Journal of Data Science 743-770 , DOI: 10.6339/JDS.201810_16(4).00005

A NEW EXTENSION OF FRÉCHET DISTRIBUTION WITH REGRESSION MODELS, RESIDUAL ANALYSIS AND CHARACTERIZATIONS

Haitham M. Yousof[1], Emrah Altun[*], G.G. Hamedani[3],

[1]*Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt.*
[*]*Department of Statistics,Hacettepe University, Turkey.*
[3]*Department of Mathematics, Statistics and Computer Science, Marquette University, USA.*

ABSTRACT

A new log location-scale regression model with applications to voltage and Stanford heart transplant data sets is presented and studied. The martingale and modified deviance residuals to detect outliers and evaluate the model assumptions are defined. The new model can be very useful in analysing and modeling real data and provides more better fits than other regression models such as the log odd log-logistic generalized half-normal, the log beta generalized half-normal, the log generalized half-normal, the log-Topp-Leone odd log- logistic-Weibull and the log-Weibull models. Characterizations based on truncated moments as well as in terms of the reverse hazard function are presented. The maximum likelihood method is discussed to estimate the model parameters by means of a graphical Monte Carlo simulation study. The flexibility of the new model illustrated by means of four real data sets.

Keywords: Regression Modeling; Residual Analysis; Martingale Residual; Modified Deviance Residual; Fréchet Distribution; Monte Carlo Simulation.

* Corresponding author: Emrah Altun
Email:emrahaltun@hacettepe.edu.tr

1 Introduction

Let $Y_1, \dots, Y_n, \dots$ be a sequence of independent and identically distributed random variables (iid rv's) with common cumulative distribution function (cdf), $F(y)$. One of the most interesting statistics is the sample maximum $M_n = \max\{Y_1, \dots, Y_n\}$. One may be interested in the behavior of M_n as the sample size n increases to infinity.

$$\begin{aligned} P_r\{M_n \leq y\} &= P_r\{Y_1 \leq y, \dots, Y_n \leq y,\} \\ &= P_r\{Y_1 \leq y\} \dots P_r\{Y_n \leq y,\} \\ &= F(y)^m. \end{aligned}$$

Assume there are sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P_r\{[(M_n - b_n)/a_n] \leq y\} \to G(y) \text{ as } n \to \infty.$$

Now, if $G(y)$ is a non-degenerate distribution function, then it will belong to one of the three following fundamental types of classic extreme value family: Gumbel model which is Type **I**; Type **II** (Fréchet model); Type **III** (Weibull model).The extreme value theory concentrates on the behavior of the block maxima or minima. The extreme value theory was introduced first by Fréchet (1927) and Fisher and Tippett (1928), then followed by Von Mises (1936) and completed by Gnedenko (1943),Von Mises (1964) and Kotz and Johnson (1992), among others.

The so called Fréchet ('Fr' for short) distribution is one of the important distributions in extreme value theory and it has applications ranging from accelerated life testing through earthquakes, rainfall, floods, horse racing, queues in supermarkets, wind speeds and sea waves. For more details about the Fréchet distribution and its applications, see Kotz and Nadarajah (2000). Furthermore, applications of this model in various fields are given in Harlow (2002).

As of late, some new important extensions of the Fr distribution were considered.The exponentiated Fréchet (EFr) by Nadarajah and Kotz (2003), Beta Fréchet (Beta-Fr) by Nadarajah and Gupta (2004), Marshall-Olkin Fréchet (MOFr) by Krishna et.al. (2013),exponentiated exponential Fréchet (EEFr) by Mansoor et al. (2016)and Weibull Fréchet (W-Fr) by Afify et al. (2016).

The probability density function (pdf) and the cdf of the Fr distribution are given by (for $x \geq 0$).

$$g(x) = g(x; a, b) = b a^b x^{-(b+1)} exp[-\left(\frac{a}{x}\right)^b] \tag{1}$$

And

$$G(x) = G(x, a, b) = exp[-\left(\frac{a}{x}\right)^b], \tag{2}$$

respectively, where $a > 0$ is a scale parameter and $b > 0$ is a shape parameter. Gleaton and Lynch (2006) defined the cdf of the so called odd log-logistic-G (OLL-G) family by

$$F(x) = F(x;\theta,\xi) = \frac{G(x,\xi)^{\theta}}{G(x,\xi)^{\theta}+G(x,\xi)^{\theta}}. \quad (3)$$

The OLL-G density function becomes

$$f(x) = f(x;\theta,\xi) = \frac{\theta_g(x,\xi)[G(x,\xi)\bar{G}(x,\xi)]^{\theta-1}}{[G(x,\xi)^{\theta}+\bar{G}(x,\xi)^{\theta}]^2}, \quad (4)$$

where $\theta > 0$ is the shape parameter and $\xi = \xi_k = (\xi_1, \xi_2 \ldots)$ is a parameter vector. A rv X with pdf (4) is denoted by $X \sim OLL - G(\theta,\xi)$.The cdf and pdf of Odd log-logistic Fréchet (OLLFr) distribution are obtained by replacing $G(x,\xi)$ and $g(x,\xi)$ with cdf and pdf of Fréchet distribution in Equations (3) and (4). Then, the cdf and the pdf of OLLFr distribution are given, respectively, by

$$F(x;\theta,a,b) = \frac{exp[-\theta(\frac{a}{x})^b]}{exp[-\theta\left(\frac{a}{x}\right)^b] + \{1 - exp[-\left(\frac{a}{x}\right)^b]\}^{\theta}}, x \geq 0, \quad (5)$$

and

$$f(x;\theta,a,b) = \theta b a^b x^{-(b+1)}$$
$$\exp\left[-\left(\frac{a}{x}\right)^b\right]\left(\exp\left[-\left(\frac{a}{x}\right)^b\right]\left\{1-\exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}\right)^{\theta-1}$$
$$\times\left(\exp\left[-\theta\left(\frac{a}{x}\right)^b\right]+\left\{1-\exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}^{\theta}\right)^{-2}, x \geq 0, \quad (6)$$

The justification for the practicalness of this model is based on introducing a new flexible extension of Fr distribution with only three parameter. We are motivated to introduce the OLLFr distribution because it contains a number aforementioned known lifetime models as sub models like the OLL inverse exponential (OLLIE) model when $b = 1$ and the OLL inverse Rayleigh model (OLLIR) when $b = 2$. It exhibits decreasing as well as upside-down rates as illustrated in Section 2. It is shown that the OLLFr distribution can be expressed as a double linear mixture of Fr densities. It can be viewed as a suitable model for fitting the right-skewed and symmetric data sets. The OLLFr distribution outperforms several of the well-known distributions with respect to four real data applications as illustrated in Section 7. The new log-location regression model based on the OLLFr distribution provides better fits than log OLL generalized half-normal, log beta generalized half-normal, log generalized half-normal, log Topp-Leone odd log-logistic-Weibull and log-Weibull models for voltage and Stanford heart transplant data sets. Based on the residual analysis (martingale and modified deviance residuals) for the new log-location regression model, we conclude that none of the observed values appear as possible outliers as well as based on the index plot of the modified deviance residual and the Q-Q plot for modified deviance residual we note that OLLFr model is more appropriate for voltage and Stanford heart transplant data sets than all the existing regression models.

