, \rho)]^\alpha. \tag{11}$$

Note that for $\beta = -1$, the corresponding cdf $P(s; \beta, \rho)$ is the uniform distribution in $(0, 1)$. Also, it is easy to verify that $F(x)$ in (11) includes three recent flexible cumulative distribution functions which will be discussed in the next section, i.e., $\lim_{\beta \longrightarrow 0} P(G(x); \beta, \rho) = \frac{1 - e^{-\rho G(x)}}{1 - e^{-\rho}}$, for $\beta = -1/\rho$, where ρ is the mean number of risk factors after the intervention of the parameter ω, we obtain $P(G(x); \beta, \rho) = 1 - [1 - G(x)]^\rho$, and for $\beta = 1$, we have the Weibull-geometric distribution (Barreto-Souza *et al.*, 2011) given by

$$f(x; p) = (1 - q)\, g(x) \, [1 - qS(x)]^{-2}, \ \text{for } x > 0, \tag{12}$$

where $q = \rho/(1 + \rho)$ and $S(x) = 1 - G(x)$.

From the results presented in this section, the EEGNB cdf with power parameter

$\alpha > 0$ is given by

$$F(x;\beta,\rho,\alpha) = \begin{cases} \left\{\frac{1-[1+\beta\rho G(x)]^{-\frac{1}{\beta}}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}}\right\}^{\alpha}, & \text{if } \alpha > 0,\ \rho > 0,\ \beta > -1/\rho,\ \beta \geq -1, \\ \left[\frac{1-\mathrm{e}^{-\rho G(x)}}{1-\mathrm{e}^{-\rho}}\right]^{\alpha}, & \text{if } \alpha > 0,\ \rho > 0,\ \beta = 0, \\ \{1-[1-G(x)]^{\rho}\}^{\alpha}, & \text{if } \alpha > 0,\ \rho \geq 1,\ \beta = -1/\rho\ (\rho : \text{integer}), \\ [\frac{G(x)}{1-qS(x)}]^{\alpha}, & \text{if } \alpha > 0,\ \rho > 0,\ \beta = 1, \\ [G(x)]^{\alpha}, & \text{if } \alpha > 0,\ \rho > 0,\ \beta = -1, \end{cases} \tag{13}$$

where $\rho = \theta\omega$. Its corresponding density function, for $\rho > 0,\ \beta \geq -1$ such that $\beta\rho \geq -1$, is

$$f(x;\alpha,\beta,\rho) = \alpha\,[P(G(x);\beta,\rho)]^{\alpha-1}\,p(G(x);\beta,\rho)\,g(x),\ x > 0, \tag{14}$$

where $P(s;\beta,\rho)$ is in (10) and its corresponding density function is given by

$$p(s;\beta,\rho) = \frac{\omega\frac{dA_N(u)}{du}\ |_{u=h_\omega(s)}}{1-A_N(h_\omega(1))} = \frac{\rho(1+\beta\rho s)^{-(1+\frac{1}{\beta})}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}},\ 0 \leq s \leq 1, \tag{15}$$

for $\beta \geq -1$ and $\beta > -1/\rho$.

The pdf (15) is defined for $\beta \neq 0$. However, we note that

$$\lim_{\beta\longrightarrow 0} p(s;\beta,\rho) = \frac{\rho\exp(-\rho s)}{1-\exp(-\rho)},\ \text{for } 0 \leq s \leq 1.$$

From the result above, the pdf $p(s;\beta,\rho)$ is an extension of the truncated exponential density in the interval $[0,1]$. This truncated pdf was proposed by Barreto & Simas (2012) and Nadarajah *et al.* (2009) to obtain the Exp-G and the skew-symmetric family of probability distributions, respectively. For $\beta = -1/\rho,\ \rho \geq 1$, the pdf (15) becomes the Kumaraswamy pdf with shape parameters ρ and one (Kumaraswamy, 1980) given by

$$p(s;\rho) = \rho(1-s)^{\rho-1},\ \text{for } 0 \leq s \leq 1.$$

Therefore, by extending the pdf in (15) for $\beta \geq -1$ such that $\beta \geq -1/\rho$ and $\rho > 0$, we obtain a new and very flexible alternative pdf to the beta distribution given by

$$p(s;\beta,\rho) = \begin{cases} \frac{\rho(1+\beta\rho s)^{-(1+\frac{1}{\beta})}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}}, & \text{if } \beta > -1/\rho,\ \beta \geq -1,\ \rho > 0, \\ \frac{1-\exp\{-\rho s\}}{1-\exp\{-\rho\}}, & \text{if } \beta = 0,\ \rho > 0, \\ \rho(1-s)^{\rho-1}, & \text{if } \beta = -1/\rho,\ \rho \geq 1\ (\rho : \text{integer}), \\ \frac{1-q}{[1-q(1-s)]^2}, & \text{if } \beta = 1,\ q = \frac{\rho}{1+\rho},\ \rho > 0, \\ 1, & \text{if } \beta = -1,\ \rho > 0, \end{cases}$$

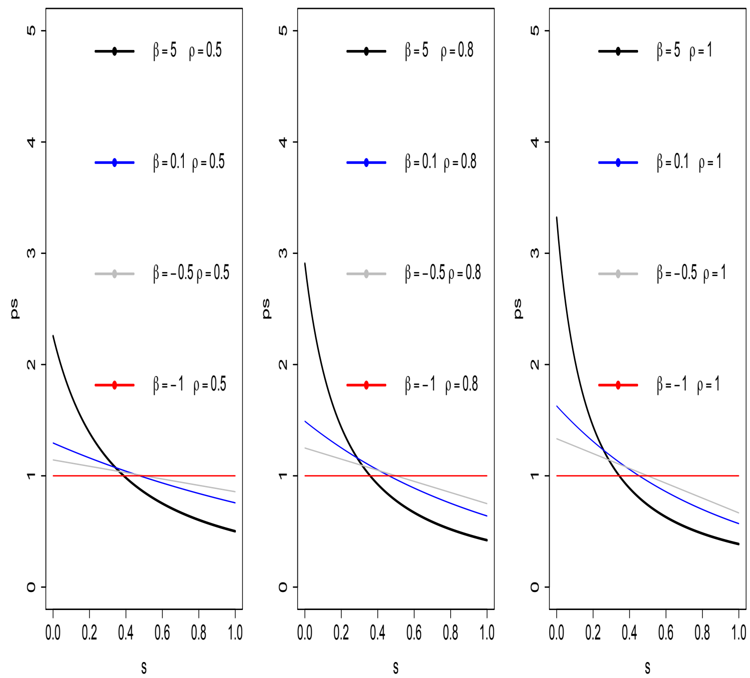

Figure 1: The EUNB pdf for $\alpha = 1$ and different values of β and ρ.

for $0 \leq s \leq 1$. From now on, this pdf is called extended-*uniform*-negative binomial (EUNB) distribution. Plots of the pdf above are displayed in Figure 1. It is clear that the uniform distribution in $(0, 1)$ has been embedded in a wider EUNB family, with an additional shape parameter β. Due to this additional shape parameter, more flexibility can be incorporated in the family when the parameter ρ goes to zero. This flexibility could be useful for data analysis purposes.

