Journal of Data Science 677-706 , DOI: 10.6339/JDS.201810_16(4).00002

A NEW DISTRIBUTION FOR EXTREME VALUES: REGRESSION MODEL, CHARACTERIZATIONS AND APPLICATIONS

H. M. Yousof[1], S. M. A. Jahanshahi[2], T. G. Ramires[3,4], G. R. Aryal[*] and G. G. Hamedani[6]

[1]*Department of Statistics, Mathematics and Insurance, Benha University, Egypt*
[2]*Faculty of Mathematics, University of Sistan and Baluchestan, Zahedan, Iran*
[3]*Department of Exact Sciences, University of Sao Paulo, Piracicaba, Brazil*
[4]*Institute for Biostatistics and Bioinformatics, University of Haslet, Belgium*
[*]*Department of Mathematics, Statistics and CS, Purdue University Northwest, USA*
[6]*Department of Mathematics, Statistics and CS, Marquette University, USA*

Abstract

A new four parameter extreme value distribution is defined and studied. Various structural properties of the proposed distribution including ordinary and incomplete moments, generating functions, residual and reversed residual life functions, order statistics are investigated. Some useful characterizations based on two truncated moments as well as based on the reverse hazard function and on certain functions of the random variable are presented. The maximum likelihood method is used to estimate the model parameters. Further, we propose a new extended regression model based on the logarithm of the new distribution. The new distribution is applied to model three real data sets to prove empirically its flexibility.

Keywords: Extreme value distribution, Regression model, Order statistics, Parameter estimation, Simulation.

[*] Corresponding author: aryalg@pnw.edu

1 Introduction

The theory of extreme value distribution is very popular in statistics and is devoted to study of stochastical series of independent and identically distributed random variables. We study the behavior of extreme values, even though these values have a very low chance to occur, but can turn out to have a very high impact to the observed system. Finance and insurance are the best fields of research to observe the importance of extreme events. The study of extreme value theory started in the last century as an equivalent theory to the central limit theory, which is dedicated to the study of the asymptotic distribution of the average of a sequence of random variables. The central limit theorem states that the sum and the mean of the random variables from an arbitrary distribution are normally distributed under the condition that the sample size is sufficiently large. However, in some practical studies we are looking for the limiting distribution of maximum or minimum values rather than the average. Assume that $X_1, X_2, \dots, X_n$ is a sequence of iid (independent and identically distributed) random variables with common cumulative distribution function (cdf) $F(x)$. One of the most interesting statistics is the sample maximum

$$M_n = \max\{X_1, X_2, \dots, X_n\}.$$

One is interested in the behavior of M_n as the sample size n increases to infinity.Suppose there are sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P_r\left\{\frac{(M_n - b_n)}{a_n} \leq x\right\} \to G(x) as\ n \to \infty.$$

If $G(x)$ is a non-degenerate distribution function, then it will belong to one of the three following fundamental types of classic extreme value distribution:

1- Type I (Gumbel distribution) ;

2- 2-Type II (Fréchet distribution) ;

3- 3-Type III (Weibull distribution).

The extreme value theory was formally introduced by Fréchet (1927) to study the asymptotic distribution of the largest value. Once Fisher and Tippett (1928) proved that the limiting distribution of the extreme values can only be one of the three types (Gumbel, Fréchet and Weibull), the theory of extreme values gained much more popularity. In recent years, several extensions of the generalized extreme value distribution have been proposed in the literature. For details see Nadarajah and Eljabri (2014) and references therein. Among the three limiting distributions of the extrema, the Fréchet distribution has applications ranging from accelerated life testing through earthquakes, floods, horse racing, rainfall, queues in supermarkets, wind speeds and sea waves. For more details about the Fréchet distribution and its applications, see Kotz and Nadarajah (2000). Moreover, applications of this distribution in various fields are given in Harlow (2002).Recently, some extensions of the Fréchet distribution were considered e.g., the exponentiated Fréchet by Nadarajah and Kotz (2003), beta Fréchet by Nadarajah and Gupta (2004), Nadarajah and Kotz (2008) and Zaharim et al.(2009), beta Fréchet by Barreto-Souza et al. (2011) and Mubarak (2013), transmuted Fréchet by Mahmoud and Mandouh (2013), Marshall-Olkin Fréchet by Krishna et al. (2013), gamma extended Fréchet by da Silva et al.(2013), transmuted exponentiated Fréchet by Elbatal et al. (2014), transmuted Marshall-Olkin Fréchet by Afify et al. (2015), transmuted exponentiated generalized Fréchet by Yousof et al.(2015), beta exponential Fréchet by Mead et al.(2016), Kumaraswamy Marshall-Olkin Fréchet by Afify et al.(2016a), Weibull Fréchet by Afify et al.(2016b),Kumaraswamy transmuted Marshall-Olkin Fréchet by Yousof et al. (2016), among others.

The aim of this paper is to propose a generalization of the Fréchet distribution using the Topp-Leone generalizer which was introduced by Rezaei et al. (2017). This study follows up on Aryal et al. (2017), where a comprehensive description of the mathematical properties and some applications of the Topp-Leone Generated Weibull distribution were provided.The probability density function (pdf) and cdf of the Topp-Leone Generated family of distribution (for $\theta > 0$) are given by

$$f(x;\alpha,\theta,\xi) = 2\alpha\theta g(x;\xi)G(x;\xi)^{\theta\alpha-1}\left[1-G(x;\xi)^{\theta}\right]\left[2-G(x;\xi)^{\theta}\right]^{\alpha-1}, x\in\mathbb{R} \tag{1}$$

and

$$F(x;\alpha,\theta,\xi) = \{G(x;\xi)^{\theta}[2-G(x;\xi)^{\theta}]\}^{\alpha}, x\in\mathbb{R} \tag{2}$$

respectively. For $\theta = 1$, we obtain the Topp-Leone family. The properties of Topp and Leone’s distribution have been studied by many authors. We mention: moments by

Nadarajah and Kotz (2003) ; reliability measures and stochastic orderings by Ghitany et al. (2005) ; distributions of sums, products and ratios by Zhou et al. (2006) ; behavior of kurtosis by Kotz and Seier (2007) ; record values by Zghoul (2011) ; moments of order statistics by Genc (2012) ; stress-strength modeling by Genc (2013) and Bayesian estimation under trimmed samples by Sindhu et al. (2013), among others. The pdf and cdf of the Fréchet (Fr) distribution are given by

$$g(x;a,b) = ba^{b}x^{-(b+1)}exp[-(\frac{a}{x})^{b}], x > 0 \tag{3}$$

and

$$G(x,a,b) = exp[-(\frac{a}{x})^{b}], x > 0 \tag{4}$$

respectively, where $a > 0$ is a scale parameter and $b > 0$ is a shape parameter. By inserting (3) and (4) into (1), we can write the pdf of the Topp-Leone Generated Fréchet (TLGFr) distribution as

$$f_{TLGFr}(x) = \frac{2\alpha\theta b a^{b}}{x^{b+1}} \exp\left[-\theta\alpha\left(\frac{a}{x}\right)^{b}\right]\{1-\exp[-\theta\left(\frac{a}{x}\right)^{b}]\}\{2-\exp[-\theta\left(\frac{a}{x}\right)^{b}]\}^{\alpha-1} \tag{5}$$

The corresponding cdf is given by

$$F_{TLGFr}(x)\{(\exp[-\theta\left(\frac{a}{x}\right)^{b}])(2-\exp[-\theta\left(\frac{a}{x}\right)^{b}])\}^{\alpha}. \tag{6}$$

For $\theta = 1$, TLGFr reduces to Topp Leone Fréchet (TLFr)distribution. Figure 1 displays some plots of the TLGFr density and its cdf for selected values of the parameters α,θ,a and b. We see from the graphs that the TLGFr distribution is more flexible compare to the classical Fréchet distribution and TLFr distribution.

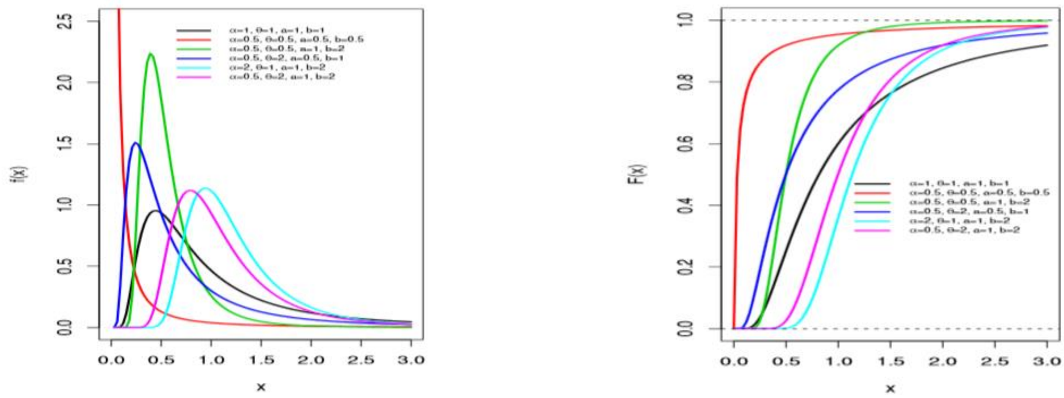

Figure 1: Probability density function(left) and Cumulative distribution function(right)of the TLGFr distribution.

