A New Extended G Family of Continuous Distributions with Mathematical Properties, Characterizations and Regression Modeling

G. G. Hamedani
Department of Mathematics, Statistics and Computer Science, Marquette University, USA.
gholamhoss.hamedani@marquette.edu

Emrah Altun
Department of Statistics, Bartin University, Bartin 74100, Turkey.
emrahaltun123@gmail.com

Mustafa Ç. Korkmaz
Department of Measurement and Evaluation, Artvin Çoruh University, Artvin, Turkey.
mcagatay@artvin.edu.tr

Haitham M. Yousof
Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt.
haitham.yousof@fcom.bu.edu.eg

Nadeem Shafique Butt
Department of Family and Community Medicine, Rabigh King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia.
nshafique@kau.edu.sa

Abstract

We propose a new extended G family of distributions. Some of its structural properties are derived and some useful characterization results are presented. The maximum likelihood method is used to estimate the model parameters by means of graphical and numerical Monte Carlo simulation study. The flexibility of the new family illustrated by means of two real data sets. Moreover, we introduce a new log-location regression model based on the proposed family. The martingale and modified deviance residuals are defined to detect outliers and evaluate the model assumptions. The potentiality of the new regression model is illustrated by means of a real data set.

Keywords: Characterization; Simulation; Regression Modeling.

1. Introduction

In the statistical literature, there are many G families which have been introduced based on the bounded models such as the beta-G family of distributions by Eugene (2002), the Kumaraswamy G family of distributions by Cordeiro and de Castro (2011) and the Topp-Leone generated family of distributions by Rezaei et al. (2017) among others. Motivated by Cordeiro and de Castro (2011) and Rezaei et al. (2017), we defne the new Extended (NE) G family of distributions. Let T be a random variable (r.v.) with the following cumulative distribution function (cdf)

$$H(t; \alpha, \beta) = 2^{\alpha^{-1}} t \left[1 + t^{\alpha}\right]^{-\alpha^{-1}} \left[1 + (1-t)^{\beta}\right]^{-\beta^{-1}}; \ 0 \le t \le 1. \qquad (1)$$

Where α and β are positive parameters. Let $h(t; \alpha, \beta)$ be the corresponding probability density function (pdf)

Then new cdf is given by

$$F(x) = F_{NEG}(x; \alpha, \beta, \xi) = \int_0^{G(x;\underline{\eta})} h(t; \alpha, \beta) dt$$

$$= \frac{2^{\alpha^{-1}} G(x;\underline{\eta})}{\left[1 + G(x;\underline{\eta})^{\alpha}\right]^{\alpha^{-1}} \left[1 + \overline{G}(x;\underline{\eta})^{\beta}\right]^{\beta^{-1}}}, x \in \mathbb{R}. \tag{2}$$

The pdf of the new family corresponding to (2) is given by

$$f(x) = f_{NEG}(x; \alpha, \beta, \xi) = \frac{2^{\alpha^{-1}} g(x;\underline{\eta}) \left\{ 1 + \overline{G}(x;\underline{\eta})^{\beta-1} \left[1 + G(x;\underline{\eta})^{\alpha+1} \right] \right\}}{\left[1 + G(x;\underline{\eta})^{\alpha}\right]^{\alpha^{-1}+1} \left[1 + \overline{G}(x;\underline{\eta})^{\beta}\right]^{\beta^{-1}+1}}, x \in \mathbb{R}, \tag{3}$$

Where $g\left(x; \underline{\eta}\right)$ is the pdf corresponding to $G\left(x; \underline{\eta}\right)$.

Hereafter, a r.v. X with pdf (3) will be denoted by $X \sim$ NE-G (α, β, ξ). Further, we may write $g\left(x; \underline{\eta}\right)$ as $g(x)$ and $G\left(x; \underline{\eta}\right)$ as $G(x)$ sometimes. The corresponding reverse hazard function is

$$r_F(x) = r_F(x; \alpha, \beta, \xi) = \frac{g\left(x; \underline{\eta}\right) \left\{ 1 + \overline{G}\left(x; \underline{\eta}\right)^{\beta-1} \left[1 + G\left(x; \underline{\eta}\right)^{\alpha+1} \right] \right\}}{G(x) \left[1 + G\left(x; \underline{\eta}\right)^{\alpha}\right] \left[1 + \overline{G}\left(x; \underline{\eta}\right)^{\beta}\right]}, \quad x \in \mathbb{R}.$$

The differential equation in terms of the reverse hazard function is

$$r_F'(x) - \frac{g'\left(x; \underline{\eta}\right)}{g\left(x; \underline{\eta}\right)} r_F(x) = g\left(x; \underline{\eta}\right) \frac{d}{dx} \left\{ \frac{1 + \overline{G}\left(x; \underline{\eta}\right)^{\beta-1} \left[1 + G\left(x; \underline{\eta}\right)^{\alpha+1} \right]}{G(x) \left[1 + G\left(x; \underline{\eta}\right)^{\alpha}\right] \left[1 + \overline{G}\left(x; \underline{\eta}\right)^{\beta}\right]} \right\}, x \in \mathbb{R},$$

with boundary condition

$$\lim_{x \to \infty} r_F(x) = \lim_{x \to \infty} g\left(x; \underline{\eta}\right).$$

For $\alpha = \beta$, we have

$$F(x; \alpha) = \frac{2^{\alpha^{-1}} G\left(x; \underline{\eta}\right)}{\left\{ [1 + G\left(x; \underline{\eta}\right)^{\alpha}] \left[1 + \overline{G}\left(x; \underline{\eta}\right)^{\alpha}\right] \right\}^{\alpha^{-1}}}, \quad x \in \mathbb{R}.$$

There is a special characterization for this submodel. For $\alpha = \beta = 1$, we have

$$F(x) = \frac{2 G\left(x; \underline{\eta}\right)}{[1 + G\left(x; \underline{\eta}\right)] \left[2 - G\left(x; \underline{\eta}\right)\right]}, \quad x \in \mathbb{R}.$$

We are motivated to introduce the new family since it exhibits the increasing, decreasing, upside-down as well as the bathtub hazard rates as illustrated in Figure 2. It is shown in Section 2 that the new family can be viewed as a mixture representation of the exponentiated G densities. The new family can also be viewed as a suitable model for fitting the right-skewed, the symmetric and bimodal data (see Section 9). The NE-Weibull and NE-normal models outperform several of the well-known lifetime Weibull and -normal distributions with respect to the 1st and the 2nd real data applications

respectively. The new log-location regression model based on the NE Weibull distribution provides better fits than the log-Weibull regression model in modelingthe HIV survival data.

2. Useful expansions

First, we consider two power series

$$(1 + z)^{-a} = \sum_{i=0}^{\infty} (-1 + z)^i 2^{-a-i} \binom{-a}{i} \qquad (4)$$

and

$$(1 - z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{j! \Gamma(b-j)} z^j, |z| < 1 \ and \ b > 0 \ is \ a \ real \ non-integer. \quad (5)$$

The cdf in (2) can be written as

$$F_{new}(x) = 2^{\alpha^{-1}} G\left(x; \underline{\eta}\right) \underbrace{\left[1 + G\left(x; \underline{\eta}\right)^\alpha\right]^{-\alpha^{-1}}}_{A} \left[1 + \overline{G}\left(x; \underline{\eta}\right)^\beta\right]^{-\beta^{-1}}. \qquad (6)$$

By Expanding the quantity A in power series (4), the NE-G cdf in (6) can be expressed as

$$F_{new}(x) = G\left(x; \underline{\eta}\right) \sum_{i=0}^{\infty} \underbrace{\left[-1 + G\left(x; \underline{\eta}\right)^\alpha\right]^i}_{B} 2^{-i} \binom{-\alpha^{-1}}{i} \left[1 + \overline{G}\left(x; \underline{\eta}\right)^\beta\right]^{-\beta^{-1}}.$$

Second, using the binomial expansion to expand B, the last equation can be expressed as

$$F_{new}(x) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} 2^{-i} \binom{-\alpha^{-1}}{i} \binom{i}{j} G\left(x; \underline{\eta}\right)^{1+(i-j)\alpha} \underbrace{\left[1 + \overline{G}\left(x; \underline{\eta}\right)^\beta\right]^{-\beta^{-1}}}_{C}.$$

Third, expanding the quantity C in power series (4), the last equation can be written as

$$F_{new}(x) = \sum_{i,l=0}^{\infty} \sum_{j=0}^{i} 2^{-\beta^{-1}-i-l} \binom{-\alpha^{-1}}{i} \binom{i}{j} G\left(x; \underline{\eta}\right)^{1+(i-j)\alpha}$$

$$\times \underbrace{\left[-1 + \overline{G}\left(x; \underline{\eta}\right)^\beta\right]^l}_{D} \binom{-\beta^{-1}}{l}.$$

Fourthly, using the binomial expansion to expand quantity D, the above equation will be

$$F_{new}(x) = \sum_{i,l=0}^{\infty} \sum_{j=0}^{i} \sum_{h=0}^{l} 2^{-\beta^{-1}-i-l} \binom{-\alpha^{-1}}{i} \binom{-\beta^{-1}}{l} \binom{i}{j} \binom{l}{h}$$

$$\times G\left(x; \underline{\eta}\right)^{1+(i-j)\alpha} \underbrace{\overline{G}\left(x; \underline{\eta}\right)^{(l-h)\beta}}_{E}.$$