From (10) and (15), we have for inferential purposes in the next section an alternative expression for the pdf (14) of the FHT random variable X as follows:

$$f(x;\alpha,\beta,\rho) = \rho\alpha g(x)\frac{[1+\beta\rho G(x)]^{-(1+\frac{1}{\beta})}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}}\left\{\frac{1-[1+\beta\rho G(x)]^{-\frac{1}{\beta}}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}}\right\}^{\alpha-1}, \; x>0, \tag{16}$$

for $\beta \geq -1$ and $\beta > -1/\rho$. A random variable X having density function (16) is denoted by $X \sim \text{EEGNB}(\rho, \alpha, \beta)$.

The hazard rate function of X takes the form

$$r(x;\alpha,\beta,\rho) = \frac{\alpha\,[P(G(x);\beta,\rho)]^{\alpha-1}\,p(G(x);\beta,\rho)\,g(x)}{1-[P(G(x),\beta,\rho)]^{\alpha}}. \tag{17}$$

Four plots of the EE-*Weibull*-NB (EEWNB) density function (16) and three plots of the hazard rate function (17) for selected values of α, β and ρ are displayed in Figure

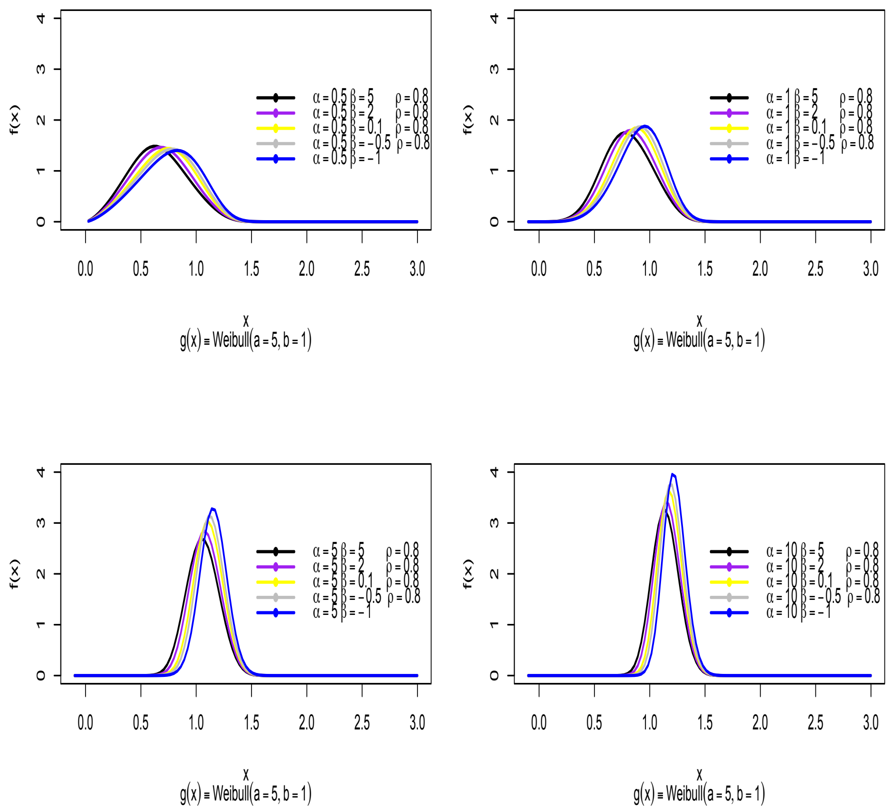

Figure 2: Effect of latent threshold variable on the Weibull distribution for $\alpha = 0.5, 1, 5, 10$.

2 and 3, respectively. For these examples, the values of the shape and scale parameters of the Weibull distribution are $a = 0.5, 4, 1, 5$ and $b = 1, 100$, respectively. These plots show the effect of the power parameter α and the first-activation mechanism through the parameter β on the Weibull distribution. It is clear that the Weibull distribution, or, the exp-W distribution are limiting distributions of the EEWNB family when the parameter β approaches minus one for different α-values and fixed ρ. Figure 4 reveals an important feature of the EEWNB distribution that is its considerable flexibility in providing hazard rates of various shapes.

4 Some remarks and specific recent lifetime models

The EEGNB family (13) unifies most recent lifetime distributions as an alternative to the classical models in survival analysis. Now, we provide some examples.

- Exponentiated Kumuraswamy-G distribution with power parameter α

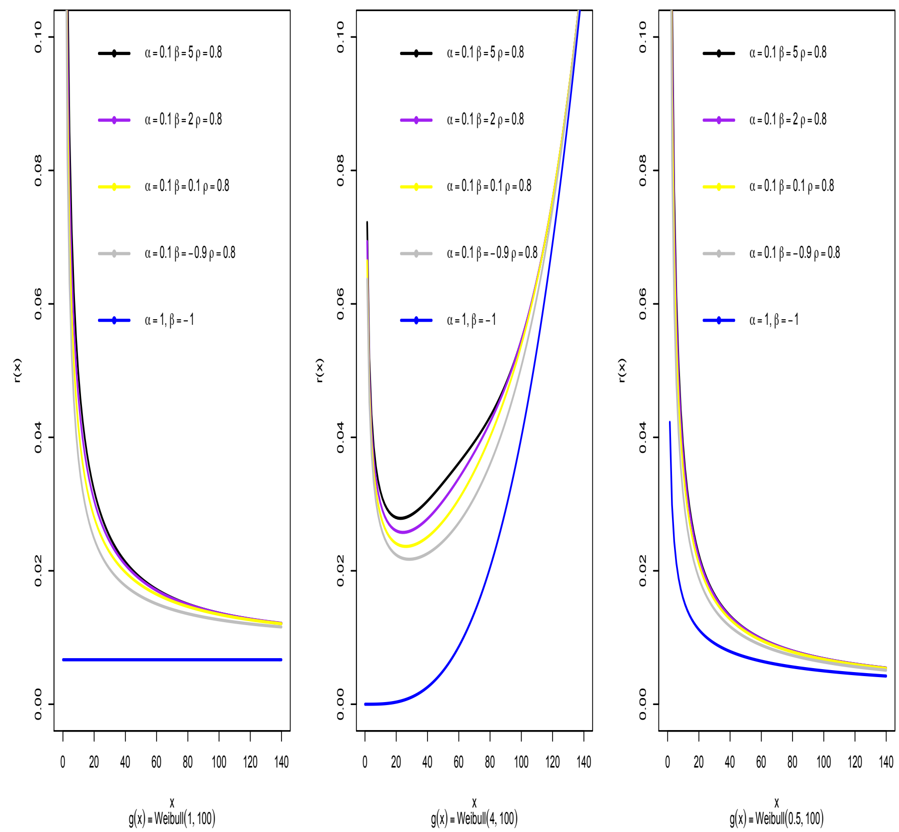

Figure 3: The hazard rate function of the EEWNB distribution (17) for some parameter values.

For $\beta = -1/\rho,$ and $\rho \geq 1$, we obtain from (13) the cdf

$$F(x;\rho,\alpha) = [1-(1-G(x))^{\rho}]^{\alpha}\,,\; x>0. \tag{18}$$

- The exponentiated Kumuraswamy distribution with power parameter α (Lemonte *et al.*, 2013)

  Assuming that the FHT random variable X has an exponentiated uniform distribution in the interval $(0,1)$ with an arbitrary power $\gamma > 0$, we obtain from (18) the exponentiated Kumuraswamy distribution given by

$$F(x;\gamma,\rho,\alpha) = [1-(1-x^{\gamma})^{\rho}]^{\alpha}\,,\; \text{for } x \in (0,1).$$

- The extended-G-geometric (EGGe) distribution.