The expression for hazard rate function (hrf, $h(x) = f(x)/[1-F(x)]$) and the reversed hazard rate function (rhrf, $r(x) = f(x)/F(x)$) of TLGFr distribution can be easily obtained using (5) and (6). Figure 2 displays some plots of the hrf and rhrf of the TLGFr distribution for selected values of the parameters α,θ,a and b..

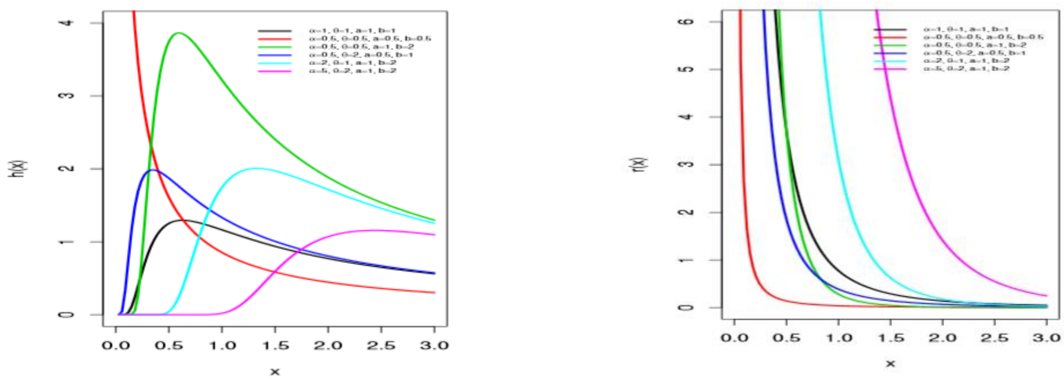

Figure 2: Hazard rate function (left) and the reversed hazard rate function (right) of the TLGFr distribution.

Most of the mathematical and statistical properties of the TLGFr distribution can not be expressed in simplified form so it is more convenient to express them in terms of Fréchet distribution. In order to do so we express the pdf and cdf of TLGFr distribution as a mixture of Fréchet distribution as below. The cdf in (6) can be expressed as

$$\begin{aligned} F(x) &= \left\{\left(\exp\left[-\theta\left(\frac{a}{x}\right)^{b}\right]\right)\left(2-\exp\left[-\theta\left(\frac{a}{x}\right)^{b}\right]\right)\right\}^{\alpha} \\ &= \exp\left\{-\theta\alpha\left(\frac{a}{x}\right)^{b}\right\}2^{\alpha}\left[1-\frac{1}{2}\exp\left\{-\theta\left(\frac{a}{x}\right)^{b}\right\}\right]^{\alpha} \\ &= 2^{\alpha}\exp\left\{-\theta\alpha\left(\frac{a}{x}\right)^{b}\right\}\sum_{k=0}^{\infty}(-1)^{k}\binom{\alpha}{k}\left(\frac{1}{2}\exp\left\{-\theta\left(\frac{a}{x}\right)^{b}\right\}\right)^{k} \qquad (7) \\ &= \sum_{k=0}^{\infty}\Upsilon_{k}\,\Pi_{(\alpha+k)\theta}(x), \end{aligned}$$

where

$$\Upsilon_k = (-1)^k 2^{\alpha-k}\binom{\alpha}{k}$$

and

$$\Pi_{(\alpha+k)\theta}(x) = G(x;a,b)^{(\alpha+k)\theta} = exp\{-[(\alpha+k)\theta](\frac{a}{x})^{b}\}$$

is the cdf of the Fréchet distribution with scale parameter $a[(\alpha+k)\theta]^{\frac{1}{b}}$ and shape parameter b. The corresponding TLGFr density function is obtained by differentiating (7) and is given by

$$f(x) = \sum_{k=0}^{\infty}\Upsilon_k\,\pi_{(\alpha+k)\theta}(x), \qquad (8)$$

where

$$\pi_{[(\alpha+k)\theta]}(x) = [(\alpha+k)\theta]\,\underbrace{ba^{b}x^{-(b+1)}\exp[-(\frac{a}{x})^{b}]}_{g(x;a,b)}\,\underbrace{\{\exp[-(\frac{a}{x})^{b}]\}^{[(\alpha+k)\theta]-1}}_{G(x;a,b)^{[(\alpha+k)\theta]-1}}$$

is the Fréchet density with scale parameter $a[(\alpha+k)\theta]^{\frac{1}{b}}$ and shape parameter b. Thus,the TLGFr density can be expressed as a double linear mixture of Fréchet densities. The rest of the paper is outlined as follows. In Section 2, we derive some of its mathematical properties including moments, generating function, residual life and reversed residual life functions and order statistics. Some characterizations results are provided in Section 3. Maximum likelihood estimation of the model parameters is addressed in Section 4. In Section 5, simulation results to assess the performance of the proposed estimation procedure are discussed. In Section 6, the log Topp-Leone Fréchet regression model is presented. In Section 7, we provide the applications to real data sets to illustrate the importance of the new family. Finally, we offer some concluding remarks in Section 8.

2 Mathematical properties

In this section we study the statistical properties of the TLGFr distribution, specifically quantile function, moments, incomplete moment, residual life and order statistics will be discussed.

2.1 Quantile function

The quantile function of a distribution is the real solution of $F\left(x_q\right) = q \; for \; 0 \leq q \leq 1$. The quantile function for a probability distribution has many uses in both the theory and application. It may be used to generate values of a random variable from an arbitrary distribution with the aid of a uniform random number generator. On inverting (6), the TLGFr quantile function is given by

$$x_q = a[-\frac{1}{\theta}\ln\{1-\sqrt{1-q^{\frac{1}{\alpha}}}\}]^{-\frac{1}{b}}.$$

2.2 Moments and cumulants

Let X be a TLGFr random variable. The rth order ordinary moment of X is given by

$$\mu_r' = \mathrm{E}(X^r) = \int_0^\infty x^r \, f(x) dx$$

$$= \int_0^\infty x^r \frac{2\alpha\theta b a^b}{x^{b+1}}$$

$$\exp\left\{-\theta\alpha\left(\frac{a}{x}\right)^b\right\}\left\{1-\exp\left[-\theta\left(\frac{a}{x}\right)^b\right]\right\}\left\{2-\exp\left[-\theta\left(\frac{a}{x}\right)^b\right]\right\}^{\alpha-1} dx \qquad (9)$$

$$= 2\alpha\theta b a^b \int_0^\infty x^{r-b-1} \exp\left\{-\theta\alpha\left(\frac{a}{x}\right)^b\right\}$$

$$\left\{1-\exp\left[-\theta\left(\frac{a}{x}\right)^b\right]\right\}\left\{2-\exp\left[-\theta\left(\frac{a}{x}\right)^b\right]\right\}^{\alpha-1} dx$$

One can calculate this integral numerically but it will be more convenient to use the mixture representation of the TLGFr and derive the rth order ordinary moment of X. In this case the rth moments are given by

$$\mu_r' = \sum_{k=0}^{\infty} \Upsilon_k \int_0^\infty x^r \, \pi_{(\alpha+k)\theta}(x) dx.$$

For any $r < b$, the moments of TLGFr distribution is given by

$$\mu_r' = a^r \sum_{k=0}^{\infty} \Upsilon_k \,[(\alpha+k)\theta]^{\frac{r}{b}} \Gamma(1-\frac{r}{b}), \tag{10}$$

where $\Gamma(a) = \int_0^{\infty} t^{a-1}\, exp(-t)dt$ is the gamma function. Table 1 lists the first four moments of X for two sets of parameters obtained from equations (9) and (10).In the expression (10) the number of terms (n) to be included in the sum varies from first 5 to first 100. It can be observed that the numerical values of the moments are equal if we take more terms in expression (10).

Table 1: First four moments of TLGFr (α,θ,a,b) with $b = 5$

		$\alpha = 2.5, \theta = 0.5,$	$a = 0.5$	$\alpha = 2.5, \theta = 2.5,$	$a = 2.5$
n	Moments	Expression(9)	Expression(10)	Expression(9)	Expression(10)
5	μ_1'	0.485263	0.487291	3.347659	3.361651
10	μ_1'	0.485263	0.485267	3.347659	3.347686
50	μ_1'	0.485263	0.485263	3.347659	3.347659
100	μ_1'	0.485263	0.485263	3.347659	3.347659
5	μ_2'	0.2418885	0.2435912	11.5118	11.59283
10	μ_2'	0.2418885	0.2418921	11.5118	11.51197
50	μ_2'	0.2418885	0.2418885	11.5118	11.5118
100	μ_2'	0.2418885	0.2418885	11.5118	11.5118
5	μ_3'	0.1244182	0.126083	40.84848	41.39507
10	μ_3'	0.1244182	0.1244221	40.84848	40.84977
50	μ_3'	0.1244182	0.1244182	40.84848	40.84848
100	μ_3'	0.1244182	0.1244182	40.84848	40.84848
5	μ_4'	0.06645564	0.06871811	150.5178	155.6421
10	μ_4'	0.06645564	0.06646157	150.5178	150.5312
50	μ_4'	0.06645564	0.06645564	150.5178	150.5178
100	μ_4'	0.06645564	0.06645564	150.5178	150.5178

Henceforth,$Y_{[(\alpha+k)\theta]}$denotes the Fréchet distribution with scale parameter $a[(\alpha+k)\theta]^{\frac{1}{b}}$ and shape parameter b.Setting $r = 1$ in (10), we have the mean of X. The central moments, cumulants, skewness and kurtosis measures can be calculated from the ordinary moments of order r for $r < b$.