Fifthly, applying (5) for E, we have

$$F_{new}(x) \sum_{i,l,k=0}^{\infty} \sum_{j=0}^{i} \sum_{h=0}^{l} (-1)^k 2^{-\beta^{-1}-i-l} G\left(x; \underline{\eta}\right)^{1+(i-j)\alpha+k} \binom{-\alpha^{-1}}{i} \binom{-\beta^{-1}}{l} \binom{i}{j} \binom{l}{h} \binom{[l-h]\beta}{k}.$$

Finally

$$F_{new}(x) = \sum_{i,k=0}^{\infty} \sum_{j=0}^{i} v_{i,k,j} \Pi_{1+(i-j)\alpha+k}(x), \qquad (7)$$

where

$$v_{i,k,j} = \sum_{l=0}^{\infty} \sum_{h=0}^{l} (-1)^k 2^{-\beta^{-1}-i-l} \binom{-\alpha^{-1}}{i} \binom{-\beta^{-1}}{l} \binom{i}{j} \binom{l}{h} \binom{[l-h]\beta}{k},$$

and $\Pi_\delta(x) = G\left(x; \underline{\eta}\right)^\delta$ is the cdf of Exp-G with power parameter δ. The corresponding pdf can be written as

$$f_{new}(x) = \sum_{i,k=0}^{\infty} \sum_{j=0}^{i} v_{i,k,j} \pi_{1+(i-j)\alpha+k}(x), \tag{8}$$

where $\pi_\delta(x) = \delta g(x, \underline{\eta}) G(x, \underline{\eta})^{\delta-1}$ is the pdf of the Exp-G distribution with power parameter δ. Equation (7) and (8) reveal that pdf of NE-G is a linear combination of Exp-G densities. Thereby, some properties of the proposed family such as moments and generating function can be determined from those of Exp-G distribution.

3. Two Special Members of the Family

The NE-G family generates alternative extended distributions. Now, we present the two important sub-models of this new family.

3.1 The NE-normal (NEN) distribution

As a first example, we extend the ordinary normal distribution which has symmetrical and bell-shaped. We define the NEN distribution by taking $G(x; \mu, \sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ for $x, \mu \in \mathfrak{R}$ and $\sigma > 0$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution. Hence, the cdf of the NEN distribution is given by

$$F(x; \alpha, \beta, \mu, \sigma) = \frac{2^{1/\alpha} \Phi\left(\frac{x-\mu}{\sigma}\right)}{\left[1 + \Phi\left(\frac{x-\mu}{\sigma}\right)^\alpha\right]^{1/\alpha} \left[1 + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\beta\right]^{1/\beta}}.$$

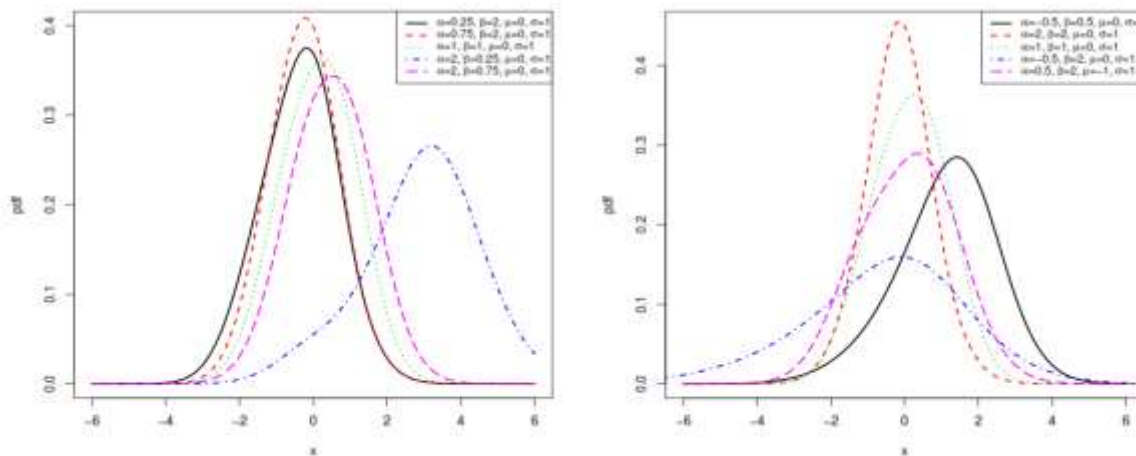Some possible plots of the NEN density for selected parameter values are displayed in Figure 1.

Figure 1: The pdf plots of NEN distributions for some selected parameter values.

740

Pak.j.stat.oper.res. Vol.XIV No.3 2018 pp737-758

3.2 The NE-Weibull (NEW) distribution

As the second example, we define the NEW distribution by employing the Weibull cdf with shape $\gamma > 0$ and scale $\theta > 0$ parameters, defined by $G(x; \theta, \gamma) = 1 - \exp(-(\theta x)^\gamma)$ (for $x > 0$). The cdf of the NEW distribution is given by

$$F(x; \alpha, \beta, \theta, \gamma) = \frac{2^{1/\alpha}\left(1 - e^{-(\theta x)^\gamma}\right)}{[1 + (1 - e^{-(\theta x)^\gamma})^\alpha]^{1/\alpha}[1 + e^{-\beta(\theta x)^\gamma}]^{1/\beta}}.$$

Possible plots of the NEW density and hrf for selected parameter values are displayed in Figure 2. From these Figures, we can say that pdf shapes of the NEW distribution can be bi-modal, uni-modal and decreasing-increasing-decreasing shaoed. Also, its hrf can be both monotone and non-monotone shaped.
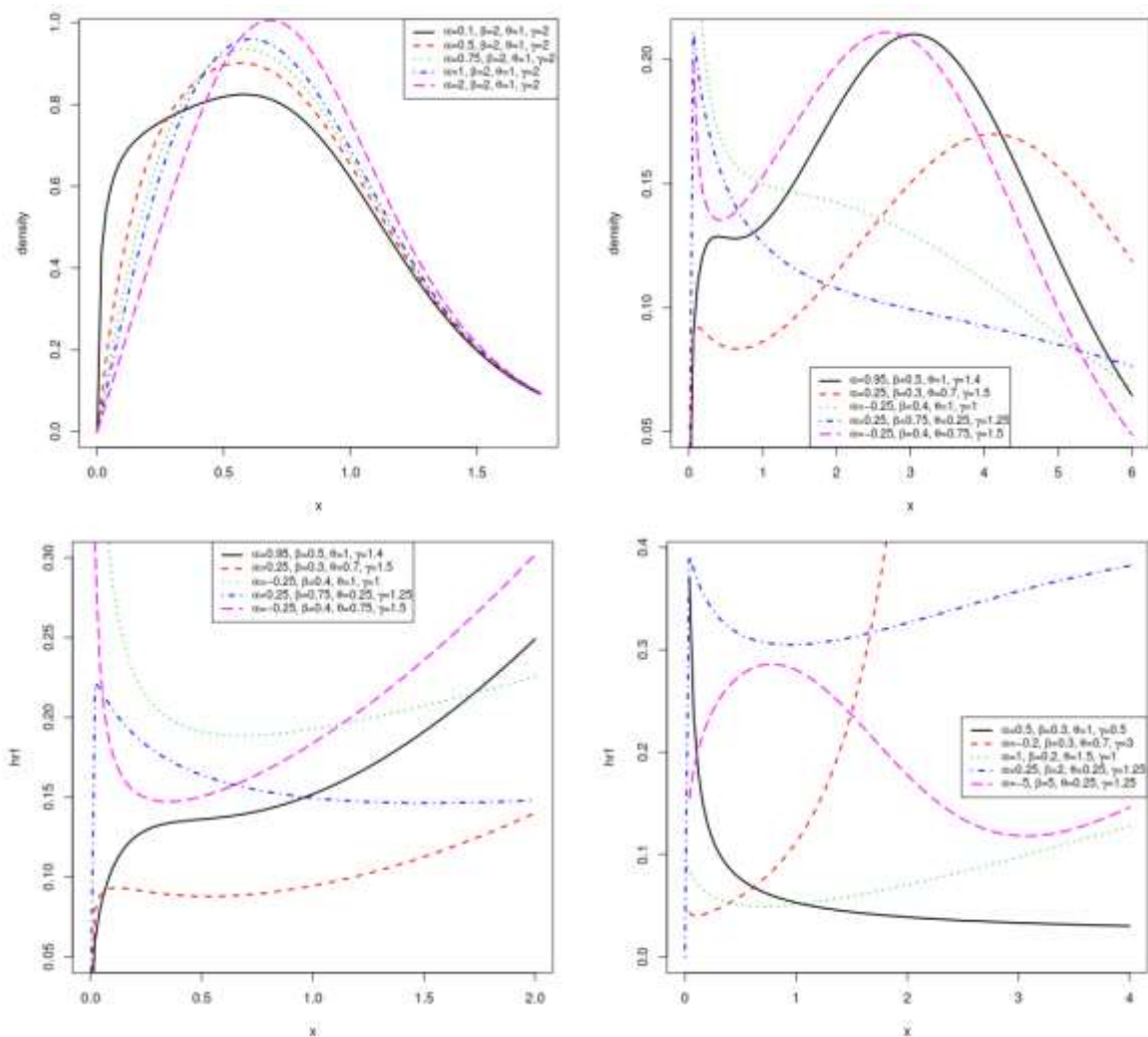

Figure 2: Possible pdf and hrf plots of NEW distributions for some selected parameter values.