  If we take $\beta = 1$ and $\alpha = 1$, the EEGNB distribution (16) becomes the geometric-minimum stable distribution (Marshall & Olkin, 1997). It is an extension of the Weibull-geometric distribution (Barreto-Souza *et al.*, 2011) given by

$$f(x;p) = (1-q)\,g(x)\,[1-qS(x)]^{-2},\; \text{for } x>0, \tag{19}$$

  where $q = \rho/(1+\rho)$ and $S(x) = 1 - G(x)$. The mathematical properties of this distribution was studied by Cordeiro & Lemonte (2013). It has a nice stability property (Marshall & Olkin, 1997) and a strong biological interpretation provided by the EEGNB family (16). If we take $G(x) = \Phi(v)$ in (9), where $v = v(x) = \rho^{\star}(x/\eta^{\star})/\alpha^{\star}$, $\rho^{\star}(z) = z^{1/2} - z^{-1/2}$, $\alpha^{\star} > 0$, $\eta^{\star} > 0$ and $\Phi(\cdot)$ is the standard normal cdf, we obtain a geometric version of the Marshall-Olkin extended Birnbaum-Saunders distribution (Lemonte, 2013) given by

$$F(x;\eta^{\star},\alpha^{\star},\rho) = \frac{\Phi(x)}{1-(1-p)\Phi(-v)},$$

  where $p = 1/(1+\rho)$.

- The extended generalized exponential (EGE) family(Kundu & Gupta, 2011)

  Assuming that X in (18) has an uniform distribution in the interval $(0,\rho/\lambda)$, the EGE cdf becomes

$$F(x;\alpha,\rho,\lambda) = \left[1-\left(1-\frac{\lambda}{\rho}x\right)^{\rho}\right]^{\alpha}.$$

- The log-exponentiated Kumuraswamy (log-exp-Kw) distribution (Lemonte *et al.*, 2013)

Consider that X has an exponentiated exponential random variable with power parameter γ. It follows from (18) that the log-EK distribution is

$$F(x,\gamma,\rho,\alpha) = \left\{1-\left[1-\left(1-\mathrm{e}^{-x}\right)^{\gamma}\right]^{\rho}\right\}^{\alpha}.$$

- The exponentiated Lomax distribution

  If we take $G(x)=\frac{x}{\gamma+x}$ in (18), we obtain a new extended Lomax distribution given by

$$F(x;\gamma,\rho,\alpha) = \left[1-\left(\frac{\gamma}{\gamma+x}\right)^{\rho}\right]^{\alpha}, \text{ for } x>0.$$

  If $\alpha=1$, we obtain the Lomax distribution with parameters $\gamma>0$ and $\rho>0$ discussed and generalized by Lemonte & Cordeiro (2011).

- The modified first hitting time G (MFHT-G) distribution

  The MFHT-G model can be obtained as follows:

$$\lim_{\beta\to 0} F(x;\beta,\rho,\alpha) = \left[\frac{1-\mathrm{e}^{-\rho G(x)}}{1-\mathrm{e}^{-\rho}}\right]^{\alpha}.$$

  If $\alpha=1$, the MFHT-G is the Exp-G model formulated by Barreto & Simas (2012).

- The exponentiated-G distribution (exp-G) (Pewsey & Gomes, 2013)

  If $\beta=-1$, the EEGNB reduces to the ExpG with power parameter α, namely $F(x;-1,\rho,\alpha)=G(x)^{\alpha}$.

5 Identifiability of the EEGNB distributions

A simple but powerful characterization of the exp-G distribution with a strong biological meaning can be obtained from Theorem 1(a) in Ferreira & Steel (2006) as follows:

$$\beta=-1 \text{ if and only if } F(x)=G(x)^{\alpha}.$$

The parameters β and ω play a crucial role as measure of the impact of the activation mechanism R to change the baseline distribution G, or, to change the mean parameter θ to ρ . If we take $\omega=1$, or, $\rho=\theta$, it means that we are totally convinced that the change on the baseline distribution G is provided by the first-activation mechanism, and the EEGNB distribution function will be given by

$$F(x;\beta,\alpha,\theta) = \left[\frac{1-(1+\beta\theta G(x))^{-\frac{1}{\beta}}}{1-(1+\beta\theta)^{-\frac{1}{\beta}}}\right]^{\alpha} \quad \alpha>0,\ \beta>-1/\theta. \tag{20}$$

On the other hand, the parameter β unifies the parent distribution function $G(x)$ and recent lifetime models for $\beta = -1, 0, 1$, respectively. Unfortunately, the parameter ω is not identifiable for inferential purposes. However, from (3) and the definition of the FHT X, we can obtain an important relationship between the geometric distribution of the latent threshold R and the negative binomial distribution (1) given by

$$\frac{1-\omega}{\omega} \leq \frac{\rho}{2}, \text{ or, } \omega \geq \frac{1}{1+\frac{\rho}{2}}, \tag{21}$$

for $\rho > 0$. The above expression and the dispersion parameter β suggest a helpful measure of the impact or effect of the first-activation mechanism as follows:

$$\omega^{\star} = \begin{cases} 0, & \text{if } \beta = -1, \\ \frac{1}{1+\frac{\rho}{2}}, & \text{if } \beta > -1 \text{ and } \beta \neq 0, \\ 1, & \text{if } \beta = 0. \end{cases} \tag{22}$$

The quantity $\omega^{\star}$ in (22) shows a strong relationship among the first-activation mechanism and the dispersion parameter β. For instance, if $\beta > 0$, we have an aggregation or contagious (Neyman, 1939; Feller, 1943) of the risk factors in space or time (over-dispersion) which is undesirable and harmful to the first-activation mechanism in the sense of obtaining small values for ω, whereas the reverse is not true if $\beta = 0$ (Poisson distribution). If $-1 < \beta < 0$ (under-dispersion), we can also to harmful the first-activation mechanism. In conclusion, our model suggests that we must have to consider the effect of the dispersion of the risk factors on the first-activation mechanism. We think that this problem could be deserve more attention when suggesting or formulating alternative or flexible models in survival analysis. In the next sections, we study some mathematical properties of the EEGNB model and how the dispersion of the risk factors and the first-activation mechanism change the baseline distribution G thorough the Box-Cox transformation.

6 A useful expansion

From (9) the cdf $H(x) = H(x;\beta,\rho) = \frac{1-[1+\beta\rho G(x)]^{-\frac{1}{\beta}}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}}$. Then, the corresponding pdf is given by

$$h(x) = \rho\, g(x)\, \frac{[1+\beta\rho G(x)]^{-(1+\frac{1}{\beta})}}{1-(1+\beta\rho)^{-\frac{1}{\beta}}}.$$

Then, equation (16) reduces to a simple exponentiated-H ("exp-H") distribution with power parameter $\alpha > 0$ given by $f(x;\alpha,\beta,\rho) = \alpha\, h(x)\, H(x)^{\alpha-1}$.

In order to obtain a useful expansion for $f(x;\alpha,\beta,\rho)$, we require two expansions. First, for $z \in (0,1)$ and any real non-integer α, we have

$$z^{\alpha} = \sum_{r=0}^{\infty} s_r(\alpha)\, z^r, \tag{23}$$

where $s_r(\alpha) = \sum_{m=r}^{\infty} (-1)^{m+r} \binom{\alpha}{m} \binom{m}{r}$. Second, for any real ν, we can expand ν^{λ} in Taylor series

$$\nu^{\lambda} = \sum_{k=0}^{\infty} (\lambda)_k \, \frac{(\nu-1)^k}{k!} = \sum_{j=0}^{\infty} f_j(\lambda)\, \nu^j, \tag{24}$$

where

$$f_j(\lambda) = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} (\lambda)_k / k!$$

and $(\lambda)_k = \lambda(\lambda-1)\ldots(\lambda-k+1)$ is the descending factorial.