The rest of the paper is organized as follows. In Section 2, we present a graphical presentation for the OLLFr model. In Section 3, we introduce some of its mathematical properties. Section 4 deals with some useful Characterizations based on truncated moments as well as in terms of the reverse hazard function. In Section 5, the maximum likelihood method is used to estimate the model parameters by means of a Monte Carlo simulation study. A new Log-location regression model as well as residual analysis are presented and displayed in Section 6. Four applications to real data sets prove empirically the importance of the new model in Section 7. Finally, some concluding remarks are given in Section 8.

2 Graphical representation

In this section, we investigate the possible hrf and pdf shapes of OLLFr distribution. Figure 1 displays the pdf and hrf shapes of OLLFr distribution. Figure 1(a) reveals that the OLLFr distribution is good candidate to model both symmetric and left-skewed data sets. Based on the Figure 1(b), OLLFr has the following hrf shapes: increasing and upside-down.

In Figure 2, we investigate the shapes of hrf of OLLFr distribution based on parameter values $0 < \theta < 2$ and $0 < a < 2$ for fixed $b = 0.5$. Figure 2 also shows the regions for shape of hrf where it is increasing or upside-down.

(a) (b)

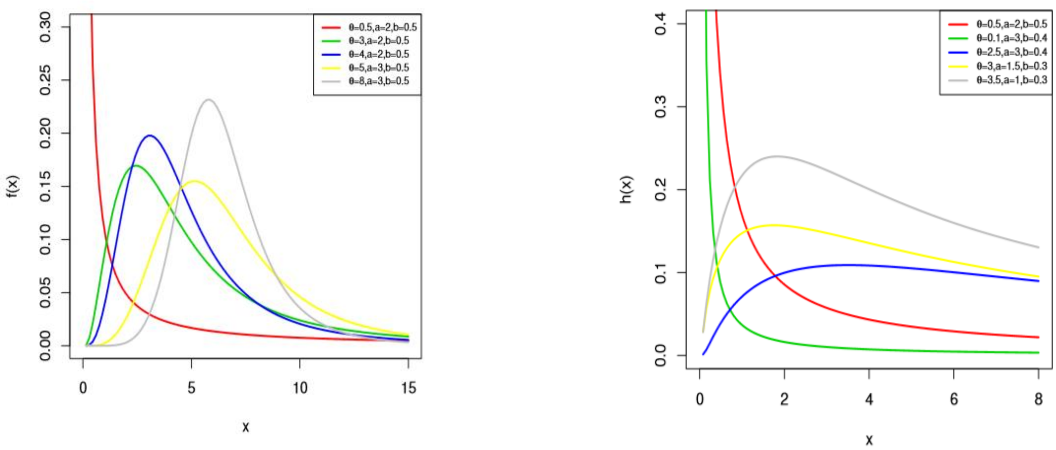

Figure 1: The pdf(a) and hrf(b) plots of OLLFr distribution

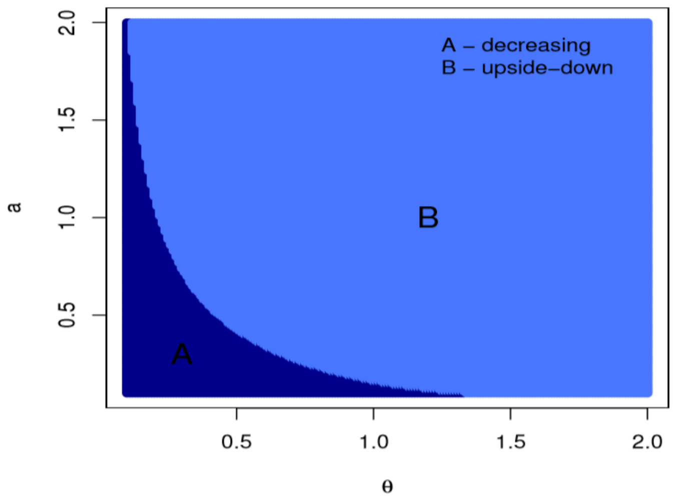

Figure 2: The hrf regions of OLLFr distribution

3 Mathematical properties

3.1 Mixture representation

We provide a useful linear representation for the OLLFr density function. First, we use a power series for $G(x,a,b)^{\theta}(\theta > 0\, real)$ given by

$$\{\exp[-(\frac{a}{x})^{b}]\}^{\theta} = \sum_{k=0}^{\infty} a_k \{\exp[-(\frac{a}{x})^{b}]\}^{k}, \quad (7)$$

where

$$a_k = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{\theta}{j} \binom{j}{k}.$$

For any real $\theta > 0$, we consider the generalized binomial expansion

$$\{1 - \exp[-\theta(\frac{a}{x})^{b}]\}^{\theta} = \sum_{k=0}^{\infty} (-1)^{k} \binom{\theta}{k} \{\exp[-(\frac{a}{x})^{b}]\}^{k}. \quad (8)$$

Inserting (7) and (8) in equation (5), we obtain

$$F(x) = \frac{\sum_{k=0}^{\infty} a_k \{\exp[-(\frac{a}{x})^{b}]\}^{k}}{\sum_{k=0}^{\infty} b_k \{\exp[-(\frac{a}{x})^{b}]\}^{k}},$$

where

$$b_k = a_k + (-1)^{k} \binom{\theta}{k}.$$

the ratio of the two power series can be expressed as

$$F(x) = \sum_{k=0}^{\infty} c_k \{exp[-(\frac{a}{x})^{b}]\}^{k} = \sum_{k=0}^{\infty} c_k\, \Pi_k(x; a, b), \quad (9)$$

where $\Pi_k(x;a,b) = G(x,a,b)^k$ is the Fr cdf with scale parameter $ak^{\frac{1}{b}}$ and shape parameter b and the cofficients $c_k{}'s$ (for $k \geq 0$) are determined from the recurrence equation

$$c_k = \frac{1}{b_0}(a_k + \frac{1}{b_0}\sum_{r=1}^{k} b_r\, c_{k-r}).$$

Upon differentiating (9), the pdf of X can be expressed as

$$f(x) = \sum_{k=0}^{\infty} c_{k+1}\, \pi_{k+1}(x;a,b), \tag{10}$$

where $\pi_{k+1}(x;a,b)$ is the Fr density with scale parameter $a(k+1)^{\frac{1}{b}}$ and shape parameter b. Thus, the OLLFr density can be expressed as a double linear mixture of Fr densities. Consequently, several of its structural properties can be obtained from Equation (10) and those properties of the Fr distribution.