4. Mathematical properties

4.1 Moments, incomplete moments and generating function

The rth ordinary moment of X is given by $\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$. Then we obtain

$$\mu'_r = \sum_{i,k=0}^{\infty} \sum_{j=0}^{i} \nu_{i,k,j} Y_{1+(i-j)\alpha+k}^r. \tag{9}$$

Henceforth, Y_δ denotes the Exp-G model with power parameter δ. For $\delta > 0$, we have $E(Y_\delta^r) = \delta \int_{-\infty}^{\infty} x^r g\left(x; \underline{\eta}\right) G\left(x; \underline{\eta}\right)^{\delta-1} dx$, which can be computed numerically in terms of the baseline quantile function (qf) $Q_G\left(u; \underline{\eta}\right) = G^{-1}\left(u; \underline{\eta}\right)$ as $E(Y_\delta^n) = \delta \int_0^1 Q_G\left(u; \underline{\eta}\right)^n u^{\delta-1} du$. Setting $r = 1$ in (9), we have the mean of X. The last integration can be carried out numerically for most parent distributions. The skewness and kurtosis measures can be calculated from the ordinary moments using well-known relationships. The nth central moment of X, say M_n, is

$$M_n = E(X - \mu)^n = \sum_{h=0}^{n} (-1)^h \binom{n}{h} (\mu'_1)^n \mu'_{n-h}.$$

The rth incomplete moment, say $I_r(t)$, of X can be expressed from (6) as

$$I_r(t) = \int_{-\infty}^{t} x^r f(x) dx = \sum_{i,k=0}^{\infty} \sum_{j=0}^{i} \int_{-\infty}^{t} \nu_{i,k,j} x^r \pi_{1+(i-j)\alpha+k}(x) dx. \tag{10}$$

The mean deviations about the mean and about the median are

$$E(|X - E(X)|) = 2\mu'_1 F\left(E(X)\right) - 2I_1\left(E(X)\right),$$
$$E(|X - Q(0.5)|) = E(X) - 2I_1\left(Q(0.5)\right),$$

where $Q(0.5) = Median(X)$ is the median, $F(E(X)) = F(\mu'_1)$ is easily calculated using (2) and $I_1(t)$ is the first incomplete moment given by (10) with $r = 1$. A general equation for $I_1(t)$ can be derived from (10) as

$$I_1(t) = \sum_{i,k=0}^{\infty} \sum_{j=0}^{i} \nu_{i,k,j} J_{1+(i-j)\alpha+k}(x)|_{-\infty}^t,$$

where $J_\delta(x)|_{-\infty}^t = \int_{-\infty}^{t} x \pi_\delta(x) dx$ is the first incomplete moment of the Exp-G model. The moment generating function (mgf) $M_X(t) = E(e^{tX})$ of X can be derived using equation (9) as

$$M_X(t) = \sum_{i,k=0}^{\infty} \sum_{j=0}^{i} \nu_{i,k,j} M_{1+(i-j)\alpha+k}(t),$$

where $M_\delta(t)$ is the mgf of Y_δ. Hence, $M_X(t)$ can be determined from the Exp-G generating function.

4.2 Moments of residual and reversed residual life

The nth moment of the residual life say, $z_n(t) = E[(X - t)^n | X > t], n = 1, 2, ...$ uniquely determines $F(x)$. The nth moment of the residual life of X is given by

$$z_n(t) = [1 - F(t)]^{-1} \int_t^\infty (x - t)^n dF(x).$$

Therefore

$$z_n(t) = [1 - F(t)]^{-1} \sum_{i,k=0}^\infty \sum_{j=0}^i \sum_{r=0}^n v_{i,k,j} (1-t)^n \int_t^\infty x^r \pi_{1+(i-j)\alpha+k}(x) dx.$$

The n th moment of the reversed residual life say, $Z_n(t) = E[(t-X)^n | X \leq t]$ for $t > 0$ and $n = 1,2,...$ uniquely determines $F(x)$. We obtain

$$Z_n(t) = F(t)^{-1} \int_0^t (t - x)^n dF(x).$$

Then, the n th moment of the reversed residual life (RRL) of X becomes

$$Z_n(t) = F(t)^{-1} \sum_{i,k=0}^\infty \sum_{j=0}^i \sum_{r=0}^n v_{i,k,j} (-1)^r \binom{n}{r} t^{n-r} \int_0^t x^r \pi_{\alpha(j+1)+k}(x) dx.$$

The mean residual life (MRL) function or the life expectation at age t defined by $z_1(t) = E[(X-t)|X > t]$, which represents the expected additional life length for a unit which is alive at the age t. The MRL of X can be obtained by setting $n = 1$ in $z_n(t)$ equation.

5. Characterizations

This section is devoted to the characterizations of the NE-G family of distributions in different directions: (i) based on the ratio of two truncated moments; (ii) in terms of the reverse hazard function. Note that (i) can be employed also when the cdf does not have a closed form. We would also like to mention that due to the nature of NE-G family of distributions, our characterizations may be the only possible ones. We present our characterizations $(i) - (ii)$ in two subsections.

5.1 Characterizations based on two truncated moments

This subsection deals with the characterizations of NE-G family of distributions based on the ratio of two truncated moments. Our first characterization employs a theorem due to Glänzel (1987), see Theorem 1 of Appendix A . The result, however, holds also when the interval H is not closed, since the condition of the Theorem is on the interior of H.

Proposition 5.1: Let $X: \Omega \to \mathbb{R}$ be a continuous random variable and let

$$q_1(x) = G(x;\eta)^{\alpha-1}\{1 + \overline{G}(x;\eta)^{\beta-1}[1 + G(x;\eta)^{\alpha+1}]\}^{-1}[1 + \overline{G}(x;\eta)^\beta]^{\frac{1}{\beta}+1},$$

and

$$q_2(x) = 2^{\frac{1}{\alpha}+1} q_1(x)[1 + G(x;\eta)^\alpha]^{-1/\alpha} \text{ for } x \in \mathbb{R}.$$

The random variable X has pdf (3) if and only if the function ξ defined in Theorem 1 is of the form

$$\xi(x) = 2^{1/\alpha}[1 + G(x;\eta)^\alpha]^{-1/\alpha} + 1, \quad x \in \mathbb{R}.$$

Proof: Suppose the random variable X has pdf (3), then

$$\left(1 - F(x)\right) E[q_1(X)|X \geq x] = 2^{1/\alpha}[1 + G(x;\eta)^\alpha]^{-1/\alpha} - 1, \quad x \in \mathbb{R},$$

and
$$(1 - F(x))E[q_2(X)|X \geq x] = 2^{2/\alpha}[1 + G(x;\eta)^\alpha]^{-2/\alpha} - 1, \quad x \in \mathbb{R}.$$

Further,
$$\xi(x)q_1(x) - q_2(x) = q_1(x)\{1 - 2^{1/\alpha}[1 + G(x;\eta)^\alpha]^{-1/\alpha}\} > 0, \quad for \ x \in \mathbb{R}.$$

Conversely, if ξ is of the above form, then

$$s'(x) = \frac{\xi'(x)q_1(x)}{\xi(x)q_1(x) - q_2(x)} = \frac{-2^{1/\alpha}g(x;\eta)G(x;\eta)^{\alpha-1}[1 + G(x;\eta)^\alpha]^{-\frac{1}{\alpha}-1}}{1 - 2^{1/\alpha}[1 + G(x;\eta)^\alpha]^{-1/\alpha}}, \quad x \in \mathbb{R},$$

and consequently
$$s(x) = -\ln\{1 - 2^{1/\alpha}[1 + G(x;\eta)^\alpha]^{-1/\alpha}\}, \quad x \in \mathbb{R}.$$

Now, according to Theorem 1, X has density (3).

Corollary 5.1: Let $X:\Omega \to \mathbb{R}$ be a continuous random variable and let $q_1(x)$ be as in Proposition 5.1. The random variable X has pdf (3) if and only if there exist functions q_2 and ξ defined in Theorem 1 satisfying the following differential equation

$$\frac{\xi'(x)q_1(x)}{\xi(x)q_1(x) - q_2(x)} = \frac{-2^{\frac{1}{\alpha}}g(x;\eta)G(x;\eta)^{\alpha-1}[1 + G(x;\eta)^\alpha]^{-\frac{1}{\alpha}-1}}{1 - 2^{\frac{1}{\alpha}}[1 + G(x;\eta)^\alpha]^{-\frac{1}{\alpha}}}, \quad x \in \mathbb{R}.$$

Corollary 5.2: The general solution of the differential equation in Corollary 5.1 is

$$\xi(x) = \{1 - 2^{1/\alpha}[1 + G(x;\eta)^\alpha]^{-1/\alpha}\}^{-1} \times$$
$$\left[\int 2^{1/\alpha}g(x;\eta)G(x;\eta)^{\alpha-1}[1 + G(x;\eta)^\alpha]^{-\frac{1}{\alpha}-1}(q_1(x))^{-1}q_2(x)dx + D\right],$$

where D is a constant. We like to point out that one set of functions satisfying the above differential equation is given in Proposition 5.1 with $D = \frac{1}{2}$. Clearly, there are other triplets (q_1, q_2, ξ) which satisfy conditions of Theorem 1.