We can write from (23) $H(x)^{\alpha-1} = \sum_{r=0}^{\infty} s_r(\alpha-1)\, H(x)^r$ and then expanding the binomial term, we have

$$H(x)^{\alpha-1} = \sum_{r=0}^{\infty} \frac{s_r(\alpha-1)}{[1-(1+\beta\rho)^{-\frac{1}{\beta}}]^r} \sum_{i=0}^{r} (-1)^i \binom{r}{i} [1+\beta\rho G(x)]^{-\frac{i}{\beta}}.$$

Based on equation (24), we can express $h(x)$ as

$$h(x) = \frac{\rho\, g(x)}{1-(1+\beta\rho)^{-\frac{1}{\beta}}} \sum_{j=0}^{\infty} f_j(-1-\beta^{-1})\, [1+\beta\rho G(x)]^j.$$

Combining the last two expressions, we obtain

$$f(x;\alpha,\beta,\rho) = g(x) \sum_{j,r=0}^{\infty} t_{j,r} \sum_{i=0}^{r} (-1)^i \binom{r}{i} [1+\beta\rho G(x)]^{j-\frac{i}{\beta}}, \tag{25}$$

where

$$t_{j,r} = t_{j,r}(\alpha,\beta,\rho) = \frac{\alpha\beta f_j(-1-\beta^{-1})\, s_r(\alpha-1)}{[1-(1+\beta\rho)^{-\frac{1}{\beta}}]^{r+1}}$$

Using again (24), we can write

$$[1+\beta\rho G(x)]^{j-\frac{i}{\beta}} = \sum_{s=0}^{\infty} f_s(j-i\beta^{-1}) \sum_{m=0}^{s} \binom{s}{m} \beta^m \rho^m G(x)^m$$

and by changing summations, we obtain

$$[1+\beta\rho G(x)]^{j-\frac{i}{\beta}} = \sum_{m=0}^{\infty} v_{m,j,i}\, G(x)^m,$$

where

$$v_{m,j,i} = v_{m,j,i}(\beta,\rho) = \sum_{s=m}^{\infty} f_s(j-i\beta^{-1}) \binom{s}{m} \beta^m \rho^m.$$

Inserting the expansion for $[1+\beta\rho G(x)]^{j-\frac{i}{\beta}}$ in (25), and after some algebra, we obtain

$$f(x;\alpha,\beta,\rho) = g(x)\sum_{m=0}^{\infty} p_{m+1}\, G(x)^m, \tag{26}$$

where

$$p_{m+1} = p_{m+1}(\alpha,\beta,\rho) = \sum_{j,r=0}^{\infty} t_{j,r} \sum_{i=0}^{r} (-1)^i \binom{r}{i} v_{m,j,i}.$$

An alternative to equation (26) is given by

$$f(x;\alpha,\beta,\rho) = \sum_{m=0}^{\infty} q_{m+1}\, h_{m+1}(x), \tag{27}$$

where $q_{m+1} = p_{m+1}(\alpha,\beta,\rho)/(m+1)$ for $m \geq 0$ and $h_{m+1}(x)$ denotes the exp-G distribution with power parameter $m+1$. Equation (27) reveals that the density function of the EUGNB family can be expressed as a linear combination of exp-G densities. So, some mathematical properties of the EUGNB family can be obtained by knowing those of the exp-G distributions. See, for example, Mudholkar *et al.* (1995), Gupta & Kundu (2001a) and Nadarajah & Kotz (2006a), among others. For example, the ordinary and incomplete moments and moment generating function (mgf) of X can be obtained immediately from those quantities of the exp-G distribution.

The cdf corresponding to (27) is given by $F(x;\alpha,\beta,\rho) = \sum_{m=0}^{\infty} q_{m+1}\, H_{m+1}(x)$, where $H_{m+1}(x)$ denotes the exp-H distribution with power parameter $m+1$.

The formulae derived in the next sections can be easily handled in most symbolic computation software platforms such as Maple, Mathematica and Matlab. These platforms have currently the ability to deal with analytic expressions of formidable size and complexity. Established explicit expressions to calculate statistical measures can be more efficient than computing them directly by numerical integration. The infinity limit in these sums can be substituted by a large positive integer such as 20 or 30 for most practical purposes.

7 Mathematical properties

7.1 Baseline quantile function

Quantile functions are in widespread use in general statistics and often find representations in terms of lookup tables for key percentiles. By inverting the first equation in (13), we obtain the EEGNB quantile function

$$Q(u) = Q_G\left(\frac{[1-(1-p)u^{1/\alpha}]^{-\beta}-1}{\beta\rho}\right), \tag{28}$$

where $p=(1+\beta\rho)^{-\frac{1}{\beta}}$. Simulations of the new family from (28) are straightforward.

For some baseline distributions with closed-form quantile function $x = Q_G(u) = G^{-1}(u)$, it is possible to invert $G(x)$. However, for some other models the solution is not possible. Power series methods are at the heart of many aspects of applied mathematics and statistics. When the function $x = Q_G(u)$ does not have a closed-form expression, it can usually be written in terms of a power series expansion of u, namely

$$x = Q_G(u) = \sum_{i=0}^{\infty} a_i\, u^i, \tag{29}$$

where the coefficients a_i are suitably chosen real numbers. For several important distributions, such as the normal, Student t, gamma and beta distributions, $Q_G(u)$ does not have a closed-form expression but it can be expanded as in equation (29). Here, we obtain some mathematical properties of the EEGNB family using a power series for $x = Q_G(u)$.

First, we use an equation in Section 0.314 of Gilchrist (2000) for a power series raised to any natural power n

$$Q_G(u)^n = \left(\sum_{i=0}^{\infty} a_i\, u^i\right)^n = \sum_{i=0}^{\infty} c_{n,i}\, u^i, \tag{30}$$

where the coefficients $c_{n,i}$ (for $i = 1, 2, \ldots$) are easily obtained from the recurrence equation

$$c_{n,i} = (i\, a_0)^{-1} \sum_{m=1}^{i} [m(n+1) - i]\, a_m\, c_{n,i-m}, \tag{31}$$

where $c_{n,0} = a_0^n$. The coefficient $c_{n,i}$ can be determined from $c_{n,0}, \ldots, c_{n,i-1}$ and hence from the quantities $a_0, \ldots, a_i$. Equations (30) and (31) are used throughout this paper.