3.2 Moments and cumulants

The n^{th} ordinary moment of X is given by

$$\mu_n' = E(X^r) = \sum_{k=0}^{\infty} c_{k+1} \int_{-\infty}^{\infty} x^n\, \pi_{k+1}(x;a,b)dx,$$

and for any $n < b$,

$$\mu_n' = \sum_{k=0}^{\infty} c_{k+1}\, a^n (k+1)^{\frac{n}{b}} \Gamma\left(1 - \frac{n}{b}\right). \tag{11}$$

Setting $r = 1$ in (11), we have the mean of X. The last integration can be computed numerically for most parent distributions. The skewness and kurtosis measures can be calculated from the ordinary moments using well-known relationships. The n^{th} central moment of X, say M_r, is

$$M_n = E(X-\mu)^n = \sum_{hh=0}^{n} (-1)^h \binom{r}{h} \mu_1'^h \mu_{n-h}'.$$

The cumulants (K_n) of X follow recursively from $K_n = \mu_n' - \sum_{r=0}^{n-1}\binom{n-1}{r-1} K_r \mu_{n-r}'$,where $K_1 = \mu_1', K_2 = \mu_2' - \mu_1'^2, K_3 = \mu_3' - 3\mu_2'\mu_1' + \mu_1'^3$ and so on.

3.3 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 as

$$M_X(t) = \sum_{k=0}^{\infty} c_{k+1}\, M_{k+1}(t) = \sum_{k,n=0}^{\infty} c_{k+1} \frac{t^n a^n}{n!\,([\![k+1)]\!]^{\frac{n}{b}}} \Gamma\left(1-\frac{n}{b}\right), \forall_n < b.$$

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

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

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

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

where the gamma function is well-defined for any non-integer b. Consider the Wright generalized hypergeometric function defined by

$${}_p\mathbf{\Psi}_q \left[\begin{matrix}(\alpha_1, A_1), \dots, (\alpha_p, A_P)\\ (\beta_{1,}B_1), \dots, (\beta_q, B_q)\end{matrix}; x\right] = \sum_{n=0}^{\infty} \frac{\prod_{j=1}^{p} \Gamma(\alpha_j + A_j n)}{\prod_{j=1}^{q} \Gamma(\beta_j + B_j n)} \frac{x^n}{n!}.$$

Then, we have

$$M(t;a,b) = {}_1\mathbf{\Psi}_0 \left[\left(1, -\frac{1}{b}\right); at\right]. \tag{12}$$

Combining expressions (10) and (12), we obtain the mgf of X as

$$M(t) = \sum_{k=0}^{\infty} c_{k+1}\, {}_1\mathbf{\Psi}_0 \left[\left(1, -\frac{1}{b}\right); at(k+1)^{\frac{1}{b}}\right].$$

3.4 Incomplete moment

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 s^{th} incomplete moment, say $I_s(t)$, of X can be expressed from (10), for $n < b$, as

$$\begin{aligned} I_n(t) &= \sum_{k=0}^{\infty} c_{k+1} \int_{-\infty}^{t} x^n \pi_{k+1}(x) dx \\ &= \sum_{k=0}^{\infty} c_{k+1}\, a^n (k+1)^{\frac{n}{b}} \gamma\left(1-\frac{n}{b}, (k+1)\left(\frac{a}{t}\right)^b\right), \forall_n < b, \end{aligned}$$

where

$$\gamma(\tau, t) = \int_{-\infty}^{t} z^{\tau-1} e^{-z} dz$$

denotes the complementary incomplete gamma function. The mean deviations about the mean

$$\delta_1 = E(|X-\mu_1'|) = 2\mu_1' F(\mu_1') - 2I_1(\mu_1')$$

and about the median

$$\delta_2 = E(|X-M|) = \mu_1' - 2I_1(M),$$

where $\mu_1' = E(X)$, $M = \text{Median}(X) = Q(0.5)$ is the median, $F(\mu_1')$ is easily calculated from (5) and $I_1(t)$ is the first incomplete moment given by the last Equation with $s = 1$. The general formula for $I_1(t)$ is

$$I_1(t) = \sum_{k=0}^{\infty} c_{k+1} a (k+1)^{\frac{1}{b}} \gamma(1-\frac{1}{b}, (k+1)(\frac{a}{t})^b).$$

3.5 Residual life function and life expectation at age *t*

The n^{th} moment of the residual life is given by

$$m_n(t) = E[(X-t)^n | X > t|] \; \forall \, n = 1,2,...$$

The n^{th} moment of the residual life of X can be expresed as

$$m_n(t) = \frac{1}{1-F(t)} \int_t^{\infty} (x-t)^n \, dF(x),$$

therefore

$$m_n(t) = \sum_{k=0}^{\infty} \frac{c_{k+1}^{\star} a^n (k+1)^{\frac{n}{b}}}{1-F(t)} \gamma(1-\frac{n}{b}, (k+1)(\frac{a}{t})^{b}), \forall n < b,$$

where

$$c_{k+1}^{\star} = c_{k+1} + \sum_{r=0}^{n} \binom{n}{r} (-t)^{n-r}.$$

The mean residual life (MRL) function or the life expectation at age t is 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 equation.

3.6 Reversed residual life function and mean inactivity time

The n^{th} moment of the reversed residual life can be expressed as

$$M_n(t) = E[(X-t)^n | X \le t|], \forall t > 0 \; and \; n = 1,2,...$$

We obtain

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

Then, the n^{th} moment of the reversed residual life of X becomes

$$M_n(t)=\sum_{k=0}^{\infty}\frac{c_{k+1}^{\star\star}a^n(k+1)^{\frac{n}{b}}}{F(t)}\gamma(1-\frac{n}{b},(k+1)(\frac{a}{t})^{b}),\forall n<b,$$

where

$$c_{k+1}^{\star\star}=c_{k+1}\sum_{r=0}^{n}(-1)^r\binom{n}{r}t^{n-r}.$$

The mean inactivity time or the mean reversed residual life function, is given by $M_1(t)=E[(t-X)|X\leq t]$ and represents the waiting time elapsed since the failure of an item on condition that this failure had occurred in $(0,t)$. The mean inactivity time of the OLLFr model can be obtained easily via setting $n=1$ in the above equation; the n^{th} moment of the residual life and n^{th} moment of the reversed residual life uniquely determines $F(x)$.

4 Characterizations

This section takes up various characterizations of OLLFr distribution.These characterizations are presented in two directions: (i) based on a simple relationship between two truncated moments and (ii) in terms of the reverse hazard function. It should be pointed out that due to the nature of the OLLFr distribution our characterizations may be the only possible ones for this distribution. We present our characterizations (i) and (ii) in two subsections.

4.1 Characterizations based on truncated moments

We employ 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 Theorem 1 is on the interior of H. We like to mention that this kind of characterization based on a truncated moment is stable in the sense of weak convergence (see, Glänzel 1990).