5.2 Characterization in terms of the reverse hazard function

The reverse hazard function, r_F, of a twice differentiable distribution function, F, is defined as

$$r_F(x) = \frac{f(x)}{F(x)}, \quad x \in support of F.$$

In this subsection we present a characterization of NE-G family of distributions for $\alpha = \beta = 1$, in terms of the reverse hazard function. In this case

$$r_F(x) = \frac{2g(x;\eta)(2 + G(x;\eta)^2)}{G(x;\eta)(1 + G(x;\eta))(2 - G(x;\eta))}, \quad x \in \mathbb{R}.$$

Proposition 5.2. Let $X:\Omega \to \mathbb{R}$ be a continuous random variable. The random variable X has pdf (3), for $\alpha = \beta = 1$, if and only if its reverse hazard function $r_F(x)$ satisfies the following differential equation

$$r_F'(x) - \frac{g'(x;\eta)}{g(x;\eta)}r_F(x)$$

$$= 2g(x;\eta)^2\left\{\frac{2G(x;\eta)^4 - G(x;\eta)^3 + 6G(x;\eta)^2 - 4G(x;\eta) - 2}{G(x;\eta)^2[(1 + G(x;\eta))(2 - G(x;\eta))]^2}\right\}, \quad x \in \mathbb{R},$$

with the boundary condition $\lim_{x\to\infty}r_F(x) = 3\lim_{x\to\infty}g(x;\eta)$.

744

Pak.j.stat.oper.res. Vol.XIV No.3 2018 pp737-758

Proof: If X has pdf (3), then clearly the above differential equation holds. Now, if the differential equation holds, then

$$\frac{d}{dx}\{r_F(x)g(x;\eta)^{-1}\} = 2\frac{d}{dx}\left\{\frac{2 + G(x;\eta)^2}{G(x;\eta)\big(1 + G(x;\eta)\big)\big(2 - G(x;\eta)\big)}\right\},$$

or

$$r_F(x) = \frac{2g(x;\eta)(2 + G(x;\eta)^2)}{G(x;\eta)\big(1 + G(x;\eta)\big)\big(2 - G(x;\eta)\big)}, \quad x \in \mathbb{R},$$

which is the reverse hazard function of (3) with the boundary condition $\lim_{x\to\infty} r_F(x) = 3\lim_{x\to\infty} g(x;\eta)$.

Examples:

a) Take the baseline cdf to be exponential with parameter η , then cdf will be given by

$$F(x) = \frac{2(1 - e^{-\eta x})}{[1 + e^{-\eta x}][2 - e^{-\eta x}]}, \quad x \geq 0.$$

b) Take the baseline cdf to be uniform $(0,1)$, then cdf will have the form

$$F(x) = \frac{2x}{[1+x][2-x]}, \quad 0 \leq x \leq 1.$$

6. Estimation and inference

Several approaches for parameter estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The MLEs enjoy desirable properties and can be used for constructing confidence intervals and also for test statistics. The normal approximation for these estimators in large samples can be easily handled either analytically or numerically. Here, we consider the estimation of the unknown parameters of the new family from complete samples only by maximum likelihood. Let $x_1, \ldots, x_n$ be a random sample from the NE-G models with a $(q + 2) \times 1$ parameter vector $\mathbf{\Phi} = (\alpha, \beta, \underline{\eta}^{\text{ú}})^{\text{ú}}$, where $\underline{\eta}$ is a $q \times 1$ baseline parameter vector. The log-likelihood function for $\mathbf{\Phi}$ is given by

$$\ell_n(\mathbf{\Phi}) = \sum_{i=1}^{n} \log\left\{1 + \overline{G}\left(x_i; \underline{\eta}\right)^{\beta-1}\left[1 + G\left(x_i; \underline{\eta}\right)^{\alpha+1}\right]\right\}$$

$$+ n\alpha^{-1}\log 2 - (\alpha^{-1} + 1)\sum_{i=1}^{n} \log\left[1 + G\left(x_i; \underline{\eta}\right)^{\alpha}\right]$$

$$- (\beta^{-1} + 1)\sum_{i=1}^{n} \log\left[1 + \overline{G}\left(x_i; \underline{\eta}\right)^{\beta}\right] + \sum_{i=1}^{n} \log g\left(x_i; \underline{\eta}\right).$$

The above log-likelihood can be maximized numerically using R (optim), SAS (PROC NLMIXED), or Ox program (sub-routine MaxBFGS), among others.

Details for fitting univariate distributions using maximum likelihood in R for censored or non censored data can be obtained at http://www.inside-r.org/packages/cran/fitdistrplus/docs/mledist.

The log-likelihood function can be maximized by solving the following nonlinear normal equations $U(\alpha) = \frac{\partial \ell_n(\Phi)}{\partial \alpha}, U(\beta) = \frac{\partial \ell_n(\Phi)}{\partial \beta}$ and $U\left(\underline{\eta}_r\right) = \frac{\partial \ell_n(\Phi)}{\partial \underline{\eta}_r}$ (for $r = 1, \ldots, q$). Setting the nonlinear system of equations $U(\alpha) = U(\beta) = U\left(\underline{\eta}_r\right) = 0$ (for $r = 1 = \cdots, q$) and solving them simultaneously yields the MLE $\widehat{\Phi}$. To solve these equations, it is more convenient to use nonlinear optimization methods such as the quasi-Newton algorithm to maximize $\ell_n(\Phi)$ numerically . For interval estimation of the parameters, we can evaluate numerically the elements of the $(q + 2) \times (q + 2)$ observed information matrix $J(\Phi) = \{-\frac{\partial^2}{\partial \Phi_r \Phi_s}[\ell_n(\Phi)]\}$. Under standard regularity conditions when $n \to \infty$, the distribution of $\widehat{\Phi}$ can be approximated by a multivariate normal $N_p(0, J(\widehat{\Phi})^{-1})$ distribution to construct approximate confidence intervals for the parameters. Here, $J(\widehat{\Phi})$ is the total observed information matrix evaluated at $\widehat{\Phi}$. The method of the re-sampling bootstrap can be used for correcting the biases of the MLEs of the model parameters. Good interval estimates may also be obtained using the bootstrap percentile method.

7. Simulation studies

In this Section, we perform two simulation studies by using the new extended normal and Wiebull distributions to illustrate the performance of MLEs corresponding to these distribution. The random numbers generation is obtained by the inverse of their cdfs. All results related to MLEs were obtained using optim-CG routine in the R programme.

7.1 Simulation study 1

In the first simulation study, we obtain the graphical results. We generate $N = 1000$ samples of size $n = 20,25, \ldots ,1000$ from new extended normal distribution with true parameters values $\alpha = 3$, $\beta = 12$, $\mu = 0$ and $\sigma = 0.1$. In this simulation study, we empirically calculate the mean, standard deviations (sd), bias and mean squared error (MSE) of the MLEs. The bias and MSE are calculated by (for $h = \alpha, \beta, \mu, \sigma$)

$$\widehat{Bias}_h = \frac{1}{1000} \sum_{i=1}^{1000} \left(\hat{h}_i - h\right),$$

and

$$\widehat{MSE}_h = \frac{1}{1000} \sum_{i=1}^{1000} \left(\hat{h}_i - h\right)^2$$

respectively. We give results of this simulation study in Figure 3. From Figure 3, we observe that when the sample size increases, the empirical means approach the true parameter value whereas all biases, sds and MSEs approach 0 in all cases.
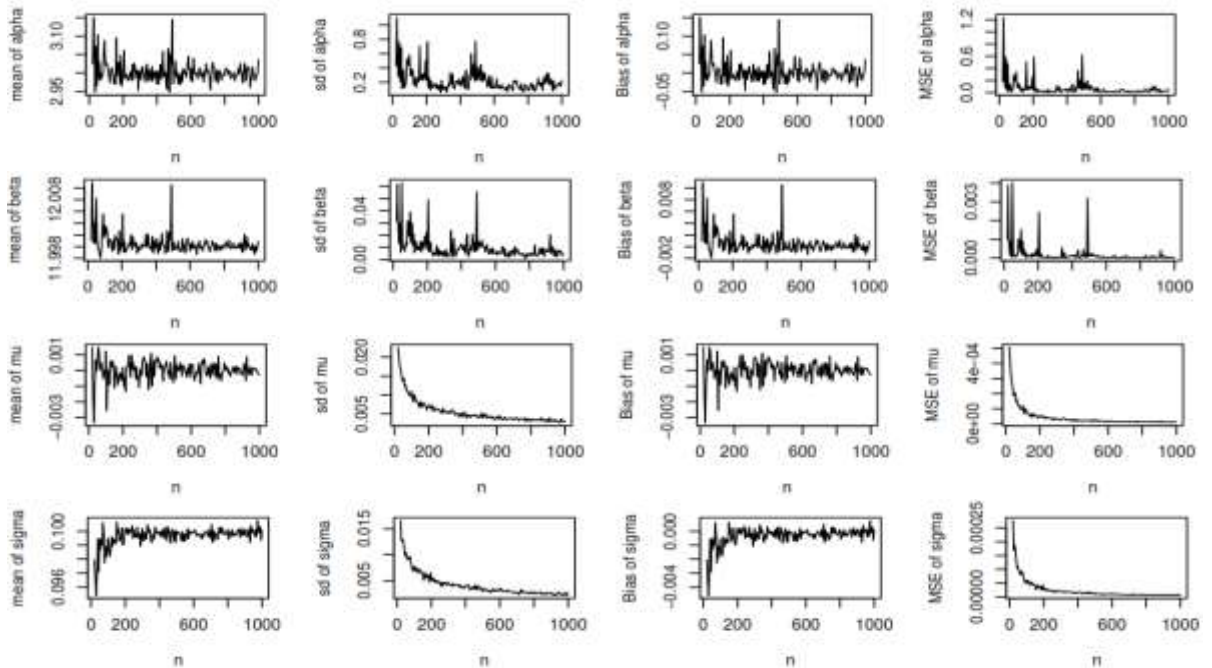
Figure 3: Simulation results of the new extended normal distribution.

7.2 Simulation study 2

In the second simulation study, we generate 1,000 samples of sizes 50, 100 and 200 from selected new extended Weibull distributions. For this simulation study, we obtain the empirical means and sd's of the MLEs. The results of this simulation study are reported in Table 1. Table 1 shows that when the sample size increases, the empirical means approach true parameter value whereas the sds decrease, as expected.