2.3 Moment generating function

Here, we provide two formulae for the moment generating function (mgf) $M_X(t) = E(e^{tX})$of X. Clearly, the first one can be derived from equation (10), for $r < b$, as

$$M_X(t) = \sum_{k=0}^{\infty} \Upsilon_k \, M_{[(\alpha+k)\theta]}(t) = \sum_{k=0}^{\infty} \Upsilon_k \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r' = \sum_{k,r=0}^{\infty} \frac{\Upsilon_k}{a^{-r}} \frac{t^r}{r!} [(\alpha+k)\theta]^{\frac{r}{b}} \Gamma(1-\frac{r}{b}).$$

As for the second formula for $M_X(t)$, setting $y = x^{-1}$ in (3), we can write this mgf as

$$M(t;a,b) = ba^b \int_0^{\infty} \exp\,(\frac{t}{y}) y^{(b-1)} \exp\{-(ay)^b\} dy.$$

By expanding the first exponential and calculating the integral, we have

$$M(t;a,b) = ba^b \int_0^{\infty} \sum_{m=0}^{\infty} \frac{t^m}{m!} exp\left(\frac{t}{y}\right) y^{b-m-1} \exp\{-(ay)^b\}\, dy$$

$$= \sum_{m=0}^{\infty} \frac{a^m t^m}{m!} \Gamma\left(\frac{b-m}{b}\right),$$

where the gamma function is well-defined for any non-integer b.

2.4 Incomplete moments and mean deviations

The main applications of the first incomplete moment refer to the mean deviations and the Bonferroni and Lorenz curves. These curves are very useful in economics, reliability, demography, insurance and medicine. The sth incomplete moment, say $\varphi_s(t)$, of X can be expressed from (8), for $r < b$, as

$$\varphi_s(t) = \sum_{k=0}^{\infty} \Upsilon_k \int_{-\infty}^{t} x^s \, \pi_{(\alpha+k)\theta}(x) dx$$

$$= \sum_{k=0}^{\infty} \frac{\Upsilon_k}{a^{-s}} [(\alpha+k)\theta]^{\frac{s}{b}} \Gamma\left(1-\frac{s}{b}, [(\alpha+k)\theta] \left(\frac{a}{t}\right)^b\right),$$

where $\Gamma(.,.)$ represent the incomplete gamma function.

The mean deviations about the mean $[\delta_1 = E(|X - \mu_1'|)]$ and about the median $[\delta_2 = E(|X - M|)]$of$X$are given by $\delta_1 = 2\mu_1' F(\mu_1') - 2\varphi_1(\mu_1')and\delta_2 = \mu_1' - 2\varphi_1(M)$,respectively, where $\mu_1' = E(X)$, $M = Median(X) = Q(0.5)$ is the median, $F(\mu_1')$ is easily calculated from (6) and $\varphi_1(t)$ is the first incomplete moment given by the last equation with $s = 1$. A general equation for $\varphi_1(t)$ can be derived from the last equation as

$$\varphi_1(t) = \sum_{k=0}^{\infty} \Upsilon_k \, a[(\alpha + k)\theta]^{\frac{1}{b}} \Gamma\left(1 - \frac{1}{b}, [(\alpha + k)\theta]\left(\frac{a}{t}\right)^b\right).$$

2.5 Residual and reversed residual life functions

The nth moment of the residual life, say $m_n(t) = E[(X - t)^n | X > t], n = 1,2, \ldots$, uniquely determines $F(x)$. Let $R(t)$ be the reliability function of a random variable then the nth moment of the residual life of T is given by $m_n(t) = \frac{1}{R(t)} \int_t^{\infty} (x - t)^n \, dF(x)$. Therefore the nth moment of the residual life of TLGFr distribution is given by

$$m_n(t) = \frac{1}{R(t)} \sum_{k=0}^{\infty} \frac{\Upsilon_k^{\star}}{a^{-r}} [(\alpha + k)\theta]^{\frac{r}{b}} \Gamma\left(1 - \frac{r}{b}, [(\alpha + k)\theta]\left(\frac{a}{t}\right)^b\right),$$

where $\Upsilon_k^{\star} = \Upsilon_k \sum_{r=0}^{n} \binom{n}{r} (-t)^{n-r}$.Another interesting function is the mean residual life (MRL) function or the life expectation at age t defined by $m_1(t) = E[(X - t) | X > t]$,which represents the expected additional life length for a unit which is alive at age t.The MRL of X can be obtained by setting $n = 1$ in the last quation.The nth moment of the reversed residual life, say $M_n(t) = E[(t - X)^n | X \le t]$ for $t > 0$ and $n = 1,2, \ldots$ uniquely determines $F(x)$.We obtain $M_n(t) = \frac{1}{F(t)} \int_0^t (t - x)^n \, dF(x)$. Then, the nth moment of the reversed residual life of X becomes.

$$M_n(t) = \frac{1}{F(t)} \sum_{k=0}^{\infty} \frac{\Upsilon_k^{\star\star}}{a^{-r}} [(\alpha + k)\theta]^{\frac{r}{b}} \Gamma\left(1 - \frac{r}{b}, [(\alpha + k)\theta]\left(\frac{a}{t}\right)^b\right),$$

where $\Upsilon_k^{\star\star} = \Upsilon_k \sum_{r=0}^{n} (-1)^r \binom{n}{r} t^{n-r}$.The mean inactivity time (MIT) or mean waiting time (MWT) also called the mean reversed residual life function is given by $M_1(t) = E[(t - X) | X \le t]$, and it represents the waiting time elapsed since the failure of an item on condition that this failure had occurred in $(0, t)$. The MIT of the TLGFr distribution can be obtained easily by setting $n = 1$ in the above equation.

2.6 Order statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample from the TLGFr distribution and let $X_{(1)}, \ldots, X_{(n)}$ be the corresponding order statistics. The pdf of ith order statistic, say $X_{i:n}$, can be written as

$$f_{i:n}(x) = \frac{f(x)}{B(i, n - i + 1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} F^{j+i-1}(x), \qquad (11)$$

where B(.,.) is the beta function defined by $B(a,b)=\int_0^1 t^{a-1}\,(1-t)^{b-1}dt$.Substituting (5) and (6) in equation (11) the pdf of $X_{i:n}$ can be expressed as

$$f_{i:n}(x)=\sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} b_{j,r,k}\,\pi_{r+k}(x),$$

where

$$b_{j,r,k}=\frac{r(-1)^j b_r f_{j+i-1,k}}{B(i,n-i+1)(r+k)}$$

and $f_{j+i-1,k}$ can be obtained recursively from

$$f_{j+i-1,k}=\frac{1}{(kb_0)}\sum_{m=0}^{k}[m(j+i)-k]\,b_m f_{j+i-1,k-m}, \text{for } k\geq 1,$$

where $f_{j+i-1,0}=b_0^{j+i-1}$. Then, the density function of the TLGFr order statistics is a mixture of exp-W densities. Based on the last equation, we note that the properties of $X_{i:n}$ follow from those properties of Y_{r+k}. For example, the moments of $X_{i:n}$ can be expressed as (for $q<b$)

$$\mathrm{E}\left(X_{i:n}^q\right)=\sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} b_{j,r,k}\,E\left(Y_{r+k}^q\right)=\sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty}\frac{b_{j,r,k}}{a^{-q}}[r+k]^{\frac{1}{b}}\,\Gamma\left(1-\frac{q}{b}\right). \qquad (12)$$

3 Characterizations

In designing a stochastic model for a particular modeling problem, an investigator will be vitally interested to know if their model fits the requirements of a specific underlying probability distribution. To this end, the investigator will rely on the characterizations of the selected distribution. Thus, the problem of characterizing a distribution is an important problem in various fields and has recently attracted the attention of many researchers. Consequently, various characterization results have been reported in the literature. These characterizations have been established in different directions. This section deals with various characterizations of TLGFr distribution. These characterizations are based on: (i) a simple relationship between two truncated moments; (ii) the reverse (or reversed) hazard function; (iii) certain functions of the random variable. It should be mentioned that for characterization (i) the cdf may not have a closed form. We present our characterizations $(i)-(iii)$ in three subsections.

3.1 Characterizations based on two truncated moments

In this subsection we present characterizations of TLGFr distribution in terms of a simple relationship between two truncated moments. This characterization result employs a theorem due to Glänzel (1987),see Theorem 3.1 below. Note that the result holds also when the interval H is not closed. Moreover, as mentioned above, it could be also applied when the cdf F does not have a closed form. As shown in Glänzel (1990), this characterization is stable in the sense of weak convergence.

Theorem 3.1.

Let $(\Omega, \mathcal{F}, \boldsymbol{P})$ be a given probability space and let $H = [d, e]$ be an interval for some $d < e(d = -\infty, e = \infty$ might as well be allowed).Let $X: \Omega \to H$ be a continuous random variable with the distribution function F and let g and h be two real functions defined on H such that

$$E[g(X)|X \geq x] = E[h(X)|X \geq x]\xi(x), x \in H,$$

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

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

where the function s is a solution of the differential equation $s' = \frac{\xi' h}{\xi h - g}$ 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 1990), 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 $\xi_n (n \in \mathbb{N})$ satisfy the conditions of Theorem 3.1 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 , 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 $\alpha \to \infty$, as was pointed out in Glänzel and Hamedani(2001).