Proposition 4.1. Let X : $\Omega\to(0,\infty)$ be a continuous random variable and let $q_1(x)=\exp\left[(\theta+1)\left(\frac{a}{x}\right)^b\right][\exp\left[-\theta\left(\frac{a}{x}\right)^b\right]+\{1-\exp[-\left(\frac{a}{x}\right)^b]\}^{\theta}]$ and $q_2(x)=q_1(x)\{1-\exp[-\left(\frac{a}{x}\right)^b]\}$ for $x>0$. The random variable X belongs to the family (6) if and only if the function η defined in Theorem1 has the form

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

Proof. Let X be a random variable with pdf (6), then

$$\left(1-F(x)\right)E[q_1(X)|X\geq x]=\{1-\exp[-\left(\frac{a}{x}\right)^b]\}^{\theta},$$

and

$$\left(1-F(x)\right)E[q_2(X)|X\geq x]=\frac{\theta}{\theta+1}\{1-\exp[-\left(\frac{a}{x}\right)^b]\}^{\theta+1}.$$

Further,

$$\eta(x)q_1(x)-q_2(x)=-\frac{q_1(x)}{\theta+1}\left\{1-exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}<0\ for\ x>0.$$

Conversely, if η is given as above, then

$$s'(x)=\frac{\eta' q_1(x)}{\eta(x)q_1(x)-q_2(x)}=\frac{\theta ba^b x^{-(b+1)}\exp[-\left(\frac{a}{x}\right)^b]}{1-\exp[-\left(\frac{a}{x}\right)^b]},x>0,$$

and hence

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

Now, according to Theorem 1, X has density (6)

Corollary 4.1. Let X : Ω → (0, ∞) be a continuous random variable and let q1 be as in Proposition 4.1. Then, X has pdf (6) if and only if there exist functionsq_2 and η defined in Theorem 1 satisfying the differential equation

$$\frac{\eta'(x)q_1(x)}{\eta(x)q_1(x)-q(x)}=\frac{\theta ba^b x^{-(b+1)}\exp[-(\frac{a}{x})^b]}{1-\exp[-(\frac{a}{x})^b]},x>0.$$

The general solution of the differential equation in Corollary 4.1 is

$$\eta(x)=\{1-exp[-(\frac{a}{x})^b]\}^{-1}[-\int\theta ba^b x^{-(b+1)}\,exp\left[-\left(\frac{a}{x}\right)^b\right]\left(q_1(x)\right)^{-1}q_2(x)dx+D],$$

where D is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 4.1 with $D=0$. Clearly, there exist other triplet of functions (q_1,q_2,η) satisfying the conditions of Theorem 1.

4.1 Characterization in terms of the reverse hazard function

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

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

In this subsection we present characterization of OLLFr distribution in terms of the reverse hazard function.

Proposition 4.2. Let X : Ω → (0, ∞) be a continuous random variable. Then, X has pdf (6) if and only if its reverse hazard function $r_F(x)$ satisfies the differential equation

$$r_F'(x) + (b+1)x^{-1}r_F(x) = \theta b a^b x^{-(b+1)}$$
$$\times \frac{d}{dx}\left\{\frac{\left\{1-\exp[-\left(\frac{a}{x}\right)^b]\right\}^{\theta-1}}{\exp[-\theta\left(\frac{a}{x}\right)^b] + \left\{1-\exp[-\left(\frac{a}{x}\right)^b]\right\}^{\theta}}\right\}, x > 0.$$

Proof: If X has pdf (6), then clearly the differential equation holds. Now, if the differential equation holds, then

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

from which we arrive at the reverse hazard function of (6).

Remark 4.1. For $\theta = 1$, we have the following simple differential equation

$$r_F'(x) + (b+1)x^{-1}r_F(x) = 0.$$

5 Estimation and Simulation

5.1 Maximum likelihood estimation

If X follows the OLLFr distribution with parameter vector $\mathbf{\Phi} = (\theta, a, b)^T$, the log-likelihood for $\mathbf{\Phi}$ from a single observation x of X is given by

$$\ell(\mathbf{\Phi}) = log(\theta) + log(b) + blog(a) - (b+1)\,log(x) + log\omega$$
$$+(\theta-1)\log[w(1-w)] - 2\log\left[\omega^{\theta} + (1-\omega)^{\theta}\right],$$

where $\omega = exp[-(\frac{a}{x})^b]$. The components of the unit score vector $U = U(\mathbf{\Phi}) = (\partial\theta/\partial\ell, \partial a/\partial\ell, \partial b/\partial\ell)^T = (U(\theta), U(a)U(b))^T$ are given by

$$U(\theta) = \frac{1}{\theta} + \log[\omega(1-\omega)] - 2\frac{\omega^{\theta}\log(\omega) + (1-\omega)^{\theta}\log(1-\omega)}{\omega^{\theta}(1-\omega)^{\theta}},$$

$$U(a) = \frac{b}{a} + \frac{m}{\omega} + (\theta-1)\frac{m-2m\omega}{\omega(1-\omega)} - 2\frac{\theta m\omega^{\theta-1} - \theta m(1-\omega)^{\theta-1}}{\omega^{\theta} + (1-\omega)^{\theta}}$$

and

$$U(b) = \frac{1}{b} + \log(a) - \log(x) + \frac{q}{\omega} + (\theta-1)\frac{q-2q\omega}{\omega(1-\omega)} - 2\frac{\theta q\omega^{\theta-1} - \theta q(1-\omega)^{\theta-1}}{\omega^{\theta} + (1-\omega)^{\theta}}$$

where

$$m = -ba^{b-1}x^{-b}\exp[-\left(\frac{a}{x}\right)^b] and\ q = -\left(\frac{a}{x}\right)^b \exp\left[-\left(\frac{a}{x}\right)^b\right]\log\left(\frac{a}{x}\right).$$

For a random sample $x = (x_1, \dots, x_n)^T$ of size n from X, the total log-likelihood is

$$\ell_n(\mathbf{\Phi}) = \sum_{i=0}^{n} \ell^{(i)}(\mathbf{\Phi}),$$

where $\ell^{(i)}(\mathbf{\Phi})$is the log-likelihood for the i^{th} observation. The total score function is

$$U_n = \sum_{i=0}^{n} U^{(i)},$$

where $U^{(i)}$ has the form given before. Maximization of $\ell(\mathbf{\Phi})$ (or$\ell_n(\mathbf{\Phi})$) can be easily performed using well-established routines such as the nlm or optimize in the R statistical package. Setting these equations to zero,$\mathrm{U}(\mathbf{\Phi}) = 0$,and solving them simultaneously gives the MLE$\widehat{\Phi}$ of $\mathbf{\Phi}$.

These equations cannot be solved analytically and statistical software can be used to evaluate them numerically using iterative techniques such as the Newton-Raphson algorithm.

5.1 Simulation Study

Here, the performance of the maximum likelihood method is investigated for estimating the OLLFr parameters by means of Monte Carlo simulation study with 10,000 replications. The coverage probabilities (CPs), mean square error (MSES) and the bias of the parameter estimates,estimated average lengths (ALs) are calculated by means of R software. We generate $N = 10{,}000$ samples of sizes $N = 50, 55, \ldots, 1000$ from the OLLFr distribution with $\alpha = 2, a = 2, b = 2$. Let $\hat{\alpha}, \hat{a}, \hat{b}$ be the MLEs of the new model parameters and $(s_{\hat{\alpha}}, s_{\hat{a},} s_{\hat{b}})$ be the standard errors of the MLEs. The equations for the above measures are given by,

$$\widehat{Bias}_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}_i - \epsilon),$$

$$\widehat{MSE}_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}_i - \epsilon)^2,$$

$$CP_\epsilon(n) = \frac{1}{N}\sum_{i=1}^{N} I\left(\hat{\epsilon}_i - 1.95996 s_{\hat{\epsilon}_i}, \hat{\epsilon}_i + 1.95996 s_{\hat{\epsilon}_i}\right),$$

$$AL_\epsilon(n) = \frac{3.919928}{N}\sum_{i=1}^{N}\hat{\epsilon}_i .$$

where $\epsilon = \alpha, a, b$. Figure 3 displays the numerical results for these measures. Based on Figure 3, we conclude:

- ✓ The estimated biases decrease when the sample size n increases,
- ✓ The estimated MSEs decay toward zero as n increases,
- ✓ The CPs are near 0.95 and approach the nominal value when the sample size increases,

✓ The ALs decrease for all parameters when the sample size increases.
These results reveal the consistency property of the MLEs.