Table 1: Empirical means and standard deviations (in parentheses) for the new extended Weibull distributions.

Parameters	n = 50				n = 100				n = 200			
$\alpha, \beta, \lambda, \gamma$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\gamma}$
0.5,5,0.5,5	0.5592	4.9945	0.5046	5.1534	0.4702	4.9941	0.5012	5.0517	0.5099	5.0001	0.5011	5.0170
	(0.4102)	(0.0686)	(0.0177)	(0.4318)	(0.2294)	(0.0537)	(0.0131)	(0.3009)	(0.1160)	(0.0140)	(0.0093)	(0.1097)
1.5,3,0.1,3	1.5026	2.9980	0.1005	3.0064	1.5030	3.0087	0.1003	3.0381	1.5008	3.0021	0.1006	3.0018
	(0.0534)	(0.0329)	(0.0046)	(0.1870)	(0.0914)	(0.0558)	(0.0035)	(0.1282)	(0.0355)	(0.0298)	(0.0025)	(0.0689)
0.8,15,0.5,5	0.8342	15.0008	0.5043	5.1451	0.7902	15.0002	0.4995	5.0394	0.8112	14.9989	0.5009	5.0124
	(0.3962)	(0.0093)	(0.0167)	(0.4057)	(0.2883)	(0.0033)	(0.0132)	(0.2384)	(0.0857)	(0.0006)	(0.0084)	(0.1026)
1,10,2,1	1.1831	10.0009	2.0111	1.1161	0.9237	10.0373	2.0085	1.0767	1.0539	10.0085	1.9938	1.0466
	(1.2940)	(0.3448)	(0.3924)	(0.1894)	(0.8701)	(0.2054)	(0.2987)	(0.1384)	(0.7699)	(0.1062)	(0.2306)	(0.1256)
1,5,0.1,5	0.9868	4.9986	0.1015	5.0461	0.9904	5.0014	0.0998	4.9930	1.0049	4.9989	0.0999	4.9967
	(0.2149)	(0.0356)	(0.0033)	(0.2812)	(0.0681)	(0.0075)	(0.0022)	(0.0973)	(0.0594)	(0.0120)	(0.0014)	(0.0612)

8. Log-NEW regression model

Let X be a random variable having the NEW density function with four parameters $\alpha > 0, \beta > 0, \theta > 0$ and $\gamma > 0$ as discussed in Subsection 2.3. By replacing $\gamma = 1/\sigma$ and $\theta = \exp(\mu)$, the density function of $Y = \log(X)$ is given by (for $y \in \Re$)

$$f(y) = \frac{\frac{2^{\alpha^{-1}}}{\sigma}\exp\left[\left(\frac{y-\mu}{\sigma}\right)-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\left\{1+\left(\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right)^{\beta-1}\left[1+\left\{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha+1}\right]\right\}}{\left[1+\left\{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha}\right]^{\alpha^{-1}+1}\left[1+\left(\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right)^{\beta}\right]^{\beta^{-1}+1}} \quad ...(11)$$

where $\mu \in \Re$ is the location parameter, $\sigma > 0$ is the scale parameter, $\alpha > 0$ and $\beta > 0$ are the shape parameter. We refer to Equation (11) as the log-NEW (LNEW) distribution and write $Y \sim LNEW(\alpha, \beta, \sigma, \mu)$. Figure 4 displays plots of this density function for some parameter values. They reveal that the LNEW density can be very flexible for modeling left skewed and symmetric data.
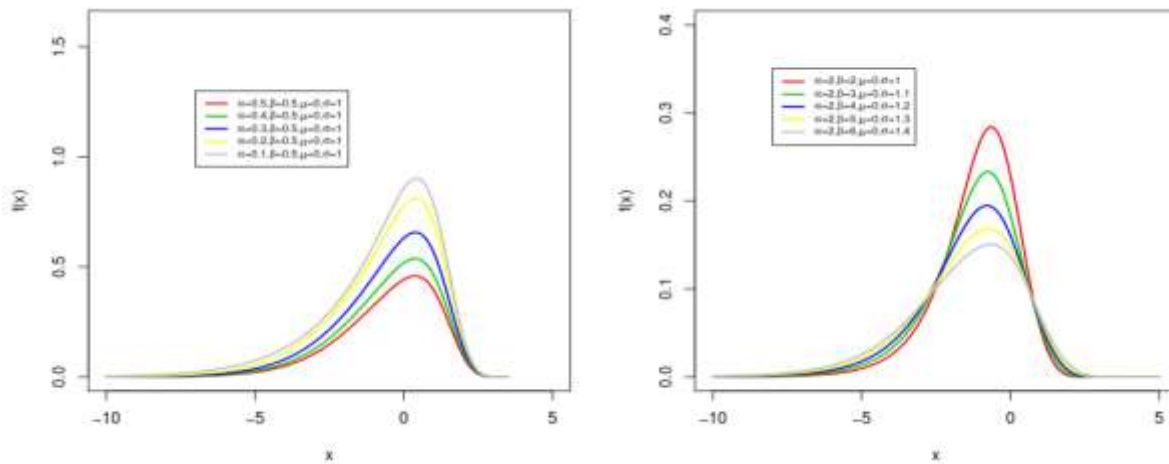

Figure 4: Plots of the LNEW density for selected parameter values.

The corresponding survival function is

$$S(y) = 1 - \frac{2^{\alpha^{-1}}\left\{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}}{\left[1+\left\{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha}\right]^{\alpha^{-1}}+\left[1+\left(\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right)^{\beta}\right]^{\beta^{-1}}} \quad ...(12)$$

The standardized random variable $Z = (Y - \mu)/\sigma$ has density function

$$f(z) = \frac{2^{\alpha^{-1}}\exp[(z)-\exp(z)]\left\{1+(\exp[-\exp(z)])^{\beta-1}\left[1+\{1-\exp[-\exp(z)]\}^{\alpha+1}\right]\right\}}{[1+\{1-\exp[-\exp(z)]\}^{\alpha}]^{\alpha^{-1}+1}\left[1+(\exp[-\exp(z)])^{\beta}\right]^{\beta^{-1}+1}}. \quad (13)$$

Based on the LNEW distribution, a linear location-scale regression model is proposed by linking the response variable y_i and the explanatory variable vector $\mathbf{v}_i^T = (v_{i1}, ..., v_{ip})$ by

$$y_i = \mathbf{v}_i^T \boldsymbol{\beta} + \sigma z_i, \quad i = 1, ..., n, \quad (14)$$

where the random error z_i has density function (13), $\beta = (\beta_1, ..., \beta_p)^T$, $\sigma > 0$, $\alpha > 0$ and $\boldsymbol{\beta} \in \Re^p$ are unknown parameters. The parameter $\mu_i = \mathbf{v}_i^T \beta$ is the location of y_i. The location parameter vector $\mu = (\mu_1, ..., \mu_n)^T$ is represented by a linear model $\mu = V\boldsymbol{\beta}$, where $V = (v_1, ..., v_n)^T$ is a known model matrix.

Consider a sample $(y_1, v_1), \cdots, (y_n, v_n)$ of n independent observations, where each random response is defined by $y_i = \min\{\log(x_i), \log(c_i)\}$. We assume non-informative censoring such that the observed lifetimes and censoring times are independent. Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. The log-likelihood function for the vector of parameters $\tau = (\alpha, \beta, \sigma, \boldsymbol{\beta}^T)^T$ from model (14) has the form $l(\tau) = \sum_{i \in F} l_i(\tau) + \sum_{i \in C} l_i^{(c)}(\tau)$, where $l_i(\tau) = \log[f(y_i)]$, $l_i^{(c)}(\tau) = \log[S(y_i)]$, $f(y_i)$ is the density (11) and $S(y_i)$ is the survival function (12) of Y_i. Then, the total log-likelihood function for τ is given by

$$\ell(\tau) = r\log\left(\frac{2^{\alpha^{-1}}}{\sigma}\right) + \sum_{i \in F}(z_i - u_i) + \sum_{i \in F} \log[\{1 + (\exp[-u_i])^{\beta-1}[1 + \{1 - \exp[-u_i]\}^{\alpha+1}]\}]$$

$$-\alpha^{-1}(\alpha + 1)\sum_{i \in F}\log[1 + \{1 - \exp[-u_i]\}^\alpha] - \beta^{-1}(\beta + 1)\sum_{i \in F}\log[1 + (\exp[-u_i])^\beta] \qquad (15)$$

$$+\sum_{i \in C}\log\left[1 - \frac{2^{\alpha^{-1}}\{1 - \exp[-u_i]\}}{[1+\{1-\exp[-u_i]\}^\alpha]^{\alpha^{-1}}+[1+(\exp[-u_i])^\beta]^{\beta-1}}\right],$$

where $u_i = \exp(z_i)$, $z_i = (y_i - v_i^T\boldsymbol{\beta})/\sigma$ and r is the number of uncensored observations (failures) and c is the number of censored observations. The MLE $\hat{\tau}$ of the vector of unknown parameters can be evaluated by maximizing the log-likelihood (15).

The asymptotic covariance matrix $K(\tau)^{-1}$ of $\hat{\tau}$ can be approximated by the inverse of the $(p + 2) \times (p + 2)$ observed information matrix $-\ddot{L}(\tau)$, whose elements are evaluated numerically in most statistical packages. The approximate multivariate normal distribution $N_{p+2}(0, -\ddot{L}(\tau)^{-1})$ for $\hat{\tau}$ can be used in the classical way to construct approximate confidence intervals for the parameters in τ.