7.2 Moments

Hereafter, we shall assume that $G(x)$ is the cdf of a random variable W and that $F(x)$ is the cdf of a random variable X having density function (16). A first representation for the moments of the EEGNB distribution can be obtained from the (r, k)th probability weighted moments (PWMs) of W defined by (for $r, k = 0, 1, 2, \ldots$)

$$\tau_{r,k} = \mathrm{E}[W^r\, G(W)^k] = \int_{-\infty}^{\infty} x^r\, G(x)^k\, g(x) dx. \tag{32}$$

From equation (27), we can write the nth moment of X as

$$\mu'_n = \sum_{m=0}^{\infty} q_{m+1}\,(m+1)\,\tau_{n,m}. \tag{33}$$

So, the moments of X can be expressed as an infinite weighted sum of PWMs of the parent distribution. A second formula for $\tau_{n,m}$ can be based on the parent quantile function $Q_G(x) = G^{-1}(x)$. Setting $G(x) = u$, we obtain

$$\tau_{n,m} = \int_0^1 Q_G(u)^n \, u^m du. \tag{34}$$

From equations (30) and (31), we have $\tau_{n,a} = \sum_{i=0}^{\infty} \frac{c_{n,i}}{(a+i+1)}$. Based on equation (27), we can obtain a second representation for μ'_n

$$\mu'_n = \sum_{m,i=0}^{\infty} \frac{(m+1)\, q_{m+1}\, c_{n,i}}{n+m+i+1} \tag{35}$$

Now, let $Y_{m+1} \sim$exp-G$(m+1)$. A third formula for the nth moment of X can be obtained from (27) as

$$\mu'_n = \sum_{m=0}^{\infty} q_{m+1}\, E(Y_{m+1}^n). \tag{36}$$

Expressions for moments of several exponentiated distributions are given by Nadarajah & Kotz (2006b), which can be used to obtain $E(X^n)$.

The ordinary moments of several EUGNB distributions can be determined directly from equation (36). Here, we provide only two examples. The nth moment of the EE-*Exponential*-NB (with parameter $\lambda > 0$) is given by

$$\mu'_n = n!\, \lambda^n \sum_{m,j=0}^{\infty} \frac{(-1)^{n+j}\,(m+1)\binom{m}{j}\, q_{m+1}}{(j+1)^{n+1}}.$$

For the EE-*Logistic*-NB, where $G(x) = (1 + \mathrm{e}^{-\mathrm{x}})^{-1}$, using a result from Prudnikov *et al.* (1986) (Section 2.6.13, equation 4), we obtain (for $t < 1$)

$$\mu'_n = \sum_{m=0}^{\infty} (k+1)\, q_{m+1} \left(\frac{\partial}{\partial t}\right)^n B(t+1+m, 1-t)\Big|_{t=0},$$

where $B(a,b) = \int_0^1 t^{a-1}\,(1-t)^{b-1} dt$ is the beta function.

Explicit expressions for the moments of various EEGNB distributions can be determined from equations (33)-(36). In the following, we provide some examples.

- **EE-Normal-NB (EENNB) distribution**

  The EENNB distribution is defined from (20) by taking the normal $\mathrm{N}(\mu, \sigma^2)$ as the baseline distribution. The moments of $W \sim \mathrm{N}(\mu, \sigma^2)$ can be obtained from the moments of $Z \sim \mathrm{N}(0,1)$ using $\mathrm{E}(W^r) = \sum_{k=0}^{r} \mu^{r-t}\, \sigma^r\, \mathrm{E}(Z^r)$, and then we can work with the standard normal distribution in general terms. Consider the error function $\mathrm{erf}(x)$ defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-\mathrm{t}^2} \mathrm{dt}.$$

From equation

$$\Phi(x) = \frac{1}{2}\left\{1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right\},$$

we can expand $\mathrm{erf}(x)$ (with $\mu = 0$ and $\sigma = 1$) in power series (with $\mu = 0$ and $\sigma = 1$) given by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{m=0}^{\infty} \frac{(-1)^m\, x^{2m+1}}{(2m+1)\, m!}.$$

From Cordeiro & Nadarajah (2011), (equation (11)), the normal PWMs can be expressed in terms of the Lauricella functions of type A (Aarts, 2000) defined by

$$F_A^{(n)}(a; b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n) = \\ \sum_{m1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\ldots+m_n} (b_1)_{m_1} \ldots (b_n)_{m_n}}{(c_1)_{m_1} \ldots (c_n)_{m_n}} \frac{x_1^{m_1} \ldots x_n^{m_n}}{m_1! \ldots m_n!},$$

where $(a)_i = a(a+1)\ldots(a+i-1)$ is the ascending factorial given by (with the convention that $(a)_0 = 1$). The (r, j)th PWM of the normal distribution is

$$\tau_{r,k} = 2^{r/2}\, \pi^{-(k+1/2)} \sum_{\substack{l=0 \\ (r+k-l)\text{ even}}}^{k} \binom{k}{l} (\frac{\pi}{2})^l\, \pi^l\, \Gamma\left(\frac{r+k-l+1}{2}\right) \times \\ F_A^{(k-l)}\left(\frac{r+k-l+1}{2}; \frac{1}{2}, \ldots, \frac{1}{2}; \frac{3}{2}, \ldots, \frac{3}{2}; -1, \ldots, -1\right).$$

This equation holds when $r+k-l$ is even and it vanishes when $r+k-l$ is odd. So, the EENNB moments can be expressed as an infinite weighted linear combination of Lauricella functions of type A. Finally, the EENNB moments can be obtained from (27) as

$$\mu'_n = 2^{n/2} \sum_{k=0}^{\infty} \frac{(k+1)\, q_{k+1}}{\pi^{(k+1/2)}} \sum_{\substack{l=0 \\ (n+k-l)\text{ even}}}^{k} \binom{k}{l} 2^{-l}\, \pi^l\, \Gamma\left(\frac{n+k-l+1}{2}\right) \times \\ F_A^{(k-l)}\left(\frac{n+k-l+1}{2}; \frac{1}{2}, \ldots, \frac{1}{2}; \frac{3}{2}, \ldots, \frac{3}{2}; -1, \ldots, -1\right).$$

- **EE-Gamma-NB (EEGaNB) distribution**

The EEGaNB distribution is defined in (13) by taking the gamma Ga(a, b) with shape parameter a and scale parameter b as the baseline distribution. Using the power series expansion for the incomplete gamma function

$$G_{a,b}(x) = \frac{(bx)^a}{\Gamma(a)} \sum_{m=0}^{\infty} \frac{(-b\,x)^m}{(a+m)\,m!},$$

and from the equation (9) in Cordeiro & Nadarajah (2011), the quantities $\tau_{r,k}$ can be determined as

$$\tau_{r,k} = \frac{\Gamma(r + a(k+1))}{a^k\, b^r\, \Gamma(a)^{k+1}}\; F_A^{(k)}(r + a(k+1); a, \ldots, a; a+1, \ldots, a+1, -1, \ldots, -1).$$

Finally, the moments of EEGaNB distribution can be obtained from equation (27) as follow

$$\mu_n' = \sum_{m=0}^{\infty} (m+1)\, q_{m+1}\, \frac{\Gamma(n + a(m+1))}{a^m\, b^n\, \Gamma(a)^{m+1}} \times$$
$$F_A^{(m)}(n + a(m+1); a, \ldots, a; a+1, \ldots, a+1, -1, \ldots, -1).$$

7.3 Incomplete moments

For empirical purposes, the shape of many distributions can be usefully described by what we call the incomplete moments. These types of moments play an important role for measuring inequality, for example, income quantiles and Lorenz and Bonferroni curves, which depend upon the incomplete moments of the distribution. The nth incomplete moment of X is determined as

$$m_n(y) = E(X^n | X < y) = \sum_{m=0}^{\infty} (m+1)\, q_{m+1} \int_0^{G(y)} Q_G(u)^n\, u^m du. \tag{37}$$

The last integral can be computed for most baseline G distributions. Using (30) and (31), we obtain

$$m_n(y) = E(X^n | X < y) = \sum_{m,i=0}^{\infty} \frac{(m+1)\, q_{m+1}\, c_{n,i}\, G(y)^{m+i+1}}{m+i+1}, \tag{38}$$

where $c_{n,i}$ is determined from (31).