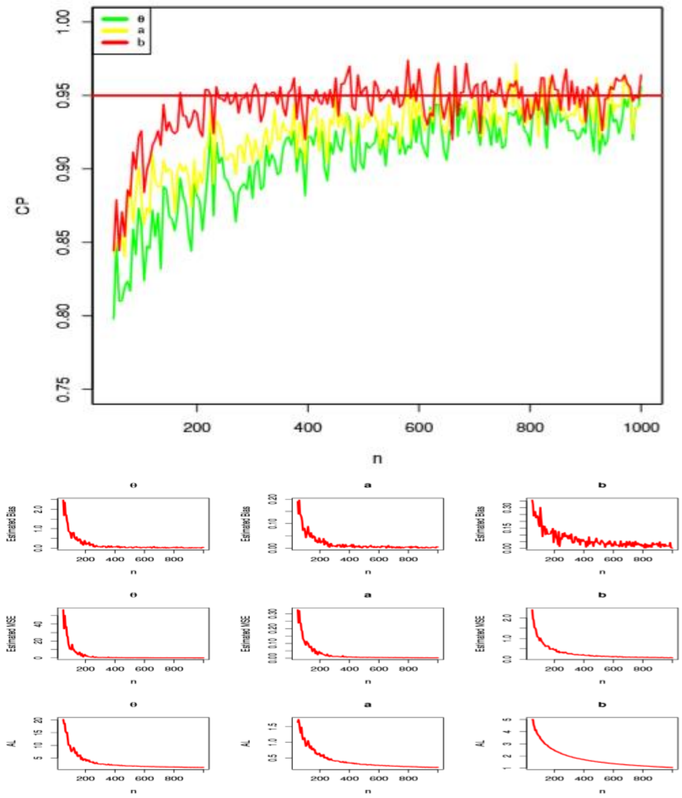

Figure 3: Estimated CPs, biases, MSEs and ALs for the selected parameter values.

6 The log odd log-logistic Fréchet regression model

Consider the OLLFr distribution with four parameters given in (6) and let X be a random variable with OLLFr distribution (6) and $Y = log(X)$. The density function of Y(for $y \in \Re$)obtained by replacing $b = 1/\sigma$ and $a = exp(\mu)$ can be expressed as

$$f(y) = \frac{\theta}{\sigma}\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\} \times \left(\left[\exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\}\right]\left(1 - \left[\exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\}\right]\right)\right)^{\theta-1} \times \{[\exp\{-\exp\{-(\frac{y-\mu}{\sigma})\}\}]^{\theta} + (1-[\exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\}])^{\theta}\}^{-2}, \quad (13)$$

where $\mu \in \Re$ is the location parameter, $\sigma > 0$ is the scale parameter and $\alpha > 0$ is the shape parameter. We refer to equation (13) as the Log-OLLFr (LOLLFr) distribution,say $Y \sim \text{LOLLFr}(\alpha, \mu, \sigma)$. The survival function corresponding to (13) is given by

$$s(y)=\frac{(1-[exp\{-exp\{-\left(\frac{y-\mu}{\sigma}\right)\}\}])^{\theta}}{[\exp\{-\exp\{-(\frac{y-\mu}{\sigma})\}\}]^{\theta}+(1-[\exp\left\{-\exp\left\{-\left(\frac{y-\mu}{\sigma}\right)\right\}\right\}])^{\theta}} \tag{14}$$

and the hrf is simply $h(y)=f(y)/S(y)$. The standardized random variable $Z=(Y-\mu)/\sigma$ has density function

$$\begin{aligned} f(z) = \theta\exp\{-z\}\exp\{-\exp\{-z\}\}([\exp\{-\exp\{-z\}\}](1 \\ -[\exp\{-\exp\{-z\}\}]))^{\theta-1}.\times\{[\exp\{-\exp\{-z\}\}]^{\theta} \\ +(1-[\exp\{-\exp\{-z\}\}])^{\theta}\}^{-2}. \end{aligned} \tag{15}$$

Based on the LOLLFr density, we propose a linear location-scale regression model linking the response variable y_i and the explanatory variable vector $v_i^T=(v_{i1},\dots,v_{ip})$ given by

$$y_i=V_i^T\beta+\sigma z_i, i=1,\dots,n, \tag{16}$$

where the random error z_i has density function (15), $\beta=(\beta_1,\dots,\beta_p)^T$, and $\sigma>0,\alpha>0$ and $\beta>0$ are unknown parameters. The parameter $\mu_i=v_i^T\beta$ is the location of y_i. The location parameter vector $\mu=(\mu_1,\dots,\mu_n)^T$ is represented by a linear model $\mu=V\beta$, where $\mathrm{V}=(v_1,\dots,v_n)^T$ is a known model matrix.

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. The log-likelihood function for the vector of parameters $\Theta=(\alpha,\beta,\sigma,\beta^T)^T$ from model (16) has the form $l(\Theta)=\sum_{i\in F}l_i(\Theta)+\sum_{i\in C}l_i^{(c)}(\Theta)$,where $l_i(T)=log[f(y_i)], l_i^{(c)}(\Theta)=log[S(y_i)], f(y_i)$ is the density (13)and $S(y_i)$ is the survival function (14) of Y_i. The total log-likelihood function for Θ is given by

$$\begin{aligned} \ell(\Theta) &= rlog\left(\frac{\theta}{\sigma}\right)-\sum_{i\in F}z_i-\sum_{i\in F}\exp(z_i) \\ &+(\theta-1)\sum_{i\in F}\log([\exp\{-\exp\{-z_i\}\}](1-[\exp\{-\exp\{-z_i\}\}])) \\ &-2\sum_{i\in F}log\{[\exp\{-\exp\{-z_i\}\}]^{\theta}+(1-[\exp\{-\exp\{-z_i\}\}])^{\theta}\}. \\ &+\theta\sum_{i\in C}^{i\in F}\log(1-[\exp\{-\exp\{-z_i\}\}]) \\ &-\sum_{i\in C}\log\{[\exp\{-\exp\{-z_i\}\}]^{\theta}+(1-[\exp\{-\exp\{-z_i\}\}])^{\theta}\}, \end{aligned} \tag{17}$$

where $z_i=(y_i-\mu_i)/\sigma$, and r is the number of uncensored observations (failures). The MLE $\widehat{\Theta}$ of the vector of unknown parameters can be evaluated by maximizing the log-likelihood (17). The R software is used to estimate $\widehat{\Theta}$.