8.1 Residual analysis

Residual analysis has critical role in checking the adequacy of the fitted model. In order to analyze departures from the error assumption, two types of residuals are considered: martingale and modified deviance residuals.

8.1.1 Martingale residual

The martingale residuals is defined in counting process and takes values between $-\infty$ and $+1$ (see, Fleming and Harrington(1994) for details). The martingale residuals for LNEW model is,

$$r_{M_i} = \begin{cases} 1 + \log\left(1 - \dfrac{2^{\alpha^{-1}}\{1-\exp[-u_i]\}}{[1+\{1-\exp[-u_i]\}^\alpha]^{\alpha^{-1}}+[1+(\exp[-u_i])^\beta]^{\beta-1}}\right) & \text{if } i \in F, \\[2ex] \log\left(1 - \dfrac{2^{\alpha^{-1}}\{1-\exp[-u_i]\}}{[1+\{1-\exp[-u_i]\}^\alpha]^{\alpha^{-1}}+[1+(\exp[-u_i])^\beta]^{\beta-1}}\right) & \text{if } i \in C, \end{cases} \qquad (16)$$

$u_i = \exp(z_i)$ and $z_i = (y_i - v_i^T\boldsymbol{\beta})/\sigma$.

8.1.2 Modified deviance residual

The main drawback of martingale residual is that when the fitted model is correct, it is not symmetrically distributed about zero. To overcome this problem, modified deviance

residual was proposed by Therneau et al. (1990). The modified deviance residual is given by

$$
r_{D_i} = \begin{cases} sign(r_{M_i})\{ -2[r_{M_i} + \log(1 - r_{M_i})]\}^{1/2}, & \text{if } i \in F \\ sign(r_{M_i})\{ -2r_{M_i}\}^{1/2}, & \text{if } i \in C, \end{cases} \tag{17}
$$

where $\hat{r}_{M_i}$ is the martingale residual.

9. Data Analysis

In this section, we provide applications to three real data sets to prove empirically the potentiality of NEW and NEN models. We also compare the fits of these models with some generalizations of the Weibull and normal distributions on two real data sets. The third data set refers to regression modeling. To determine the optimum model, we also compute Cramer von Mises (W^*) and Anderson-Darling (A^*) goodness of-fit statistics for all models. The statistics W^* and A^* are described in detail in Chen and Balakrishnan (1995). In general, it can be chosen as the best model which has the smaller values of the W^* and A^* statistics. All computations of the MLEs are performed by the maxLik routine and all goodness-of-fits statistics are calculated by the goftest routine in the R programme. The details are given by followings.

9.1. Stress data

The first real data set introduces the stress-rupture life of kevlar 49/epoxy strands which are subjected to constant sustained pressure at the 90% stress level until all had failed such that we obtain complete data with exact failure times. This data set was studied by Andrews and Herzberg (1985), Cooray and Ananda (2008) and Paraniaba et al. (2013).

The data are: 0.01, 0.01, 0.02, 0.02, 0.02, 0.03, 0.03, 0.04, 0.05, 0.06, 0.07, 0.07,0.08, 0.09, 0.09, 0.1, 0.1, 0.11, 0.11, 0.12, 0.13, 0.18, 0.19, 0.2, 0.23, 0.24, 0.24, 0.29 , 0.34, 0.35, 0.36, 0.38, 0.4, 0.42, 0.43, 0.52, 0.54, 0.56, 0.6, 0.6, 0.63, 0.65, 0.67, 0.68, 0.72, 0.72, 0.72, 0.73, 0.79, 0.79, 0.8, 0.8, 0.83, 0.85, 0.9, 0.92, 0.95, 0.99, 1, 1.01, 1.02, 1.03, 1.05, 1.1, 1.1, 1.11, 1.15, 1.18, 1.2, 1.29, 1.31, 1.33, 1.34, 1.4, 1.43, 1.45, 1.5, 1.51, 1.52, 1.53, 1.54, 1.54, 1.55, 1.58, 1.60, 1.63, 1.64, 1.8, 1.8, 1.81, 2.02, 2.05, 2.14, 2.17, 2.33, 3.03, 3.03, 3.34, 4.2, 4.69, 7.89.

In the applications, the shape of the hrf can help in selecting a particular model. For this aim, the total time on test (TTT) plot (Aarset, 1987) can be used. The TTT plot is obtained by plotting $T\left(\frac{r}{n}\right)$ against r/n where $T\left(\frac{r}{n}\right) = \left[\sum_{i=1}^{n} y_{(i)} + (n - r)y_{(r)}\right]/\sum_{i=1}^{n} y_{(i)}$, $r = 1, \dots, n$ and y_i are the order statistics of the sample. It is convex shape for decreasing hrf and is concave shape for increasing hrf. The TTT plots for this set is given by Figure 5. From Figure 5, the data set deals with convex-concave-convex shaped.
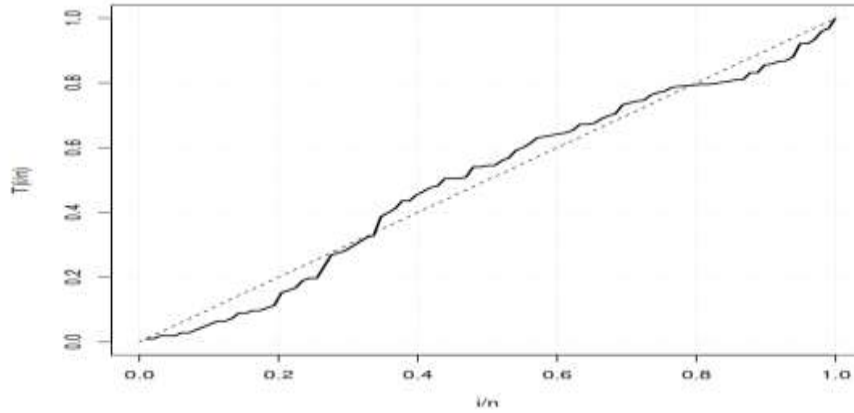
Figure 5: TTT plot for Stress data.

On this data set, we compare the NEW model with the Weibull (W), beta Weibull (BW), by Lee et al. (2007), Kumaraswamy Weibull (KwW), by Cordeiro and de Castro (2011), odd log-logistic Weibull (OLLW), by Cruz et al. (2016), and generalized odd log-logistic Weibull (GOLL2W), by Haghbin et al. (2017), under above criteria.

Table 2 lists the MLEs, their standard errors of the parameters and goodness-of-fits statistics from the fitted models. Table 2 shows that the NEW model could be chosen as the best model among the fitted models since these models have the lowest values of the W^* and A^* statistics.

Table 2: **MLEs, standard erros of the estimates (in parentheses) and goodness-of-fits statistics for the first data set**

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\gamma}$	A^*	W^*
NEW	0.2117	1.9899	0.7058	1.1436	0.6911	0.1167
	(0.2311)	(0.8222)	(0.1392)	(0.1548)		
GOLL2W	0.9201	1.0409	0.9745	0.9964	1.0247	0.1822
	(0.2069)	(3.7793)	(3.5523)	(0.1801)		
BW	0.7547	0.2398	3.4282	1.0439	0.8439	0.1387
	(0.1203)	(0.0273)	(0.0049)	(0.0136)		
KwW	0.2429	0.7264	3.4964	1.0366	0.8469	0.1355
	(0.0243)	(0.0443)	(0.0003)	(0.0001)		
OLLW	0.9200		1.0149	0.9965	1.0250	0.1824
	(0.2068)		(0.1276)	(0.1800)		
W			0.9938	0.9365	1.0882	1.1896
			(0.1125)	(0.0743)		

The plots of the fitted densities, cdfs and hrfs of all models, and probability-probability (P-P) plot of NEW model are displayed in Figure 6. These plots show that the NEW model provides the good fit to these data compared to the other models. The fitted hrf

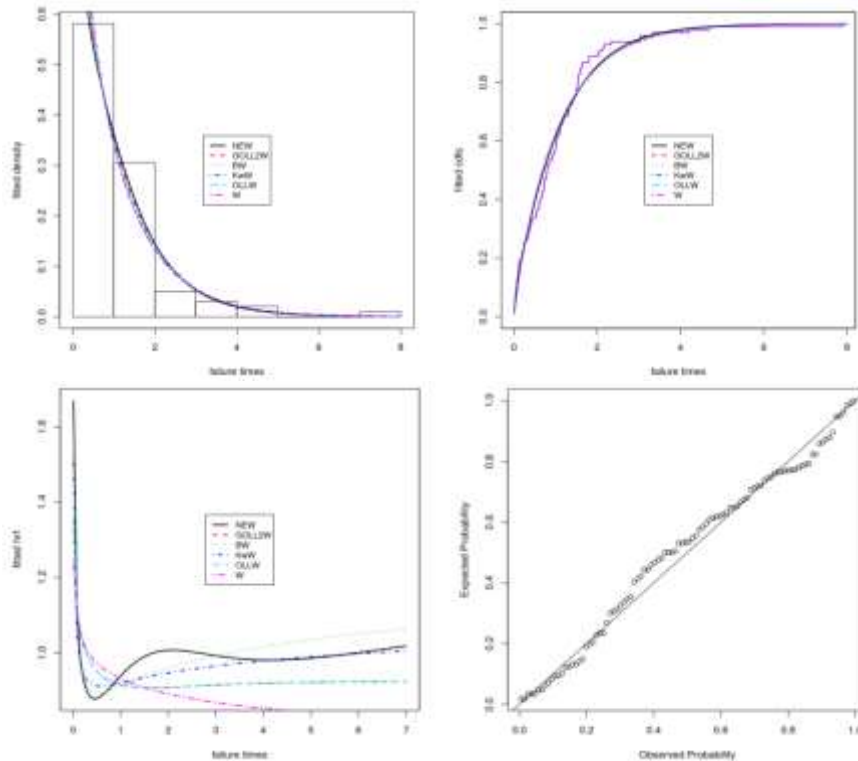shape of NEW model is close to TTT plot of data set. None of the models, except NEW, fit to Figure 5(a).