7.4 Generating function

Here, we provide three formulae for the mgf $M(s) = E(\mathrm{e}^{s\,X})$ of X. Clearly, the first formula for $M(s)$ comes from equation (27) as

$$M(s) = \sum_{m=0}^{\infty} q_{m+1}\, M_{m+1}(s), \tag{39}$$

where $M_{m+1}(s)$ is the mgf of the exp-G distribution with power parameter $m+1$. Hence, $M(s)$ can be determined from the exp-G generating function.

A second formula for $M(s)$ can be derived from equation (27) as

$$M(s) = \sum_{m=0}^{\infty} (m+1)\, q_{m+1}\, \rho_m(s), \tag{40}$$

where the quantity $\rho_m(s) = \int_{-\infty}^{\infty} \mathrm{e^{sx}\, G(x)^m\, g(x) dx}$ can be obtained from the baseline quantile function $Q_G(u) = G^{-1}(u)$ as

$$\rho_m(s) = \int_0^1 \exp[s\, Q_G(u)]\, u^m du. \tag{41}$$

A third formula for $M(s)$ can be obtained from (40) by expanding the exponential function in (41) and using (30) and (31)

$$\rho_m(s) = \sum_{n,i=0}^{\infty} c_{n,i} \int_0^1 \frac{s^n\, u^{i+m}}{n!} du = \sum_{n,i=0}^{\infty} \frac{c_{n,i}\, s^n}{(i+m+1)\, n!}. \tag{42}$$

7.5 Mean deviations

The mean deviations about the mean ($\delta_1(X) = E(|X - \mu_1'|)$) and about the median ($\delta_2(X) = E(|X - M|)$) of X can be expressed as

$$\delta_1(X) = 2\mu_1'\, F\left(\mu_1'\right) - 2m_1\left(\mu_1'\right) \qquad \text{and} \qquad \delta_2(X) = \mu_1' - 2m_1(M), \tag{43}$$

respectively, where $F(\mu_1')$ is easily calculated from (13), $\mu_1' = \mathrm{E}(X)$ can be obtained from one of the formulae derived in Section 7.2, $M = Q_G\left(\frac{[1-(1-p_0)2^{-1/\alpha}]^{-\beta}-1}{\beta\rho}\right)$ is the median of X and $m_1(z) = \int_{-\infty}^{z} x\, f(x) dx$ is the first incomplete moment obtained from (37) or (38).

Now, we provide two alternative ways to compute $\delta_1(X)$ and $\delta_2(X)$. A general equation for $m_1(z)$ can be derived from equation (37) as

$$m_1(z) = \sum_{m=0}^{\infty} q_{m+1}\, J_{m+1}(z), \tag{44}$$

where

$$J_{m+1}(z) = (m+1) \int_{-\infty}^{z} x\, h_{m+1}(x) dx. \tag{45}$$

Equation (45) is the basic quantity to compute the mean deviations for the EEGNB distributions. The mean deviations (43) depend only on the first incomplete moment of the exp-G distributions. A simple application of (44) and (45) refers to the EE-Weibull-NB distribution. The exponentiated Weibull density function (for $x > 0$) with power parameter $m+1$, shape parameter c and scale parameter β is given by

$$h_{m+1}(x) = c\,(m+1)\,\beta^c\, x^{c-1} \exp\left\{-(\beta x)^c\right\} \left[1 - \exp\left\{-(\beta x)^c\right\}\right]^m,$$

and then $J_{m+1}(z)$ can be reduced to

$$J_{m+1}(z) = \beta^{-1} \sum_{r=0}^{\infty} \frac{(-1)^r\,(m+1)\binom{m}{r}}{(r+1)^{1+1/c}}\,\gamma\left(1+c^{-1},(r+1)(\beta z)^c\right),$$

where $\gamma(a,x)=\int_0^x t^{a-1}\,\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$ is the incomplete gamma function.

A second general formula for $m_1(z)$ can be derived by setting $u=G(x)$ in (27)

$$m_1(z) = \sum_{k=0}^{\infty}(m+1)\,q_{k+1}\;T_m(z), \tag{46}$$

where $T_m(z)$ is given by

$$T_m(z) = \int_0^{G(z)} Q_G(u)\,u^m du.$$

Applications of equations (44) and (46) can be conducted to obtain Bonferroni and Lorenz curves defined for a given probability π by $B(\pi)=m_1(q)/\pi\,\mu_1'$ and $L(\pi)=m_1(q)/\mu_1'$, respectively, where $q=Q_G(\pi)$ is computed from the baseline quantile function.

8 Box-Cox transformation

The Box-Cox transformation has been widely used since it was proposed in their famous 1964's paper and it has inspired a large amount of research in many fields, especially in econometrics. Usually, the Box-Cox transformation is applied to the response variable of the linear regression model, when the normality assumption of the errors is not satisfied. An excellent review of the Box-Cox transformation technique can be found in Sakia (1992). More recently, Yin & Ibrahim (2005) have proposed a class of models though Box-Cox transformation for cure rate models, and Castro *et al.* (2010) have justified it by means of the negative binomial distribution. Here, we applied the Box-Cox transformation in a different manner as it was proposed before. Now, we are looking for a new distribution $F(x)$ which is a solution of the following equations :

$$\begin{cases} P_{BC}(h_p([F(x)]^{1/\alpha};\beta) = \rho G(x), & \text{if } \beta \neq 0,\\ P_{BC}(h_p([F(x)]^{1/\alpha};\beta) = -\rho G(x), & \text{if } \beta = 0, \end{cases} \tag{47}$$

where $h_p,\ p$ and $P_{BC}(s;\beta)$ (the Box-Cox transformation) are given by $h_p(s)=1-(1-p)s$,

$$p = \begin{cases} (1+\beta\rho)^{-1/\beta}, & \text{if } \beta \neq 0,\\ \mathrm{e}^{-\rho}, & \text{if } \beta = 0, \end{cases} \tag{48}$$

and

$$P_{BC}(s;\beta) = \begin{cases} \frac{s^{-\beta}-1}{\beta}, & \text{if } \beta \neq 0, \\ \log(s), & \text{if } \beta = 0, \end{cases} \tag{49}$$

respectively. From equation (48) we obtain the Fisher parametrization which will be useful for inferential and computational purposes as follows:

$$\rho = \begin{cases} \frac{p^{-\beta}-1}{\beta}, & \text{if } \beta \neq 0, \\ -\log(p), & \text{if } \beta = 0. \end{cases} \tag{50}$$

The solution of the equations in (47) is the EEGNB distribution proposed before. However, it provides a different interpretation how the baseline distribution $G(x)$ is changed by the latent first activation mechanism thorough the parameter ω. For example, the right-side of the first equation (47), the baseline distribution $G(x)$ is changed only by the parameter ρ (it is involving the first-activation mechanism), and the left-side the parameters β and α are changing $F(x)$ by using the Box-Cox transformation. As before, this transformation provides a natural link between the MFHT-G (Barreto & Simas, 2012) and the EEGNB formulated in (13). We believe that it is a new manner to interpret the effect on the parent distribution to obtain new and more flexible alternative models in survival analysis.