6.1 Residual analysis

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

6.1.1 Martingale residual

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

$$rM_i = \begin{cases} 1+\log\left(\dfrac{(1-[\exp\{-\exp\{-z_i\}\}])^{\theta}}{[\exp\{-\exp\{-z_i\}\}]^{\theta}+(1-[\exp\{-\exp\{-z_i\}\}])^{\theta}}\right), if\ i\in F, \\ \log\left(\dfrac{(1-[\exp\{-\exp\{-z_i\}\}])^{\theta}}{[\exp\{-\exp\{-z_i\}\}]^{\theta}+(1-[\exp\{-\exp\{-z_i\}\}])^{\theta}}\right), if\ i\in C, \end{cases} \quad (18)$$

where $u_i = 2\Phi[exp(z_i\sqrt{2}/2)]$ *and* $z_i = (y_i - \mu_i)/\sigma$.

6.1.2 Modifted 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). Th modified deviance residual for LOLLFr model is

$$rD_i = \begin{cases} sign(rM_i)\{-2[rM_i+\log(1-rM_i)]\}^{1/2}, if\ i\in F \\ sign(rM_i)\{-2rM_i\}^{1/2}, \qquad if\ i\in C, \end{cases} \quad (19)$$

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

7 Applications

In this section, four real data sets are used to compare OLL-Fr model with well-known extensions of Fréchet distribution given in Table 1.The first two data sets are used to demonstrate the univariate data fitting performance of OLL-Fr distribution. The third and fourth data sets are used to investigate the usefulness of proposed distribution in survival analysis. The optim function is used to estimate the unknown model parameters. The MLEs, estimated −A, standard errors of MLEs and Akaike Information Criteria (AIC) are reported in Table 2 and 4. The lower the values of these criteria show the better fitted model on data sets. The histograms with fitted pdfs are provided for visual comparison of the fitted distribution functions. Moreover, fitted pdfs, hrfs, survival function and P-P plots of best fitted models are displayed in Figure 4(b) and 5(b).

Table 1: The competitive distributions for OLLFr model.

Distribution	Abbreviation	References
Weibull Fr´echet Distribution	W-Fr	Afify et al. (2016)
Beta Fr´echet Distribution	Beta-Fr	Nadarajah and Gupta (2004)
Kumaraswamy Fr´echet Distribution	Kum-Fr	Mead and Abd-Eltawab (2014)
Fr´echet	Fr	-

7.1 First Application

The first data set were used by Birnbaum and Saunders (1969) and correspond to the fatigue time of 101 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second (cps). The data set is 70, 90, 96, 97, 99, 100, 103, 104, 104, 105, 107, 108,108, 108, 109, 109, 112, 112 ,113, 114, 114, 114, 116, 119, 120, 120, 120, 121, 121, 123, 124, 124,124, 124, 124 ,128, 128, 129, 129, 130, 130, 130, 131, 131, 131, 131, 131, 132, 132, 132, 133, 134,134, 134, 134, 134, 136, 136, 137, 138, 138, 138, 139, 139, 141, 141, 142, 142, 142 ,142, 142,142, 144, 144, 145, 146, 148, 148, 149, 151, 151, 152, 155, 156, 157, 157 ,157, 157, 158, 159, 162,163, 163, 164, 166, 166, 168, 170, 174, 196, 212.

Table 2 shows the estimated parameters and their standard errors,$-\ell$ and AIC values. Based on the figures in Table 2, it is clear that OLL-Fr model provides the best fit for this data set. Figure 4(a) displays the estimated pdfs of the fitted models. Figure 4(b) displays the fitted pdf, hrf, survival function and P-P plot of OLL-Fr distribution. Figure 4 proves that OLL-Fr distribution provides the superior fit to symmetric data sets.

Table 2: MLEs and their SEs (reported in second line) of the fitted models and goodness-of-fit statistics for first data set

Models	α	θ	a	b	$-\ell$	AIC
OLL-Fr		169.503	0.041	0.045	455.774	917.549
		14.098	0.003	0.001		
W-Fr	1.786	1.632	134.521	2.463	457.770	923.540
	3.705	0.513	44.409	0.911		
Kum-Fr	32.650	61.596	35.915	1.516	456.254	920.509
	180.018	86.715	130.911	0.479		
Beta-Fr	10.275	58.651	273.301	0.895	457.564	923.127
	12.303	41.159	130.136	0.333		
Fr			120.784	5.056	475.186	954.371
			2.526	0.325		

In addition, Likelihood Ratio (LR) test is used to compare the OLL-Fr distribution with its sub-models. For example, the test of $H_0: \theta = 1$ against $H_1: \theta \neq 1$ is equivalent to comparing the OLL-Fr and Fr distributions with each other. The LR test statistic can be written as follows,

$$LR = 2\left[\ell(\hat{\theta}, \hat{a}, \hat{b}) - \ell\left(1, \widehat{a^*}, \widehat{b^*}\right)\right],$$

where $\widehat{a^*}$ and $\widehat{b^*}$ are the ML estimators of a and b, respectively, obtained under H_0. Under the regularity conditions and if H_0 is assumed to be true, the LR test statistic converges in distribution to a chi square with r degrees of freedom, where r equals the difference between the number of parameters estimated under H_0 and the number of parameters estimated in general, (for $H_0: \theta = 1,$ we have $r = 1$). Table 3 shows the LR statistics and the corresponding p-values for the first data set.

Table 3: The LR test results for first data set.

	Hypotheses	LR	p-value
OLL-Fr versus Fr	$H_0: \theta = 1$	38.824	< 0.001

As seen in Table 3, the p-values are smaller than 0.05, so the null hypotheses are rejected. Hence, OLL-Fr model fits the first data set better than its sub-model according to the LR test results.