Figure 6: The fitted plots for the first data set.

9.2 Windshield data

Secondly, we consider the data on service times (in 1000 hours) for a particular model windshield given in Murthy et al. (2004). Recently, this data set has been analized by Ramos et al. (2013) and Tahir et al. (2015). The data are 0.046, 1.436, 2.592, 0.140, 1.492, 2.600, 0.150, 1.580, 2.670, 0.248, 1.719, 2.717, 0.280, 1.794, 2.819, 0.313, 1.915, 2.820, 0.389, 1.920, 2.878, 0.487, 1.963, 2.950, 0.622, 1.978, 3.003, 0.900, 2.053, 3.102, 0.952, 2.065, 3.304, 0.996, 2.117, 3.483, 1.003, 2.137, 3.500, 1.010, 2.141, 3.622, 1.085, 2.163, 3.665, 1.092, 2.183, 3.695, 1.152, 2.240, 4.015, 1.183, 2.341, 4.628, 1.244, 2.435, 4.806, 1.249, 2.464, 4.881, 1.262, 2.543, 5.140.

On this data set, we compare the NEW model with the norma (N), beta normal (BN), by Eugene et al. (2002), Kumaraswamy normal (KwN), by Cordeiro and de Castro (2011), odd log-logistic normal (OLLN), by da Silva Braga et al. (2016), and generalized odd log-logistic normal (GOLL2N), by Haghbin et al. (2017), under above criteria. Table 2 lists the MLEs, their standard errors of the parameters and goodness-of-fits statistics from the fitted models. Table 2 shows that the NEW model could be chosen as the best model among the fitted models since these models have the lowest values of the W^* and A^* statistics. Table 3 lists the MLEs, their standard errors of the parametersand goodness-of-fits statistics from the fitted models. Table 3 shows that the NEN model could be chosen

as the best model among the fitted models since these models have the lowest values of the W^* and A^* statistics.

Table 3: **MLEs, standard erros of the estimates (in parentheses) and goodness-of-fits statistics for the second data set**

Model	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	A^*	W^*
NEN	0.5682	28.6104	2.6402	1.1817	0.3075	0.0352
	(0.6472)	(8.3947)	(0.2200)	(0.1610)		
GOLL2N	1.3962	0.0890	-0.0234	0.6326	0.6313	0.0851
	(0.1508)	(0.0093)	(0.0063)	(0.00001)		
BN	4.6173	0.1146	-0.4056	0.6796	0.3511	0.0560
	(1.6998)	(0.0193)	(0.0019)	(0.0032)		
KwN	0.1161	0.4288	-0.1576	0.6490	0.3444	0.0437
	(0.0146)	(0.0008)	(0.0018)	(0.0018)		
OLLN	0.7854		2.1063	1.0145	0.4592	0.0581
	(0.4949)		(0.1651)	(0.5110)		
			2.0855	1.2354	0.4060	0.0470
			(0.1556)	(0.1101)		

The plots of the fitted densities and cdfs are given in Figure 7. P-P plots of all models are also drown in Figure 8. These plots show that the NEN model provides the good fit to these data compared to the other models.
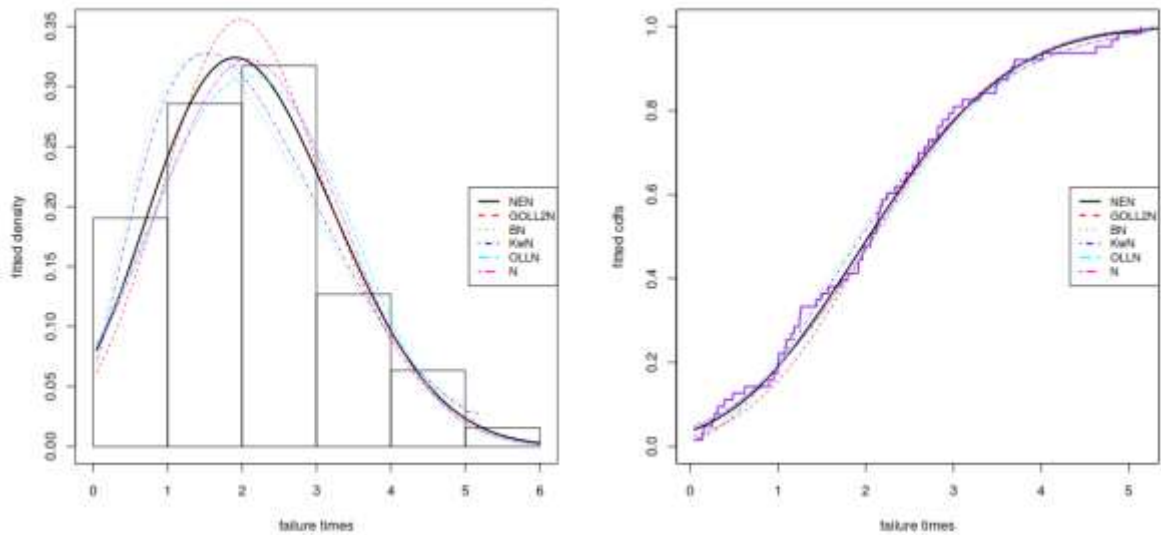

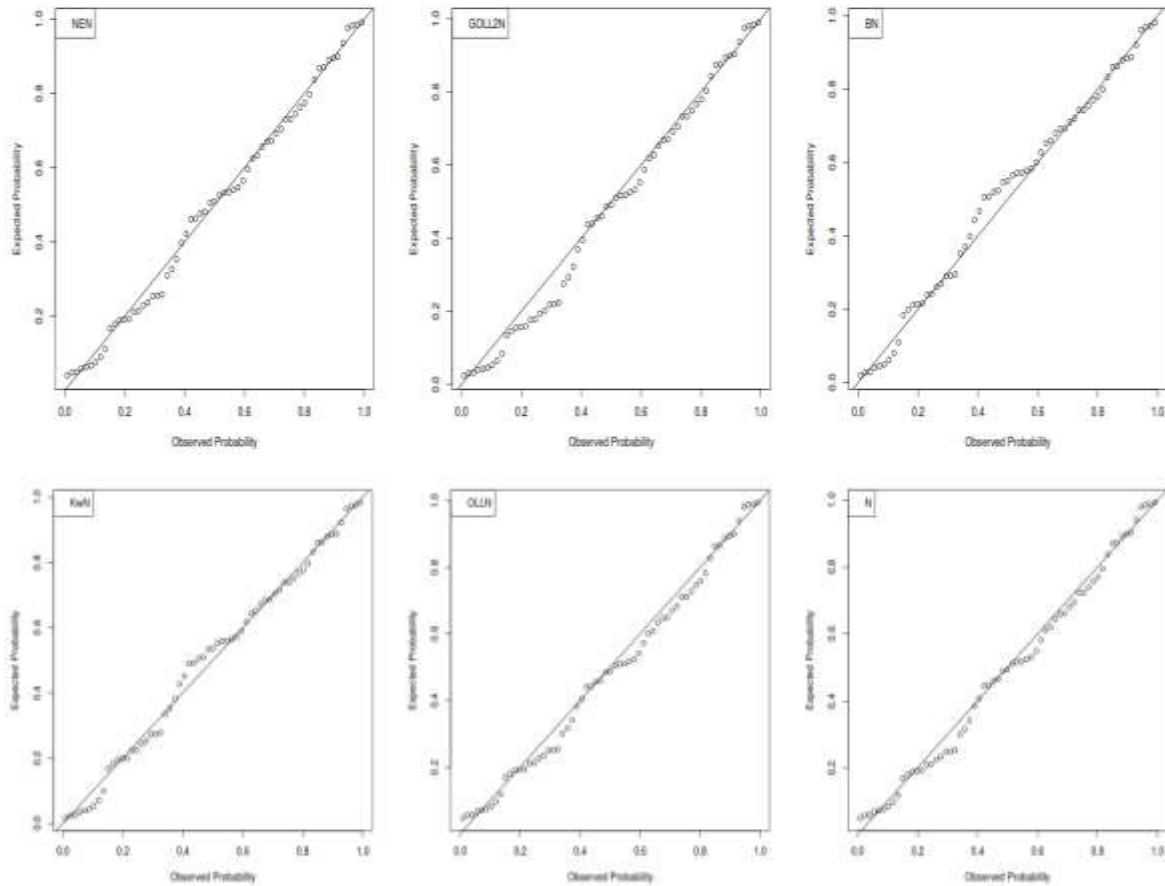Figure 7: The fitted plots for the second data set.

Figure 8: The P-P plots for the second data set.