A further consequence of the stability property of Theorem 3.1 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.

In some cases, one can take $q_1(x) = 1$, which reduces the condition of Theorem 1 to $\mathbf{E}[q_2(X)|X \geq x] = \xi(x)$,$x \in H$. We, however, believe that employing three functions q_1, q_2 and ξ will enhance the domain of applicability of Theorem 3.1.

Proposition 3.1

Let $X:\Omega \to (0,\infty)$ be a continuous random variable and let, $h(x) = \{1 - exp[-\theta(\frac{a}{x})^b]\}^{-1}\{2 - exp[-\theta(\frac{a}{x})^b]\}^{1-\alpha}$ and $g(x) = h(x)exp[-\theta\alpha(\frac{a}{x})^b]$ for $x > 0$ .The random variable X has pdf (5) if and only if the function ξ defined in Theorem 3.1 has the form

$$\xi(x) = \frac{1}{2}\left\{1 + \exp\left[-\theta\alpha\left(\frac{a}{x}\right)^b\right]\right\}, x > 0. \quad (13)$$

Proof. Let X be a random variable with pdf (5), then

$$\left(1 - F(x)\right)E[h(x)|X \geq x] = 2\left\{1 - exp\left[-\theta\alpha\left(\frac{a}{x}\right)^b\right]\right\}, x > 0,$$

and

$$\left(1 - F(x)\right)E[g(x)|X \geq x] = \left\{1 - exp\left[-2\theta\alpha\left(\frac{a}{x}\right)^b\right]\right\}, x > 0,$$

and finally

$$\xi(x)h(x) - g(x) = \frac{1}{2}h(x)\left\{1 - exp\left[-\theta\alpha\left(\frac{a}{x}\right)^b\right]\right\} > 0 \; for \; x > 0.$$

Conversely,if ξ is given as above,then

$$s'(x) = \frac{\xi'(x)h(x)}{\xi(x)h(x) - g(x)} = \frac{\theta\alpha b a^b x^{-(b+1)}\exp[-\theta\alpha\left(\frac{a}{x}\right)^b]}{1 - \exp[-\theta\alpha\left(\frac{a}{x}\right)^b]} \; x > 0,$$

and hence

$$s(x) = -\log\left\{1 - exp\left[-\theta\alpha\left(\frac{a}{x}\right)^b\right]\right\}, x > 0.$$

Now,in view of Theorem 3.1, X has density (5).

Corollary 3.1.

Let $X: \Omega \rightarrow (0, \infty)$ be a continuous random variable and let $h(x)$ be as in Proposition 3.1. The pdf of X is (5) if and only if there exist functions g and ξ defined in Theorem 3.1 satisfying the differential equation

$$\frac{\xi'(x)h(x)}{\xi(x)h(x) - g(x)} = \frac{\theta\alpha b a^b x^{-(b+1)} \exp[-\theta\alpha \left(\frac{a}{x}\right)^b]}{1 - \exp[-\theta\alpha \left(\frac{a}{x}\right)^b]}, x > 0.$$

The general solution of the differential equation in Corollary 3.1 is

$$\xi(x) = \{1 - \exp[-\theta\alpha \left(\frac{a}{x}\right)^b]\}^{-1}[-\int \theta\alpha b a^b x^{-(b+1)} \exp\left[-\theta\alpha \left(\frac{a}{x}\right)^b\right] \left(h(x)\right)^{-1} g(x) + D],$$

where D is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 3.1 with $D = \frac{1}{2}$. However, it should be also noted that there are other triplets (h, g, ξ) satisfying the conditions of Theorem 3.1.

3.2 Characterization in terms of the reverse hazard function

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

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

Proposition 3.2.

Let $X: \Omega \rightarrow (0, \infty)$ be a continuous random variable. The pdf of X is (5) if and only if its reverse hazard function $r_F(x)$ satisfies the differential equation

$$r_F'(x) + (b+1)x^{-1} r_F(x) = -\frac{\theta b a^b x^{-(b+1)} \exp[-\theta \left(\frac{a}{x}\right)^b]}{\{2 - \exp\left[-\theta \left(\frac{a}{x}\right)^b\right]\}^2}. \quad (14)$$

Proof. If X has pdf (5), then clearly (14) holds. Now, if (14) holds, then

$$\frac{d}{du}\{x^{(b+1)} r_F(x)\} = 2\alpha\theta b a^b \frac{d}{dx}\{\frac{1 - \exp[-\theta \left(\frac{a}{x}\right)^b]}{2 - \exp[-\theta \left(\frac{a}{x}\right)^b]}\},$$

or

$$r_F(x) = \frac{2\alpha\theta b a^b x^{-(b+1)} \{1 - \exp[-\theta \left(\frac{a}{x}\right)^b]\}}{2 - \exp[-\theta \left(\frac{a}{x}\right)^b]},$$

which is the reverse hazard function of the TLGFr distribution.

3.3 Characterization based on certain functions of the random variable

The following proposition has already appeared in Hamedani (2013), so we will just state it here which can be used to characterize TLGFr distribution.

Proposition 3.3.

Let $X: \Omega \to (d, e)$ be a continuous random variable and let $\psi(x)$ be a differentiable function on (d, e) with $lim_{x\to e} - \psi(x) = 1$. Then for $\delta \neq 1$,

$$E[\psi(X)|X \leq x] = \delta\psi(x)$$

implies

Remarks 3.1. It is easy to see that for certain functions, e.g.,

$$\psi(x) = \exp\left[-\theta\left(\frac{a}{x}\right)^b\right]\left\{2 - \exp\left[-\theta\left(\frac{a}{x}\right)^b\right]\right\}, \delta = \frac{\alpha}{\alpha+1}$$

and $(d, e) = (0, \infty)$,proposition 3.3 provides a characterization of TLGFr distribution. Clearly there are other suitable functions , we choose the above one for simplicity.

4 Parameter estimation

Several approaches for parameter estimation has been proposed in the literature but maximum likelihood method is the most commonly employed. The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used for constructing confidence intervals and regions and also in test statistics. The normal approximation for these estimators in large samples can be easily handled either analytically or numerically. So, we consider the estimation of the unknown parameters of this family from complete samples only by maximum likelihood. Let $x_1, \dots, x_n$ be a random sample from TLGFr distributions with parameters α,θ, a and b.Let $\eta = (\alpha, \theta, a, b)^{\mathrm{T}}$be the 4×1 parameter vector. For determining the MLE of η, we have the log-likelihood function

$$\ell = \ell(\boldsymbol{\eta})$$

$$= n\log 2 + n\log\alpha + n\log\theta + n\log b + n\, b\log a$$

$$-(b+1)\sum_{i=1}^{n}\log(x_i) - \theta\alpha\sum_{i=1}^{n}\left(\frac{a}{x_i}\right)^b + \sum_{i=1}^{n} log\left(1 - s_i^{\theta}\right)$$

$$+(\alpha - 1)\sum_{i=1}^{n} log\left(2 - s_i^{\theta}\right),$$

where $s_i = \exp[-(\frac{a}{x_i})^b]$. The components of the score vector, $\boldsymbol{U}(\boldsymbol{\eta}) = \frac{\partial\ell}{\partial\boldsymbol{\eta}} = (\frac{\partial\ell}{\partial\alpha}, \frac{\partial\ell}{\partial\theta}, \frac{\partial\ell}{\partial a}, \frac{\partial\ell}{\partial b})^T$ are given by

$$U_\alpha = \frac{n}{\alpha} - \theta \sum_{i=1}^{n} \left(\frac{a}{x_i}\right)^b + \sum_{i=1}^{n} \log\left(2 - s_i^\theta\right),$$

$$U_\theta = \frac{n}{\theta} - \alpha \sum_{i=1}^{n} \left(\frac{a}{x_i}\right)^b - \sum_{i=1}^{n} \frac{s_i^\theta logs_i}{\left(1 - s_i^\theta\right)} - (\alpha - 1) \sum_{i=1}^{n} \frac{s_i^\theta \log s_i}{\left(2 - s_i^\theta\right)},$$

$$U_a = \frac{nb}{a} - \theta\alpha a^{b-1} b \sum_{i=1}^{n} x_i^b - \sum_{i=1}^{n} \frac{\theta m_i s_i^{\theta-1}}{\left(1 - s_i^\theta\right)} - (\alpha - 1) \sum_{i=1}^{n} \frac{\theta m_i s_i^{\theta-1}}{(2 - s_i^\theta)}$$

and

$$U_b = \frac{n}{b} + n \log a - \sum_{i=1}^{n} \log(x_i) - \theta\alpha \sum_{i=1}^{n} \left(\frac{a}{x_i}\right)^b \log\left(\frac{a}{x_i}\right)$$

$$- \sum_{i=1}^{n} \frac{\theta z_i s_i^{\theta-1}}{\left(1 - s_i^\theta\right)} - (\alpha - 1) \sum_{i=1}^{n} \frac{\theta z_i s_i^{\theta-1}}{\left(2 - s_i^\theta\right)},$$

where $z_i = -\left(\frac{a}{x_i}\right)^b \log\left(\frac{a}{x_i}\right) \exp\left[-\left(\frac{a}{x_i}\right)^b\right]$ and $m_i = -ba^{b-1} x_i^{-b} exp\left[-\left(\frac{a}{x_i}\right)^b\right]$.