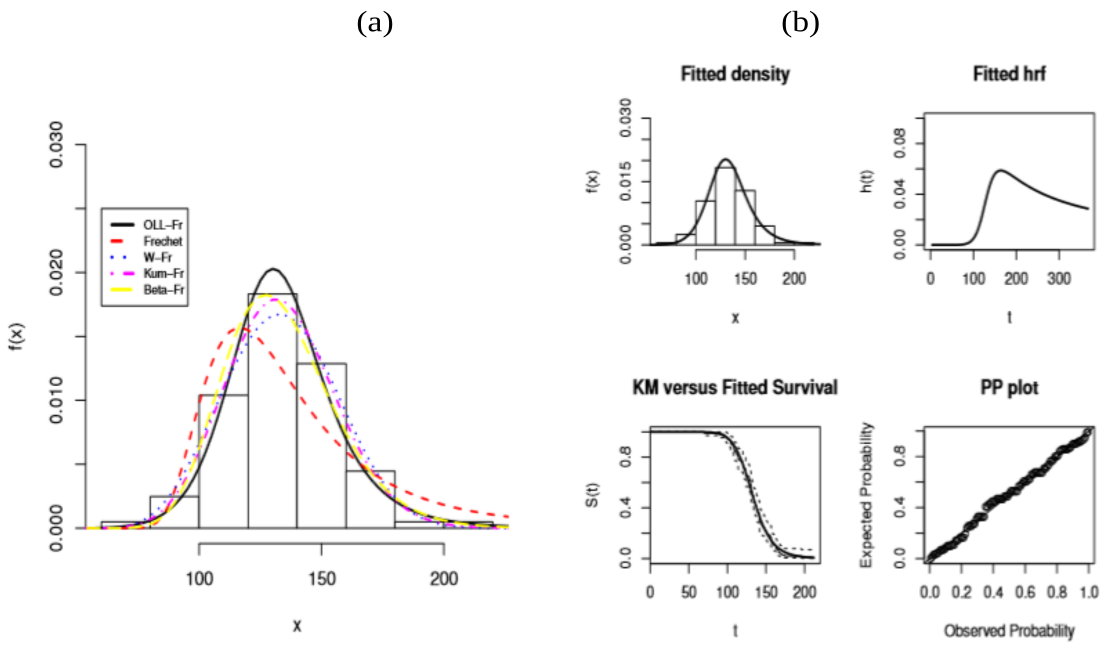

Figure 4: (a) Fitted densities of fitted models and (b) fitted functions of OLL-Fr model for first set.

7. Second Application

The second data set is due to Bjerkedal (1960) and refers to the survival times (in days) of 72 guinea pigs infected with different amout of virulent tubercle bacilli. The data are: 0.1, 0.33, 0.44, 0.56, 0.59, 0.72, 0.74, 0.77, 0.92, 0.93, 0.96, 1, 1, 1.02, 1.05, 1.07, 1.07, 1.08, 1.08,1.08, 1.09, 1.12, 1.13, 1.15, 1.16, 1.2, 1.21, 1.22, 1.22, 1.24, 1.3, 1.34, 1.36, 1.39, 1.44, 1.46, 1.53,1.59, 1.6, 1.63, 1.63, 1.68, 1.71, 1.72, 1.76, 1.83, 1.95, 1.96, 1.97, 2.02, 2.13, 2.15, 2.16, 2.22, 2.3,2.31, 2.4, 2.45, 2.51, 2.53, 2.54, 2.54, 2.78, 2.93, 3.27, 3.42, 3.47, 3.61, 4.02, 4.32, 4.58, 5.55.

Table 4 shows the estimated parameters and their standard errors, $-\ell$ and AIC values. Based on the figures in Table 4, OLL-Fr model provides the best fit for second data set. Figure 5(a) displays the estimated pdfs of the fitted models. Figure 5(b) displays the fitted pdf, hrf, survival function and P-P plot of OLL-Fr distribution. Figure 5 proves that OLL-Fr distribution provides the superior fit to left-skewed data sets.

Table 4: MLEs and their SEs (reported in second line) of the fitted models and goodness-of-fit statistics for second data set

Models	α	θ	a	b	$-\ell$	AIC
OLL-Fr		15.678	0.105	0.137	94.444	194.889
		12.188	0.216	0.106		
W-Fr	1.869	3.069	1.119	0.403	95.020	198.040
	3.705	0.513	44.409	0.911		
Kum-Fr	3.021	6.084	0.943	0.710	101.085	210.171
	1.857	2.122	0.804	0.085		
Beta-Fr	20.748	79.803	15.937	0.194	98.449	204.898
	17.246	46.119	14.213	0.060		
Fr			1.058	1.173	118.166	240.332
			0.113	0.084		

Table 5 shows the LR statistics and the corresponding p-values for the second data set. From Table 5, we conclude that the OLL-Fr model fits the second data set better than the its sub-model.

Table 5: The LR test results for second data set.

	Hypotheses	LR	p-value
OLL-Fr versus Fr	$H_0: \theta = 1$	47.444	< 0.001

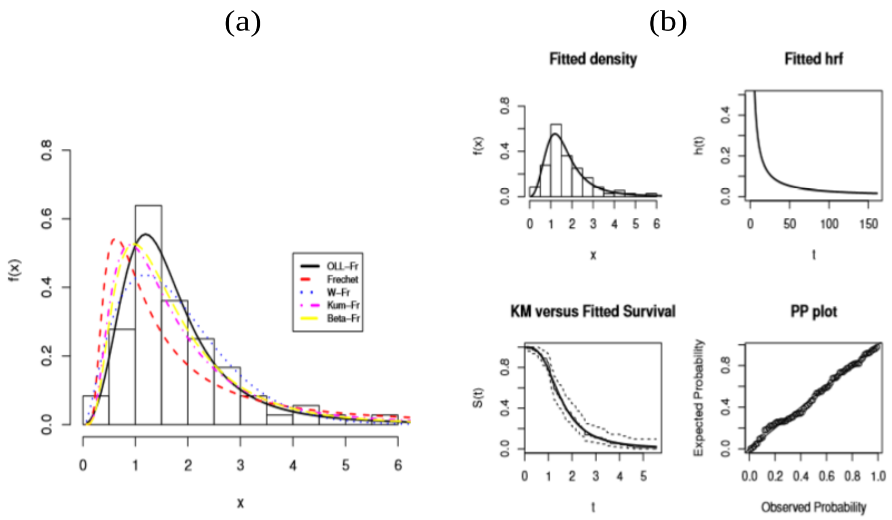

Figure 5: (a) Fitted densities of fitted models and (b) fitted functions of OLL-Fr model for first set.

7.3 Third Application

Lawless (2003) reported an experiment in which specimens of solid epoxy electrical-insulation were studied in an accelerated voltage life test. The sample size is $n =$

60, the percentage of censored observations is 10% and there are three levels of voltage: 52.5, 55.0 and 57.5. The variables involved in the study are:x_i-failure times for epoxy insulation specimens(in min);c_i-censoring indicator(0=censoring,1=lifetime observed);v_{i1}-voltage (kV).

These data set was used by Pescim et al. (2013) and Pescim et al. (2017) for illustrating the log-Beta-GHN (LBGHN) and log-odd log-logistic GHN (LOLLGHN) regression models. Pescim et al. (2013) compared the LBGHN regression model with log-GHN (LGHN) and log-Weibull models. In this section we compare the LOLLFr regression model with models reported in Pescim et al. (2013) and Pescim et al. (2017). The regression model fitted to the voltage data set are given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma z_i \qquad (20)$$

respectively, where the random variable yi follows the LOLLFr distribution given in (13). The results are presented in Table 6. The MLEs of the model parameters and their SEs and the values of the AIC and BIC statistics are listed in Table 6.

Table 6: MLEs of the parameters to the voltage data for LOLLFr, LOLL-GHN LBGHN, LGHN and Log-Weibull regression models, the corresponding SEs (in parentheses), p-values in and the AIC and BIC statistics.