9.3 HIV survival data

The hypothetical dataset, belongs to Health Maintenance Organization(HMO) is used to demonstrate the usefulness of LNEW regresion model. The data set can be found in R package *Bolstad2*. The variables involved in the study are: y_i - observed survival time (in months); $cens_i$ - censoring indicator (0= alive at study end or lost to follow-up,1=death due to AIDS or AIDS related factors), $x_{i1}(1 = yes, 0 = no)$ represents the history of drug use and x_{i2} represents the ages of patients.
We consider the following regression model
$$y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} x_{i1} + \boldsymbol{\beta_2} x_{i2} + \sigma z_i,$$
where y_i has the LOLLBXII density, for $i = 1, \dots, 100$. The MLE method is used to estimate unknown parameters of LNEW and Log-Weibull (LW) regression models. The survival function of LW regression model is given by
$$S(y) = \exp\left[-\exp\left(\frac{y - \mathbf{v}^T \beta}{\sigma}\right)\right]. \tag{18}$$

Table 4 lists the MLEs of the model parameters of the LNEW and LW regression models fitted to the current data and the estimated minus log-likelihood values. Based on the figures in Table 4, the LNEW regression model has the lower minus log-likelihood value than LW regression model. Therefore, it is concluded that LNEW regression model provides better fits than LW regression model for used data set. Based on the estimated

regression parameters, note that β_0, β_1 and β_2 is statistically significant at any significance level.

Table 4: **MLEs of the parameters, their standard errors and p-values, the estimated $-\ell$.**

Parameters	LNEW regression model			LW regression model		
	Estimates	Std.Errors	p-values	Estimates	Std.Errors	p-values
α	0.768	0.880	-	-	-	-
β	4.642	4.613	-	-	-	-
σ	0.737	0.140	-	0.839	0.072	<0.001
β_0	6.445	0.526	<0.001	6.148	0.510	<0.001
β_1	-0.090	0.013	<0.001	-0.090	0.013	<0.001
β_2	-1.008	0.199	<0.001	-1.049	0.188	<0.001
$-\ell$	127.8124			128.502		

9.3.1 Residual Analysis

Figure 9 displays the index plot of the modified deviance residuals and its Q-Q plot against to $N(0,1)$ quantiles for used data set. Based on Figure 9,it is concluded that none of the values appears as a possible outlier. Therefore, the fitted model is statistically valid.
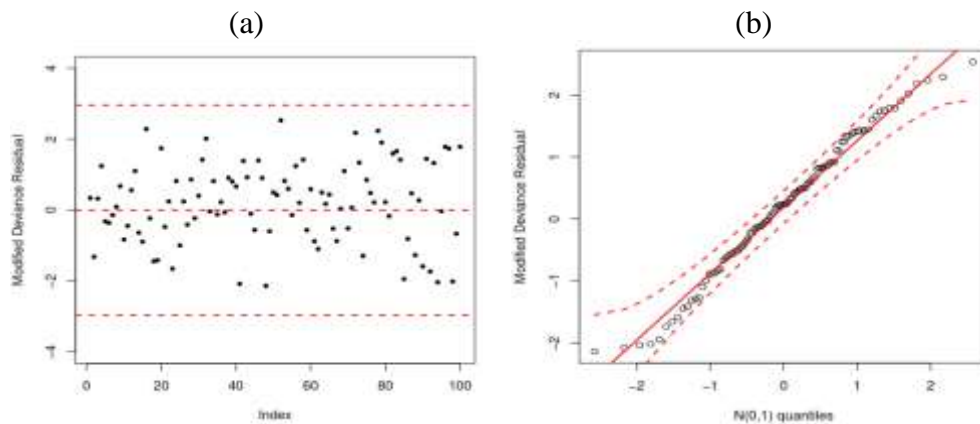

Figure 9: (a) Index plot of the modified deviance residual and (b) Q-Q plot for modified deviance residual.

10. Conclusions

In this work, we propose a new flexible exended G family of distribution. Some of its structural properties are derived and some useful characterization results are presented. The maximum likelihood method is discussed to estimate the model parameters by means of graphical and numerical Monte Carlo simulation study. Moreover, we introduce a new log-location regression model based on the proposed family. The martingale and modified deviance residuals are defined to detect outliers and evaluate the model assumptions. The potentiality of the new regression model is illustrated by means of a

real data. The flexibility of the new family illustrated by means of two real data sets. The NE-Weibull and NE-normal models outperforms several of the well-known lifetime Weibull and normal distributions with respect to the 1st and the 2nd real data applications respectively.

References

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

2. Cooray, K. and Ananda, M. M. (2008). A generalization of the half-normal distribution with applications to lifetime data. Communications in Statistics-Theory and Methods, 37(9), 1323-1337.

3. Chen, G. and Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. Journal of Quality Technology, 27(2), 154-161.

4. Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. J. Stat. Comput. Simul., 81, 883-893.

5. Cruz, J. N. D., Ortega, E. M. and Cordeiro, G. M. (2016). The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis. Journal of Statistical Computation and Simulation, 86(8), 1516-1538.

6. da Silva Braga, A., Cordeiro, G. M., Ortega, E. M. and Nilton da Cruz, J. (2016). The odd log-logistic normal distribution: Theory and applications in analysis of experiments. Journal of Statistical Theory and Practice, 10(2), 311-335.

7. Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. Communications in Statistics-Theory and methods, 31(4), 497-512.

8. Fleming, T. R. and Harrington, D.P. (1994). Counting process and survival analysis, John Wiley, New York.

9. Glänzel, W., A characterization theorem based on truncated moments and its application to some distribution families, Mathematical Statistics and Probability Theory (Bad Tatzmannsdorf, 1986), Vol. B, Reidel, Dordrecht, 1987, 75–84.

10. Glänzel, W., Some consequences of a characterization theorem based on truncated moments, Statistics: A Journal of Theoretical and Applied Statistics, 21 (4), 1990, 613–618.

11. Haghbin, H., Ozel, G., Alizadeh, M. and Hamedani, G. G. (2017). A new generalized odd log-logistic family of distributions. Communications in Statistics-Theory and Methods, 46(20), 9897-9920.

12. Lee, C., Famoye, F. Olumolade, O. (2007). Beta-Weibull distribution: some properties and applications to censored data. Journal of modern applied statistical methods, 6(1), 17.

13. Paranaiba, P. F., Ortega, E. M., Cordeiro, G. M. and Pascoa, M. A. D. (2013). The Kumaraswamy Burr XII distribution: theory and practice. Journal of Statistical Computation and Simulation, 83(11), 2117-2143.

14. Ramos, M. W. A., Marinho, P. R. D., da Silva, R. V. and Cordeiro, G. M. (2013). The exponentiated Lomax Poisson distribution with an application to lifetime data. Advances and Applications in Statistics, 34(2), 107-135.

15. Rezaei, S., Sadr, B. B., Alizadeh, M. and Nadarajah, S. (2017). Topp-Leone generated family of distributions: Properties and applications, Communications in Statistics-Theory and Methods, 46(6), 2893-2909.

16. Tahir, M. H., Cordeiro, G. M., Mansoor, M. and Zubair, M. (2015). The Weibull-Lomax distribution: properties and applications. Hacettepe Journal of Mathematics and Statistics, 44(2), 461-480.

17. Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. Biometrika, 77(1), 147-160.

Appendix A

Theorem 1: Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a given probability space and let $H = [a, b]$ be an interval for some $d < b$ $(a = -\infty, b = \infty$ mightaswellbeallowed). Let $X: \Omega \to H$ be a continuous random variable with the distribution function F and let q_1 and q_2 be two real functions defined on H such that

$$\mathbf{E}[q_2(X)|X \geq x] = \mathbf{E}[q_1(X)|X \geq x]\xi(x), \quad x \in H,$$

is defined with some real function η. Assume that $q_1, q_2 \in C^1(H)$, $\xi \in C^2(H)$ and F is twice continuously differentiable and strictly monotone function on the set H. Finally, assume that the equation $\xi q_1 = q_2$ has no real solution in the interior of H. Then F is uniquely determined by the functions q_1, q_2 and ξ, particularly

$$F(x) = \int_a^x C \left| \frac{\xi'(u)}{\xi(u)q_1(u) - q_2(u)} \right| \exp(-s(u))du,$$

where the function s is a solution of the differential equation $s' = \frac{\xi' q_1}{\xi q_1 - q_2}$ and C is the normalization constant, such that $\int_H dF = 1$. We like to mention that this kind of characterization based on the ratio of truncated moments is stable in the sense of weak convergence (see, Glänzel [2]), in particular, let us assume that there is a sequence $\{X_n\}$ of random variables with distribution functions $\{F_n\}$ such that the functions q_{1n}, q_{2n} and ξ_n $(n \in \mathbb{N})$ satisfy the conditions of Theorem 1 and let $q_{1n} \to q_1$, $q_{2n} \to q_2$ for some continuously differentiable real functions q_1 and q_2. Let, finally, X be a random variable with distribution F. Under the condition that $q_{1n}(X)$ and $q_{2n}(X)$ are uniformly integrable and the family $\{F_n\}$ is relatively compact, the sequence X_n converges to X in distribution if and only if ξ_n converges to ξ, where

$$\xi(x) = \frac{E[q_2(X)|X \geq x]}{E[q_1(X)|X \geq x]}.$$

This stability theorem makes sure that the convergence of distribution functions is reflected by corresponding convergence of the functions q_1, q_2 and ξ, respectively. It guarantees, for instance, the 'convergence' of characterization of the Wald distribution to that of the Lévy-Smirnov distribution if $\alpha \to \infty$. A further consequence of the stability property of Theorem 1 is the application of this theorem to special tasks in statistical practice such as the estimation of the parameters of discrete distributions. For such purpose, the functions q_1, q_2 and, specially, ξ should be as simple as possible. Since the function triplet is not uniquely determined it is often possible to choose ξ as a linear function. Therefore, it is worth analyzing some special cases which helps to find new characterizations reflecting the relationship between individual continuous univariate distributions and appropriate in other areas of statistics.