Setting the nonlinear system of equations $U_\alpha = 0$, $U_\theta = 0$, $U_a = 0$ and $U_b = 0$ and;solving them simultaneously yields the MLE $\hat{\boldsymbol{\eta}} = (\hat{\alpha}, \hat{\theta}, \hat{a}, \hat{b})^T$ .To solve these equations,it is usually more convenient to use nonlinear optimization methods such as the quasi-Newton algorithm to numerically maximize ℓ. For interval estimation of the parameters, we obtain the 4×4 observed information matrix $J(\boldsymbol{\eta}) = \left\{\frac{\partial^2 \ell}{\partial_r \partial_s}\right\}$ (for $r, s = \alpha, \theta, a, b$) , whose elements can be computed numerically.Under standard regularity conditions when $n \to \infty$,the distribution of $\hat{\boldsymbol{\eta}}$ can be approximated by a multivariate normal $N_4(0, J(\hat{\eta})^{-1})$ distribution to construct approximate confidence intervals for the parameters. Here, $J(\hat{\boldsymbol{\eta}})$ is the total observed information matrix evaluated at $\hat{\boldsymbol{\eta}}$. 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.

5 Simulation study

In this section, we present some simulations for different sample sizes to assess the accuracy of the MLEs. Simulating random variables from well defined probability distributions has been discussed in the computational statistics literature, e.g. the inverse transformation method, the rejection and acceptance sampling technique, etc. An ideal technique for simulating from the TLGFr distribution is the inversion method. We can simulate random variable X by

$$X = a[-\frac{1}{\theta}\ln\{1-\sqrt{1-U^{\frac{1}{\alpha}}}\}]^{-\frac{1}{b}},$$

where U is a uniform random number in $(0,1)$. For selected combinations of a,b,α and θ we generate samples of sizes $n = 50, 100, 200, 300, 500$ and $1,000$ from the TLGFr distribution. We repeat the simulations $N = 1000$ times and evaluate the mean estimates and the root mean squared errors (RMSEs). We use two combinations for the parameter values (I: $a = 2, b = 0.5, \alpha = 1.5$ and $\theta = 2$ and II: $a = 0.5, b = 1.5, \alpha = 2$ and $\theta = 0.5$).The empirical results obtained using a script AdequacyModel of the R-package are given in Table 2 and Table 3. It can be observed that as sample size increases the mean squared error decreases. Therefore, the maximum likelihood method works very well to estimate the model parameters of the TLGFr distribution.

Table 2: Empirical means and the RMSEs for $a = 2, b = 0.5, \alpha = 1.5$ and $\theta = 2$.

	Estimated Values (RMSE)			
Sample size(n)	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\theta}$
50	3.9845	0.5378	4.8042	3.1633
	(5.4382)	(0.1084)	(8.1385)	(3.0408)
100	3.5187	0.5272	3.5857	2.9926
	(4.3291)	(0.0887)	(5.7399)	(2.7028)
200	3.2358	0.5177	2.9464	2.6954
	(3.8035)	(0.0707)	(4.2470)	(2.0572)
300	3.0797	0.5146	2.5250	2.5275
	(3.4286)	(0.0642)	2.5250	2.5275
500	2.9771	0.5102	2.3531	2.3071
	(2.8980)	(0.0546)	(2.6241)	(1.3926)
1000	2.5287	0.5070	1.9724	2.2207
	(1.8546)	(0.0449)	(1.6607)	(1.0559)

Table 3: Empirical means and the RMSEs for $a = 0.5, b = 1.5, \alpha = 2\ and\ \theta = 0.5$.

Estimated Values (RMSE)				
Sample size(n)	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\theta}$
50	0.5895	1.6214	6.9156	0.6689
	(0.3412)	(0.3299)	(12.8743)	(0.5908)
100	0.5718	1.5865	5.5775	0.6359
	(0.3093)	(0.2651)	(9.4981)	(0.4894)
200	0.5654	1.5553	4.1179	0.5725
	(0.2673)	(0.2171)	(5.9609)	(0.3672)
300	0.5527	1.5453	3.9076	0.5623
	(0.2639)	(0.1968)	(5.3145)	(0.3167)
500	0.5582	1.5426	2.9794	0.5525
	(0.2220)	(0.1712)	(3.2903)	(0.2853)
1000	0.5434	1.5302	2.6175	0.5279
	(0.1798)	(0.1334)	(2.4793)	(0.2044)

6 Regression model

When the explanatory variables affect the response variable X, they need to be in-clude in the models. To include the information of the explanatory variables, parametric models are widely used. A regression model that provides a good fit to lifetime data tends to yield more precise estimates of the quantities of interest. The non-inclusion of explanatory variables in the model, when really necessary, will result in considerable part of the variability in the response variable as residual. Recently, several regression models have been proposed in the literature, considering the class of location, for example, Cordeiro et al. (2016) introduced a new class of survival regression models considering the log-gamma extended Weibull model, Ramires et al. (2013) presented the beta generalized half-normal geometric regression model to predict the myelogenous leukemia,Cordeiro et al. (2015) introduced the log-generalized Weibull-log-logistic regression model for entomological data, among others.

Let X be a random variable having the pdf (5) with $\theta = 1$. A class of regression models for location is characterized by the fact that the random variable $Y = log(X)$ has a distribution with location parameter $\mu(v)$ dependent only on the explanatory variable vector and a scale parameter $\sigma > 0$. The log-linear model can be hen written as

$$Y = \mu(v) + \sigma Z,$$

where Z has the distribution which does not depend on v. The density function of the random variable Y, re-parameterized in terms of $\mu = log(a)$ and $\sigma = 1/b$, is given by

$$\mathrm{f(y)} = \frac{2\alpha}{\sigma}\exp\left[-\alpha e^{-\frac{(y-\mu)}{\sigma}} - \frac{(y-\mu)}{\sigma}\right]$$
$$[1-\exp(-e^{-\frac{(y-\mu)}{\sigma}})]\{2-\exp[-e^{-\frac{(y-\mu)}{\sigma}}]\}^{\alpha-1}, \qquad (15)$$

where $\alpha > 0$ is skewness parameter, $\mu \in \mathbb{R}$ is the location parameter and $\sigma > 0$ is the scale parameter.

We refer to equation (15) as the *log-Topp Leone Fréchet* (LTLFr) distribution,say $Y{\sim}LTLFr(\alpha, \mu, \sigma)$ .If $X{\sim}TLFr(\alpha, a, b), then\, Y = log(X) \sim LTLFr(\alpha, \mu, \sigma)$. The survival function corresponding to (15) is given by

$$S(y) = 1 - (exp[-e^{-\frac{(y-\mu)}{\sigma}}])^{\alpha}[2 - (exp[-e^{-\frac{(y-\mu)}{\sigma}}])]^{\alpha}. \qquad (16)$$

Consider the standardized random variable given by $Z = (Y-\mu)/\sigma.$The density function of Z is given by

$$f(z) = 2\alpha\, exp[-\alpha e^{-z} - z]\, [1 - exp(-e^{-z})]\{2 - ex\,p[-e^{-z}]\}^{\alpha-1}. \qquad (17)$$

Now, we propose a linear location regression model linking the response variable y_i and the explanatory variable vector $v_i^T = (v_{i1}, \dots, v_{ip})$ as follows

$$y_i = v_i^T \tau + \sigma z_i, i = 1, \dots, n, \qquad (18)$$

where the random error z_i has density function (5),$\tau = (\tau_1, \dots, \tau_p)^T, \sigma > 0$ and $\alpha > 0$ are unknown parameters. The parameter $v_i = v_i^T \tau$ is the location of y_i. The location parameter vector $v = (v_1, \dots, v_n)^T$ is represented by a linear model $v = \boldsymbol{V_\tau}$, where $\boldsymbol{V} = (v_1, \dots, v_n)^T$ is a known model matrix. With the LTLFr model (18), it is possible to incorporate explanatory variables in the analysis, opening new possibilities for fitting different types of data set.

Consider a sample $(y_1, v_1), \dots, (y_n, v_n)$ 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. Conventional likelihood estimation techniques can be applied here. The log-likelihood function for the vector of parameters $\boldsymbol{\eta} = (\alpha, \sigma, \tau^T)^T$ from model (18) has the form $l(\boldsymbol{\eta}) = \sum_{i \in F} l_i\,(\boldsymbol{\eta}) + \sum_{i \in C} l_i^{(c)}\,(\boldsymbol{\eta})$, where $l_i(\boldsymbol{\eta}) = log[f(y_i)]\, , l_i^{(c)}(\boldsymbol{\eta}) = log[S(y_i)]\, , f(y_i)$ is the density (15) and $S(y_i)$ is the survival function (16).The total log-likehood function for $\boldsymbol{\eta}$ reduces to

$$l(\boldsymbol{\eta}) = -\sum_{i\in F}(\alpha e^{-z_i} + z_i) + \sum_{i\in F}\log[1-\exp(-e^{-z_i})] + (\alpha-1)\sum_{i\in F}\log\{2-\exp[-e^{-z_i}]\} r\log\left(\frac{2\alpha}{\sigma}\right) + \sum_{i\in C}\{1-(\exp[-e^{-z_i}])^{\alpha}[2-(\exp[-e^{-z_i}])]^{\alpha}\}, \quad (19)$$

where $z_i = (y_i - v_i^T\tau)/\sigma$ and r is the number of uncensored observations (failures).The MLE $\hat{\boldsymbol{\eta}}$ of the vector of unknown parameters can be calculated by maximizing the log-likelihood (19). We use the NLMixed procedure in SAS to calculate the estimate $\hat{\boldsymbol{\eta}}$.