9 Bayesian inference: An illustrative example

In this section, we develop a Bayesian analysis thorough the MCMC algorithm of the methodology in Section 3 to a data set corresponding to a record of 799 intervals between pulses along a nerve fibre presented in Cox & Lewis (1966). We compare the EEWNB with the results of the MOEW model and the three-parameter exponentiated Weibull (exp-W) model reported by (Cordeiro & Lemonte, 2013). The cdf of the exp-W distribution for $x > 0$ takes the form $G(x) = [F(x;\gamma_1,\gamma_2)]^{\alpha} = [1-\exp\{-(\gamma_2 x)^{\gamma_1}\}]^{\alpha}$, when $\alpha > 0$. Throughout the whole Bayesian estimation procedure, the dispersion parameter β is fixed and the optimal choice of β can be determined by the Bayesian criteria. The proposed estimation method, on one hand, avoids the intractable computational difficulty when β is treated as an unknown parameter. It is appealing to assign discrete points for the parameter β which includes specific models of the EEGNB family: exp-W, MFHT-W, EGGe models for $\beta = -1, \beta = 0$ and $\beta = 1$, respectively (Yin & Ibrahim, 2005).

For the purpose of Bayesian statistical inference, we consider the baseline ExpW distribution. Consider a random sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ of size n from the EEWNB model with parameters $(\alpha, \rho, \beta, \gamma_1, \gamma_2)$. On the basis of this random sample and (14), the log-likelihood function is

$$\begin{aligned} l(\alpha,\rho,\gamma_1,\gamma_2) &= \log(L(\alpha,\rho,\gamma_1,\gamma_2 \mid \boldsymbol{x},\beta)) \\ &= n\log(\alpha) + \sum_{i=1}^{n}[(\alpha-1)\log(P(G(x_i;\gamma_1,\gamma_2);\beta,\rho)) \\ &+ \log(p(G(x_i;\gamma);\beta,\rho)) + \log(g(x_i))], \end{aligned} \tag{51}$$

for $\beta \in A = \{-1, -0.8. -0.6, -0, 4, -0.2, 0.0, 02., 0.4, 0.6, -.8, 1.0\}$. We can take $\alpha,\ \rho,\ \gamma_1$ and γ_2 to have independent non-informative gamma priors. The joint posterior distribution of $(\alpha, \rho, \gamma_1, \gamma_2)$ given $\boldsymbol{x}$ and $\beta \in A$ is thus given by

$$\pi(\alpha,\rho,\gamma_1,\gamma_2 \mid \boldsymbol{x},\beta) \propto L(\alpha,\rho,\gamma_1,\gamma_2 \mid \boldsymbol{x},\beta)\pi(\alpha)\pi(\rho)\pi(\gamma_1)\pi(\gamma_2),\ \beta \in A. \tag{52}$$

All the Bayesian computations were based on posterior samples recorded every 10th iterations from 100,000 Gibbs samples after a burn-in of 2,000 samples. We assigned the following noninformative prior distributions: $\alpha \sim \text{Gamma}(,),\ \rho \sim \text{Gamma}(,),\ \gamma_1 \sim \text{Gamma}(,)$ and $\gamma_2 \sim \text{Gamma}(,)$. Markov Chain Monte Carlo convergence was monitored according to the methods recommended and the chains converged very quickly and the parameters mixed very well. MCMC computations were implemented in the OpenBUGS 3.0.3 system, freely available at http://mathstat.helsinki.fi/openbugs/.

There are several methodologies for comparing competing models for a given dataset. We consider three approaches to Bayesian model selection, the Deviance Information Criterion (DIC, Spiegelhalter *et al.*, 2002), the Expected Akaike Information Criteria (EAIC) and the Expected Bayesian Information Criteria (EBIC), which were proposed by Brooks (2002) and used for example in Bolfarine & Bazan (2010). These model criteria are simple to compute as the relevant quantities can be calculated directly from the MCMC output.

In order to obtain this, the following criteria are considered

$$E\left[D(\boldsymbol{\eta})\right],\quad D(E[\boldsymbol{\eta}])\ \text{ and }\ \rho_D = E\left[D(\boldsymbol{\eta})\right] - D(E[\boldsymbol{\eta}]),$$

which represent the deviance's posterior mean. This deviance is obtained by considering the mean values of the posteriori of the parameters in the model, and the effective number of parameters as given in Spiegelhalter *et al.* (2002).

These quantities can be estimated by using the MCMC output, considering the value of

$$\bar{D} = \frac{1}{B}\sum_{b=1}^{B} D\left(\boldsymbol{\eta}^b\right), \quad \widehat{D} = D\left(\frac{1}{B}\sum_{b=1}^{B}\boldsymbol{\eta}^b\right) \quad \text{and} \quad \widehat{\rho_D} = \bar{D} - \widehat{D},$$

respectively, where B represents the number of iterations, and $D(\boldsymbol{\eta}^b) = -2\log L(\boldsymbol{\eta}^b \mid \boldsymbol{Y})$ is the value of the deviance in the $b-$iteraction.

The EAIC, EBIC and DIC statistics can be estimated thorough the MCMC output by considering

$$\widehat{EAIC} = \bar{D} + 2p^{\star}, \quad \widehat{EBIC} = \bar{D} + p^{\star}log(n) \text{ and } \widehat{DIC} = \bar{D} + \widehat{\rho_D} = 2\bar{D} - \widehat{D},$$

respectively, where $p^{\star}$ is the number of parameters in the model and n is the total number of observations. Smaller values of $\bar{D}$, DIC, EBIC and EAIC indicate better model fit.

We fit several special models of the EEGBN model. The models taken are: Exp-W, Marshall-Olkin Extended Exponentiated Weibull (MOEEW), Marshall-Olkin Extended Weibull (MOEW), Marshall-Olkin Extended Exponential (MOEE), Exponentiated Weibull (exp-W), Weibull (W) and Exponential (E). To carry out a model comparison, we computed ($\bar{D}$), DIC, EAIC and EBIC for each model. The MOEW model presents the best fit considering only EAIC, and the MOEEW gives the best fit considering $\bar{D}$, DIC and EBIC with an intervention or effect of the first-activation mechanism to justify the occurrence of the event of interest given by 0.57 (see Table 1). This value is suggesting that the first-activation could not be responsible for the occurrence of the event of interest which is expected in a overdispersion scenery ($\beta = 1$). The posterior mean and standard deviation, and 95% credible interval of the parameters for the models are also reported in Table 1. We do not report the results for β values different from -1,0,1 because they did not showed the best fits.

10 Conclusions

An important problem in survival applications is to obtain new distributions of the time for the occurrence of the event of interest based an activation mechanism that only depends of the risk factors and a latent threshold variable and not on the proposed parent distributions. This random time variable is known as the first passage time, or the time of tumor detection in clinical area as detailed in Yakovlev & Tsodikov (1996). Under the proposed latent model it is possible to unify alternative models for the first hitting random time with a strong biological interpretation. We propose and study some structural