The elements of the $(p+2)\times(p+2)$ observed information matrix $-\ddot{\boldsymbol{L}}(\boldsymbol{\eta})$, say $-\boldsymbol{L}_{\alpha\alpha}, -\boldsymbol{L}_{\alpha\sigma}, -\boldsymbol{L}_{\alpha\tau j}, -\boldsymbol{L}_{\sigma\sigma}, -\boldsymbol{L}_{\sigma\tau j}$ and $-\boldsymbol{L}_{\beta_j\beta_s}(for\ j, s = 1, \dots, p)$ have to be evaluated numerically. Inference on $\boldsymbol{\eta}$ can be conducted based on the approximate multivariate normal $N_{p+2}(0, -\ddot{\boldsymbol{L}}(\hat{\boldsymbol{\eta}})^{-1})$ distribution for $\hat{\boldsymbol{\eta}}$.

7 Applications

In many statistical applications, the interest is centered on estimating some population parameters. For example, the average rainfall, the median temperature, the average income and others are based on samples taken from the same population. Sometimes, the most important values are not the average or median, but the maximum or minimum values (see Castillo (1994)). For example, the maximum flood height, maximum earthquake intensity, largest wildfire, the amounts of large insurance losses etc.. Largest values, such as loads, earthquakes, winds, floods, waves, and the smallest values, such as strength, stress, are the key to failure of engineering works, so construction engineering should be based on extreme values. Some publications related to extremes from fields such as ocean engineering, structural engineering, material strength and fatigue strength can be found in Court (1953), Goodknight and Russell (1963), Draper (1963), Earle et al. (1974), Cavanie et al. (1976), Chakrabarti and Cooley (1977), Battjes (1978), Leadbetter et al. (1983), Bryant (1983), Castillo and Sarabia (1992), Castillo and Sarabia (1994), Harlow (2002), Ferro and Segers (2003), Nadarajah and Kotz (2008), Borgman(1963, 1970, 1973) Zaharim et al. (2009), Mubarak (2013), Afify et al. (2015), Afify et al. (2016a), Afify et al. (2016b), Yousof et al. (2016), among others.

In this section, we provide two applications to real data to illustrate the importance of the TLGFr distribution presented in Section 1. We also present an application for the regression model proposed in Section 6.

For the first two examples, we compare the fits of the TLGFr distribution with other models such as Fréchet (Fr), Kumaraswamy Fréchet (KFr), exponentiated Fréchet (EFr), beta Fréchet (BFr), transmuted Fréchet (TFr), Marshal-Olkin Fréchet (MOFr) and McDonald Fréchet (McFr) distributions given by:

- KFr : $f(x;\alpha,\theta,a,b) = \alpha\theta b a^b x^{-(b+1)} \exp\left[-\alpha\left(\frac{a}{x}\right)^b\right]\left\{1-\exp\left[-\alpha\left(\frac{a}{x}\right)^b\right]\right\}^{\theta-1}$;
- EFr : $f(x;\alpha,\theta,a,b) = \alpha b a^b x^{-(b+1)} \exp\left[-\alpha\left(\frac{a}{x}\right)^b\right]\left\{1-\exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}^{\alpha-1}$;
- BFr : $f(x;\alpha,a,b) = \frac{b a^b}{B(\alpha,\theta)} x^{-(b+1)} \exp\left[-\alpha\left(\frac{a}{x}\right)^b\right]\left\{1-\exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}^{\theta-1}$;
- TFr : $f(x;\theta,a,b) = b a^b x^{-(b+1)} \exp\left[-\left(\frac{a}{x}\right)^b\right]\left\{1+\theta-2\theta\exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}$;
- MOFr : $f(x;\alpha,a,b) = \alpha b a^b x^{-(b+1)} \exp\left[-\left(\frac{a}{x}\right)^b\right]\left\{\alpha+(1-\alpha)\exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}^{-2}$;
- McFr : $f(x;\alpha,\theta,\gamma,a,b) = \frac{\gamma b a^b x^{-(b+1)}}{B(\alpha,\theta)} \exp\left[-\left(\frac{a}{x}\right)^b\right] (\exp[-\left(\frac{a}{x}\right)^b])^{\alpha\gamma-1}(1-$

  $(\exp[-\left(\frac{a}{x}\right)^b])^{\gamma})^{\theta-1}$.

The parameters of the above densities are all positive real numbers except for the TFr distributions for which $|\theta| \leq 1$. In the third example, we compare the LTLFr regression model with the LFr regression model.

To evaluate performance of considered models, the MLEs of the parameters for the considered models are calculated and five goodness-of-fit statistics are used to compare the new distribution. The measures of goodness of fit including the Akaike information criterion *(AIC)*, Bayesian information criterion *(BIC)*, Anderson-Darling (*A**), Cramér- von Mises (*W**) and Kolmogrov-Smirnov (K-S) statistics are computed to compare the fitted models. In general, the smaller the values of these statistics, the better the fit to the data. The required computations are carried out in the R language for the first two applications. In the third application (censored data) the computations were done using the subroutine nlmixed of the SAS software.

7.1 Breaking stress of carbon ftbers

This data set is an uncensored data set consisting of 100 observations on breaking stress of carbon fibers (in Gba) given by Nichols and Padgett (2006) and these data are used by Mahmoud and Mandouh (2013) to fit the transmuted Fréchet distribution.

The statistics of the fitted models are presented in Table 4 and the MLEs and corresponding standard errors are given in Table 5. We note from Table 4 that the TLGFr gives the lowest values the *AIC*, *BIC*, *A**, *W** and K-S statistics for the carbon fibres data as compared to the other generalizations of Fréchet distribution. Therefore, TLGFr distribution yields the best fit for the subject data.

Table 4: The statistics: *AIC*, *BIC*, *W**, *A**, K-S for the carbon fibers data.

Model	Goodness of fit criteria					
	AIC	*BIC*	W*	*A**	K-S	P-value(K-S)
TLGFr	113.78	124.21	0.0803	0.6111	0.0757	0.6159
Fr	114.38	124.59	0.1090	0.7657	0.0874	0.4282
KFr	113.94	124.36	0.0812	0.6217	0.0759	0.6118
EFr	113.88	124.30	0.1091	0.7658	0.0874	0.4287
BFr	113.93	124.35	0.0809	0.6207	0.0757	0.6147
TFr	114.40	124.31	0.0871	0.6209	0.0782	0.5734
MOFr	113.98	124.80	0.0886	0.6142	0.0763	0.5168
McFr	123.97	137.00	0.1333	1.0608	0.0807	0.5332

Table 5: MLEs and their standard errors (in parentheses) for the carbon fibers data.

Model	Estimates				
	$\hat{\alpha}$	$\hat{\theta}$	$\hat{a}$	$\hat{b}$	$\hat{\gamma}$
TLGFr	3.4224	0.8287	1.3264	2.7435	-
	(7.6918)	(13.1029)	(7.6681)	(0.5657)	-
Fr	-	-	1.3968	4.3724	-
			(0.0336)	(0.3278)	-
KFr	0.8489	1.6239	1.6341	3.4208	-
	(16.083)	(0.6979)	(9.049)	(0.7635)	-
EFr	1.0251	-	1.3889	4.3730	-
	(17.789)	-	(5.5116)	(0.3277)	-
BFr	0.7346	1.5830	1.6684	3.5112	-
	(1.5290)	(0.7132)	(0.7662)	(0.9683)	-
TFr	-	-0.7166	1.2656	4.7121	-
	-	(0.2616)	(0.0579)	(0.3657)	-
MOFr	0.0033	-	6.2296	1.2419	-
	(0.0009)	-	(1.0134)	(0.1181)	-
McFr	44.423	19.859	0.0203	46.974	0.8503
	(25.100)	(6.706)	(0.0060)	(21.871)	(0.1353)

The histogram of the carbon fibers data and estimated pdfs and cdfs of the TLGFr distribution and its competing models are displayed in Figure 3.

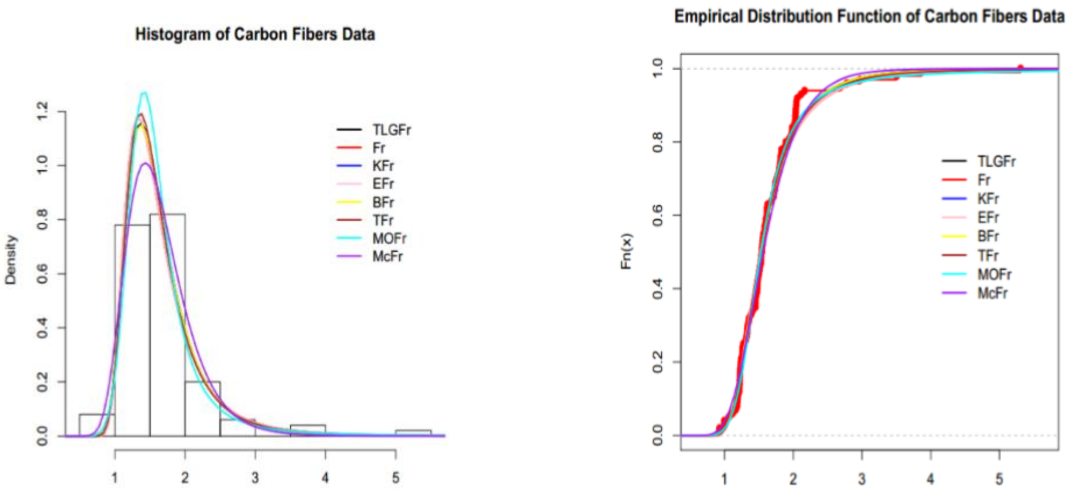

Figure 3: Histogram (left) and cdf (right) of the carbon fibers data.