Table 1: Comparison of different models for Nerve data

EEGNB		α	γ_1	γ_2	ρ	$\omega^\star$	p	q	*Dbar*	*DIC*	*EAIC*	*EBIC*
G=exp-W	Mean	2.233	0.732	10.043	-	0	-	-	13870	13870	13876	13890.05008
$\beta = -1$	SD	0.5458	0.0868	3.0402	-	-	-	-	-	-	-	-
-	P2.5	1.365	0.582	5.81	-	-	-	-	-	-	-	
-	P97.5	3.511	0.925	17.531	-	-	-	-	-	-	-	-
G=W	Mean	1	1.08	4.44	-	0	-	-	13880	13882	13884	13893.36672
$\beta = -1$	SD		0.029	0.155	-	-	-	-	-	-	-	-
	P2.5	-	1.03	4.15	-	-	-	-	-	-	-	
	P97.5	-	1.14	4.76	-	-	-	-	-	-	-	-
G=E	Mean	1	1	4.58	-	0	-	-	13887.1	13888.1	13889.1	13894.73344
$\beta = -1$	SD	-	-	0.158	-	-	-	-	-	-	-	-
	P2.5	-	-	4.28	-	-	-	-	-	-	-	-
-	P97.5	-	–	4.68	-	-	-	-	-	-	-	-
Exp-W	Mean	1.864	0.878	6.253	0.827	1	0.587	0.413	13868	13865	13876	13893.73344
$\beta = 0$	SD	0.512	0.132	2.266	0.483	-	0.156	0.156	-	-	-	-
G=W	P2.5	1.094	0.6512	3.1789	0.0655	-	0.3505	0.0615	-	-	-	-
	P97.5	3.13	1.152	11.831	1.853	-	0.939	0.65	-	-	-	-
MOEEW	Mean	1.616	1.047	4.394	1.45	0.57	0.468	0.532	13867	13850	13875	13888.05008
$\beta = 1$	SD	0.499	0.207	1.755	0.98	-	0.169	0.169	-	-	-	-
G=W	P2.5	0.846	0.74	2.038	0.146	-	0.208	0.127	-	-	-	-
	P97.5	2.716	1.509	8.512	3.802	-	0.873	0.792	-	-	-	-
MOEW	Mean	1	1.335	2.582	2.259	0.46	0.334	0.666	13868	13870	13874	13902.36672
$\beta = 1$	SD		0.0705	0.3513	0.9597	-	0.0995	0.0995	-	-	-	-
G=W	P2.5	-	1.196	1.974	0.739	-	0.183	0.425	-	-	-	-
	P97.5	-	1.476	3.408	4.475	-	0.575	0.817	-	-	-	-
MOEE	Mean	1	1	4.4429	0.0614	0.97	0.9447	0.0553	13889	13890	13893	13902.36672
$\beta = 1$	SD	-	-	0.2017	0.0575	-	0.0473	0.0473	-	-	-	-
G=E	P2.5	-	-	4.014	0.00163	-	0.82649	0.00163	-	-	-	
-	P97.5	-	-	-	0.21	-	0.998	0.174	-	-	-	-

properties of the extended exponentiated-G-negative binomial (EEGNB) distribution including moments, quantile and generating functions and mean deviations. Applications from the Bayesian viewpoint to real data sets in Section 9 indicate that the new distribution can be used quite effectively, and it encompass most recent models in survival analysis. The parameters θ and ω involved in this approach are not identifiable. However, it is possible to obtain some value of ω in order to get some idea of the effect of the first activation mechanism on the baseline distribution $G(x)$. Another possibility to avoid this problem of identifiability is to provide some link functions involving covariate variables. We believe that the random activation mechanism to extend the baseline distribution, its link with the Box-Cox transformation and the Bayesian machinery for inferential problems proposed herein are a new way to obtain and apply alternative models with potential use in many other areas such as econometrics, sociology and reliability.

References

Aarts, R. M. (2000). Lauricella functions. from mathworld, a wolfram web resource, created by eric w. weisstein. (http://mathworld.wolfram.com/lauricellafunctions.html).

Barreto, W. & Simas, A. B. (2012). The exp-G family of probability distributions. *Brazilian Journal of Probability and Statistics, DOI: 10.1214/11BJPS157*.

Barreto-Souza, W., Moraisa, A. & Cordeiro, G. M. (2011). The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation*, 81, 645–657.

Bolfarine, H. & Bazan, J. (2010). Bayesian estimation of the logistic positive exponent irt model. *Journal of Educational Behavioral Statistics*, 35, 693–713.

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. *Journal Royal Statistical Society, Series B*, 64, 616–618.

Castro, M., Cancho, V. & Rodrigues, J. (2010). A note on a unified approach for cure rate models. *Brazilian Journal of Probability and Statistics*, 24, 100–103.

Cordeiro, G. M. & Lemonte, A. (2013). On the Marshall-Olkin extended Weibull distribution. *Statistical Papers*, 54, 333–353.

Cordeiro, G. M. & Nadarajah, S. (2011). Closed-form expressions for moments of a class of beta generalized distributions. *Brazilian Journal of Probability and Statistics*, 25, 14–33.

Cox, D. R. & Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen & Co. Ltd., London.

Feller, W. (1943). On a general class of contagious distributions. *Annals of Mathematical Statistics*, 14, 389–400.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, volume I. Wiley, John Wiley & Sons, New York, Third edition.

Ferreira, J. & Steel, M. (2006). A constructive representation of univariate skewed distributions. *, Journal of the American Statistical Association*, 101, 823–829.

Gilchrist, W. (2000). *Statistical Modelling with Quantiles functions*. Chapman & Hall, New York, NY.

Gupta, R. D. & Kundu, D. (2001a). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal*, 43, 117–130.

Kumaraswamy, P. (1980). Generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 3, 107–128.

Kundu, D. & Gupta, R. D. (2011). An extension of the generalized exponential distribution. *Statistical Methodology*, 1, 485–491.

Lemonte, A. & Cordeiro, G. (2011). An extended lomax distribution. *Statistics, DOI: 10.1080/02331888.2011.568119*.

Lemonte, A., Barreto-Souza, W. & Cordeiro, G. (2013). The exponentiated kumaraswamy distribution and its log-transform. *Brazilian Journal of Probability and Statistics*, 27, 31–53.

Lemonte, A. J. (2013). A new extended Birnbaum-Saunders regression model for lifetime modeling. *Computational Statistics and Data Analysis*, 64, 34–50.

Marshall, A. & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 641–652.

Mudholkar, G. S., Srivastava, D. K. & Freimer, M. (1995). The exponentiated Weibull family. *Technometrics*, 37, 436–445.

Nadarajah, S. & Kotz, S. (2006a). The exponentiated type distributions. *Acta Applicandae Mathematicae*, 92, 97–111.

Nadarajah, S. & Kotz, S. (2006b). The beta exponential distribution. *Reliability Engineering & System Safety*, 91, 689–697.

Nadarajah, S., Nassiri, V. & Mohammadpour (2009). Truncated-exponential skew-symmetric distributions. *Technical Report*.

Neyman, J. (1939). On a new class of contagious distribution applicable in entomology and bacteriology. *Annals of Mathematical Statistics*, 10, 35–57.

Pewsey, A. & Gomes, H. (2013). Likelihood-based inference for power distributions. *Test, DOI 10.1007/s11749-011-0280-0*.

Prudnikov, A. P., Brychkov, Y. A. & Marichev, O. I. (1986). *Integrals and Series*. Gordon and Breach, Amsterdam.

Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2008). On the unification of the long-term survival models. *Statistics & Probability Letters*, 39, 753–759.

Rodrigues, J., Balakrishnan, N., Cordeiro, G. M. & de Castro, M. (2011). A unified view on lifetime distributions arising from selection mechanisms. *Computational Statistics & Data Analysis*, 55, 3311–3319.

Sakia, R. (1992). The Box-Cox transformation technique: a review. *The Statistician*, 41, 169–178.

Spiegelhalter, D. J., Best, N., Carlin, B. & Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64, 583–640.

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific Publishers, Singapore.

Yin, G. & Ibrahim, J. (2005). Cure rate models: unified approach. *The Canadian Journal of Statistics*, 33, 559–570.