7.2 Strength of glass ftbers

The second data set is generated data to simulate the strengths of glass fibers which was given by Smith and Naylor (1987). The statistics of the fitted models are presented in Table 6 and the MLEs and corresponding standard errors are given in Table 7. It is clear from Table 6 that the TLGFr gives the lowest values the *AIC*, *BIC*, *A**, *W** and K-S statistics for this data set as compared to their sub-models, and therefore these models can be chosen as the best ones.

Table 6: The statistics: *AIC*, *BIC*, *W**, *A**, and K-S for the glass fibers data.

Model	Goodness of fit criteria					
	AIC	*BIC*	W*	*A**	K-S	P-value(K-S)
TLGFr	47.765	56.338	0.0620	0.4878	0.0705	0.8905
Fr	48.127	58.414	0.0707	0.5332	0.0772	0.8185
KFr	47.866	56.438	0.0634	0.4981	0.0715	0.8810
EFr	48.127	56.557	0.0707	0.5332	0.0772	0.8187
BFr	47.880	56.452	0.0640	0.5008	0.0716	0.8804
TFr	48.738	57.168	0.0655	0.4939	0.0735	0.8470
MOFr	48.031	56.460	0.0629	0.4902	0.0813	0.7685
McFr	56.773	67.488	0.1161	0.9193	0.0831	0.7455

Table 7: MLEs and their standard errors (in parentheses) for the glass fibers data

Model	Estimates				
	$\hat{\alpha}$	$\hat{\theta}$	$\hat{a}$	$\hat{b}$	$\hat{\gamma}$
TLGFr	13.681	0.1553	1.7708	3.0833	-
	(58.034)	(3.1422)	(11.793)	(0.7578)	-
Fr	-	-	1.4108	5.4377	-
	-	-	(0.0344)	(0.5192)	-
KFr	0.2855	1.2824	1.9142	4.7731	-
	(9.1338)	(0.6388)	(12.836)	(1.3134)	-
EFr	0.9059	-	1.4367	5.4379	-
	(2.764)	-	(4.324)	(0.5193)	-
BFr	1.2996	1.2649	1.3945	4.7927	-
	(4.4378)	(0.6640)	(0.9304)	(1.4641)	-
TFr	-	0.7778	1.5491	4.3139	-
	-	(0.2477)	(0.0655)	(0.5849)	-
MOFr	0.0023	-	5.2383	1.4537	-
	(0.0004)	-	(0.8209)	(0.1650)	-
McFr	56.227	14.953	0.0073	29.104	1.1770
	(30.539)	(4.733)	(0.0013)	(11.304)	(0.1595)

The histogram of the second data and estimated pdfs and cdfs of the TLGFr distribution and its competing models are displayed in Figure 4. It is clear that the TLGFr distribution yields the best fit for this data.

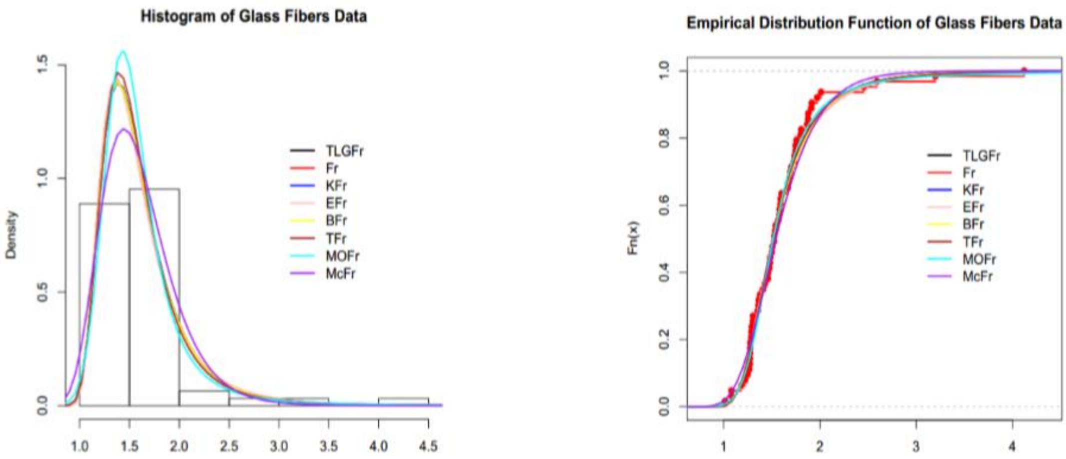

Figure 4: Histogram (left) and cdf (right) of the glass fibers data.

7.3 Application of log-TLFr regression model

In this subsection we present an application to the LTLFr regression model proposed in Section 6. The statistics used to compare the fitted models are AIC and BIC, due to the fact that the $A^{\star}$, $W^{\star}$ and K-S statistics are not appropriate for censored data.

Consider a data set reported by Batchelor and Hackett (1970) which are the results of a study of 16 acutely burned patients treated with skin allografts. In this study the patients received from one to four grafts. For each graft, the time in days x_i to rejection of the graft was recorded as well as an indicator variable v_1 which had a value of 1 if the graft was a good match and 0 if it was a poor match. The survival times of some grafts were censored by the death of the patient.

Here, we present results by fitting the model

$$y_i = \tau_0 + \tau_1 v_{i1} + \sigma z_i,$$

where the random variable $Y = log(X)$ follows the LTLFr distribution given by (15). The MLEs of the model parameters and their asymptotic SEs are listed at the end of Table 8. We also present in this table the fitted log-Fréchet (LFr) regression model. We can conclude with this table that the LTLFr regression model provides better fit than the LFr regression model.

Table 8: MLEs of the model parameters, corresponding SEs (given in parentheses), p-values in [·] and the AIC and BIC statistics.

Model	Estimates				AIC	BIC
LTLFr$(\alpha, \sigma, \tau_0, \tau_1)$	0.773	0.935	3.328	0.665	77.8	83.9
	(1.066)	(0.188)	(0.843)	(0.254)		
			[0.001]	[0.013]		
LFr(σ, τ_0, τ_1)	0.751	2.670	0.728		79.9	84.5
	(0.094)	(0.179)	(0.264)			
		[<0.0001]	[0.009]			

Figure 5 provides the plots of the empirical and estimated survival functions of the LTLFr distribution. These plots indicate that this regression model provides a satisfactory adjustment for these data.

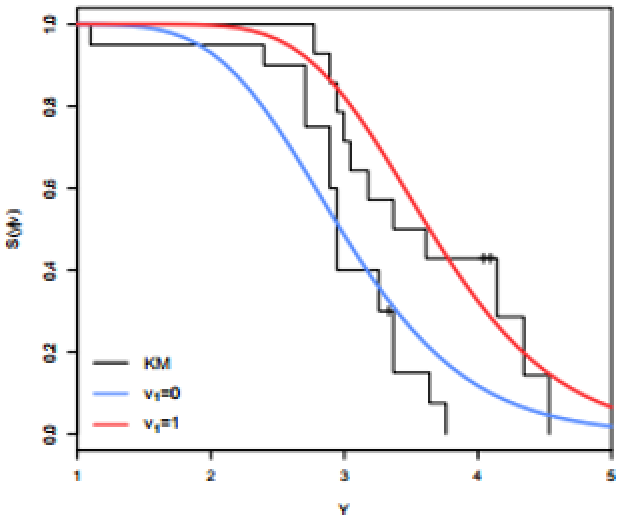

Figure 5: Estimated LTLFr survival function and empirical survival for each group of skin data.

8 Conclusions

In this paper, we introduce a new four-parameter extreme value model called the Topp-Leone generated Fréchet (TLGFr) distribution, which extends the Fréchet distribution. We study some of its statistical and mathematical properties. We derive explicit expressions for the ordinary and incomplete moments, mean deviations and generating function and moments of the residual and reversed residual life. Characterizations based on two truncated moments as well as based on reverse hazard function and on certain functions of the random variable are presented. We estimate the model parameters by maximum likelihood method. We assess the performance of the maximum likelihood estimators in terms of biases and mean squared errors by means of a simulation study. We introduce a new location-scale regression model based on the new distribution. The new distribution applied to three real data sets provide better fits than some other related models.

Acknowledgements

The authors are grateful to the editor and anonymous reviewers for their constructive comments and valuable suggestions which certainly improved the presentation and quality of the paper.

References

[1] Afify, A. Z., Hamedani, G. G., Ghosh, I. and Mead, M. E. (2015). The transmuted Marshall–Olkin Fréchet distribution: properties and applications. International Journal of Statistics and Probability, 4,132–184.

[2] Afify, A. Z., Yousof, H. M., Cordeiro, G.M. and Ahmad, M. (2016a). The Kumaraswamy Marshall-Olkin Fréchet distribution with applications. Journal of Islamic Countries Society of Statistical Sciences, 2, 1-18.

[3] Afify, A. Z., Yousof, H. M., Cordeiro, G. M., Ortega, E. M. M. and Nofal, Z. M. (2016b). The Weibull Fréchet distribution and its applications. Journal of Applied Statistics, 43, 2608–2626.

[4] Aryal, G., Ortega, E., Hamedani, G. and Yousof, H. (2017). The Topp-Leone Generated Weibull Distribution: Regression Model, Characterizations and Applications, International Journal of Statistics and Probability, 6, 126-141.

[5] Batchelor, J. R. and Hackett, M. (1970). HL-A matching in treatment of burned patients with skin allografts. The Lancet, 296, 581-583.

[6] Battjes, J. A. (1978). Probabilistic aspects of ocean waves. Technical Report no. 77- 2, Laboratory of fluid mechanics, Department of civil engineering, Delft University of Technology.

[7] Barreto-Souza, W. M., Cordeiro, G. M. and Simas, A. B. (2011). Some results for beta Fréchet distribution. Communication in Statistics-Theory and Methods, 40, 798-811.

[8] Borgman, L. E. (1963). Risk criteria. Journal of the Waterways and Harbors Division, ASCE, 89, 1–35.

[9] Borgman, L. E. (1970). Maximum wave height probabilities for a random number of random intensity storms. Technical report, University of California. Hydraulic Engineering Laboratory.

[10] Borgman, L. E. (1973). Probabilities for highest wave in hurricane. Journal of the Waterways, Harbors and Coastal Engineering Division, 99, 85–207.

[11] Bryant, P. J. (1983). Cyclic gravity waves in deep water. The Journal of the Australian Mathematical Society, Series B, Applied Mathematics, 25, 2–15.

[12] Castillo, E. and Sarabia, J. M. (1992). Engineering analysis of extreme value data: selection of models.Journal of waterway, port, coastal, and ocean engineering, 118, 129–146.

[13] Castillo, E. and Sarabia, J. M. (1994). Extreme value analysis of wave heights. Journal of Research of the National Institute of Standards and Technology, 99, 445–454.

[14] Cavanie, A., Arhan, M., Ezraty, R. (1976). A statistical relationship between individual heights and periods of storm wave. BOSS Conference, Norway, 354–360.

[15] Chakrabarti, S.K. and Cooley, R. P. (1977). Ocean wave statistics for 1961 north atlantic storm. Journal of the Waterway Port Coastal and Ocean Division, 103, 433–448.

[16] Cordeiro, G. M., Nadarajah, S., Ortega. E. M. M. and Ramires, T. G. (2018). An alternative two-parameter gamma generated family of distributions: properties and applications. Hacettepe Journal of Mathematics and Statistics, forthcoming.

[17] Cordeiro, G. M., Ortega, E. M. M. and Ramires, T. G. (2015). A new generalized Weibull family of distributions:mathematical properties and applications. Journal of Statistical Distributions and Applications, 2,1–25.

[18] Court, A. (1953). Wind extremes as design factors. Journal of the Franklin Insti- tute, 256, 39–56.

[19] Draper, L. (1963). Derivation of a design wave from instrumental records of sea waves. In ICE Proceedings, pages 291–304. Ice Virtual Library.

[20] da Silva, R., de Andrade, T. A., Maciel, D., Campos, R. P. S., and Cordeiro, G. M. (2013). A new lifetime model: The gamma extended Fréchet distribution. Journal of Statistical Theory and Applications, 12, 39-54.

[21] Earle, M. D., Ebbesmeyer, C. C. and Evans, D.J. (1974). Height-period joint probabilities in hurricane camille. Journal of the Waterways, Harbors and Coastal Engineering Division, 100,257–264.

[22] Elbatal, I. Asha, G. and Raja, V. (2014). Transmuted exponentiated Fréchet distribution: properties and applications. Journal of Statistics Application and Probability, 3, 379-394.

[23] Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In Mathematical Proceedings of the Cambridge Philosophical Society, pages 180–190. Cambridge University Press.

[24] Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65, 545–556.

[25] Fréchet, M. (1927). Sur la loi de probabilité de lécart maximum. Ann. de la Soc. Polonaise de Math.,6, 93–116.

[26] Genc, A. (2012). Moments of order statistics of ToppLeone distribution, Statistical Papers, 531, 117–131.

[27] Genc, A. (2013). Estimation of the Prob (X > Y) with Topp Leone distribution. Journal of Statistical Computation and Simulation, 83, 326-339.

[28] Ghitany, M. E., Kotz, S. and Xie, M. (2005). On some reliability measures and their stochastic orderings for the Topp Leone distribution. Journal of Applied Statistics, 32, 715-722.

[29] Glänzel, W. (1987). 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, 75–84.

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

[31] Glänzel, W. and Hamedani, G.G. (2001). Characterizations of univariate continuous distributions, Studia Scientiarum Mathematicarum Hungarica, 37, 83-118.

[32] Goodknight, R. C. and Russell, T. L. (1963) Investigation of the statistics of wave heights. ASCE Journal of the Waterways and Harbors Division, 89, 29–52.

[33] Hamedani, G. G. (2013). On certain generalized gamma convolution distributions II, Technical Report,No. 484, Marquette University.

[34] Harlow, D. G. (2002). Applications of the Fréchet distribution function, International Journal of Materials and Product Technology, 17, 482–495.

[35] Kotz, S. and Nadarajah, S. (2000). Extreme Value Distributions: Theory and Applications, Imperial College Press, London.

[36] Kotz, S. and Seier, E. (2007). Kurtosis of the Topp Leone distributions. Interstat, 1-15.

[37] Krishna, E., Jose, K. K., Alice, T. and Risti, M. M. (2013). The Marshall-Olkin Fréchet distribution. Communications in Statistics-Theory and Methods, 42, 4091- 4107.

[38] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). Extremes and related properties of random sequences and processes. Springer Verlag,Volume 11.

[39] Mahmoud, M. R. and Mandouh, R. M. (2013). On the transmuted Fréchet distribution. Journal of Applied Sciences Research, 9, 5553-5561.

[40] Mead, M. E., Afify, A. Z., Hamedani, G. G. and Ghosh, I. (2017). The beta exponential Fréchet distribution with applications. Austrian Journal of Statistics, 46, 41-63.

[41] Mubarak, M. (2012). Parameter estimation based on the Fréchet Progressive Type II censored data with binomial removals. International Journal of Quality, Statistics and Reliability, 2012, Article ID: 245910.

[42] Nadarajah, S. and Gupta, A. K. (2004). The beta Fréchet distribution. Far East Journal of Theoretical Statistics, 14, 15-24.

[43] Nadarajah, S. and Kotz, S. (2003). Moments of some J-shaped distributions. Journal of Applied Statistics, 30, 311-317.

[44] Nadarajah, S. and Kotz, S. (2003). The exponentiated Fréchet distribution. Available online at http://interstat.stat journals.net/YEAR/2003/abstracts/0312001.php

[45] Nicholas M. D. and Padgett W. J, (2006). A bootstrap control chart for Weibull percentiles, Quality and Reliability Engineering International. Vol. 22, 141-151.

[46] Nadarajah, S. and Kotz, S. (2008). Sociological models based on Fréchet random variables, Quality and Quantity, International Journal of Methodology, 42, 89–95.

[47] Nadarajah, S. and Eljabri, S. (2014). On Chen et al.'s Extreme Value Distribution, Journal of Data Science, 12, 87-106.

[48] Ramires, T. G., Ortega, E. M. M., Cordeiro, G. M. and Hamedani, G. G. (2013). The beta generalized half-normal geometric distribution. Studia Scientiarum Mathematicarum Hungarica, 50,523–554.

[49] 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.

[50] Sindhu, T. N., Saleem, M. and Aslam, M. (2013). Bayesian estimation for Topp Leone distribution under trimmed samples. Journal of Basic and Applied Scientific Research, 3, 347-360.

[51] Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and bayesian estimators for the three-parameter Weibull distribution. Applied Statistics, 36, 358–369.

[52] Yousof, H. M., Afify, A. Z., Alizadeh, M., Butt, N. S., Hamedani, G. G. and Ali, M.M. (2015). The transmuted exponentiated generalized-G family of distributions. Pakistan Journal of Statistics and Operation Research, 11, 441-464.

[53] Yousof, H. M., Afify, A. Z., Ebraheim, A. N., Hamedani, G. G. and Butt, N.S.(2016). On six-parameter Fréchet distribution: properties and applications, Pakistan Journal of Statistics and Operation Research 12, 281-299.

[54] Zaharim, A., Najid, S.K., Razali, A.M. and Sopian, K. (2009). Analysing Malaysian wind speed data using statistical distribution, Proceedings of the 4th IASME/WSEAS International Conference on Energy and Environment, Cambridge.

[55] Zghoul, A. A. (2011). Record values from a family of J-shaped distributions. Statistica, 71, 355-365.

[56] Zhou, M., Yang, D. W., Wang, Y. and Nadarajah, S. (2006). Some J-shaped distributions: Sums, products and ratios. Proceedings of the Annual Reliability and Maintainability Symposium, 175-181.

[57] Zografos, K., Balakrishnan,N.(2009).On families of beta and generalized gamma generated distributions and associated inference. Statistical Methodology, 6 (4), 344-362.