Model	α	θ	σ	β0	β1	AIC	BIC
LOLLFr		1.871	1.436	16.068	-0.184	164.3978	172.7751
		(1.529)	(1.072)	(3.311)	(0.061)		
				[<0.001]	[<0.001]		
LOLLGHN	34.255		22.028	19.279	-0.522	166.400	174.800
	(6.869)		(4.423)	(2.461)	(0.140)		
				[<0.001]	[<0.001]		
LBGHN	102.140	1.564	5.306	10.632	-0.201	167.100	177.500
	(3.989)	(0.672)	(0.666)	(3.304)	(0.056)		
				[0.002]	[0.001]		
LGHN			0.778	23.637	-0.301	178.800	185.100
			(0.089)	(2.928)	(0.053)		
				[<0.001]	[<0.001]		
Log-Weibull			0.845	22.032	-0.275	173.400	179.700
			(0.090)	(3.046)	(0.055)		
				[<0.001]	[<0.001]		

Based on the figures in Table 6, we conclude that the fitted LOLLFr regression model has the lower AIC and BIC values. Figure 6 provides the plots of the empirical and estimated survival function for the LOLLFr regression model. We can conclude with these plot that LOLLFr regression model provides a good fit to these data.

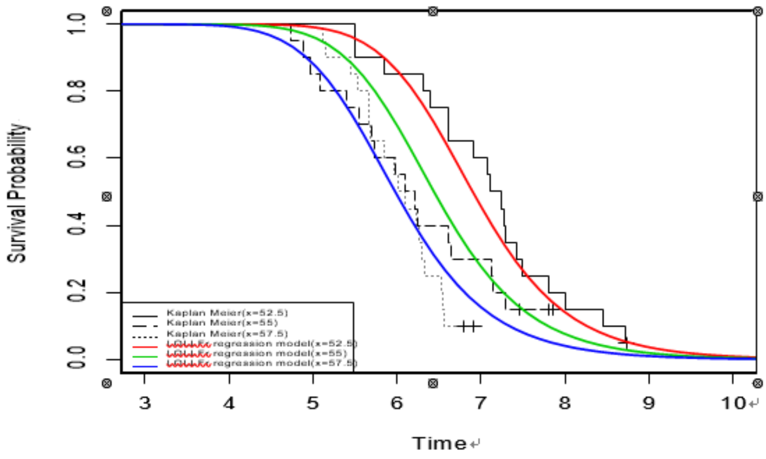

Figure 6: Estimated survival function of OLLFr regression model and empirical survival for the voltage data considering the voltage levels: $x_{i1} = 52.5; 55.0\ and\ 57.5$.

Residual Analysis of LOLLFr model for voltage data set

Figure 7 displays the index plot of the modified deviance residuals and its Q-Q plot against to $N(0,1)$ quantiles. Based on the Figure 7, we conclude that none of observed values appears as a possible outliers. Thus, it is clear that the fitted model is appropriate for these data set.

(a) (b)

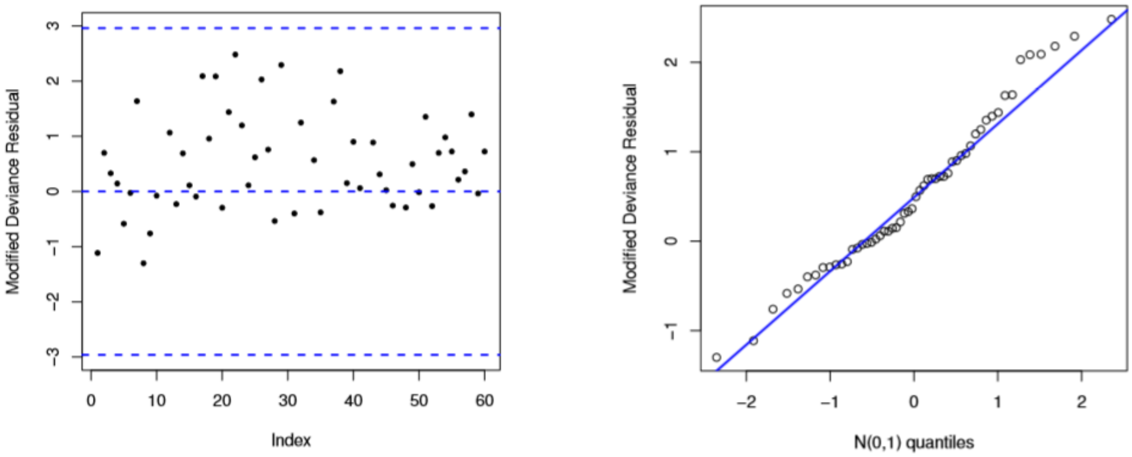

Figure 7:(a) Index plot of the modified deviance residual and (b) Q-Q plot for modified deviance residual.

7.4 Fourth Application

Recently, Brito et al. (2017) introduced the Log-Topp-Leone odd log-logistic-Weibull (Log- TLOLL-W) regression model. Brito et al. (2017) used the Stanford heart transplant data set to prove the usefulness of Log-TLOLL-W regression model. Here, we use the same data set to demonstrate the flexibility of LOLLFr regression model against to Log-TLOLL-W regression model. These data set is available in p3state.msm package of R software. The sample size is $n = 103$, the percentage of censored observations is 27%. The aim of this study is to relate the survival times (t) of patients with the following explanatory variables: x_1-year of acceptance to the program;x_2-age of patient (in years);x_3- previous surgery status $(1 = yes, 0 = no)$;x_4-transplant indicator (1 = yes, 0 = no); c_1- censoring indicator (0 =censoring, 1 =lifetime observed).

The regression model fitted to the voltage data set is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \sigma z_i \quad (21)$$

respectively, where the random variable y_i follows the LOLLFr distribution given in (13).

The results for above regression models are presented in Table 7. The MLEs of the model parameters and their SEs, p values and $-\ell$, AIC and BIC statistics are listed in Table 7. Based on the figures in Table 7, LOLLFr model has the lowest values of the $-\ell$, AIC and BIC statistics. Therefore, it is clear that LOLLFr regression model outperforms among others for these data set. According to results of LOLLFr regression model, β_0, β_1 and β_2 are statistically significant at 5% level.

Table 7: MLEs of the parameters to Stanford Heart Transplant Data for Log-Weibull, Log- TLOLL-W and LOLLFr regression models with corresponding SEs, p-values and $-\ell$, AIC and BIC statistics.

	Models								
	Log-Weibull			Log-TLOLL-W			LOLLFr		
Parameters	Estimate	S.E.	P-value	Estimate	S.E.	P-value	Estimate	S.E.	P-value
α	-	-	-	2.34	3.546	-	-	-	-
θ	-	-	-	24.029	3.015	-	2.078	0.79061	-
σ	1.478	0.133	-	9.68	12.526	-	2.886	0.954557	-
β_0	1.639	6.835	0.811	-0.645	8.459	0.939	1.252	0.5616	0.02579
β_1	0.104	0.096	0.279	0.074	0.097	0.448	0.181	0.09682	0.96156
β_2	-0.092	0.02	<0.001	-0.053	0.02	0.009	-0.047	0.0183	0.01021
β_3	1.126	0.658	0.087	1.676	0.597	0.005	-0.151	0.50161	0.7633
β_4	2.544	0.378	<0.001	2.394	0.384	<0.001	0.551	0.26825	0.03997
$-\ell$	171.2405			164.684			160.9328		
AIC	354.481			345.368			335.9656		
BIC	370.2894			366.4458			354.3088		

Residual Analysis of LOLLFr model for Stanford heart transplant data set

Figure 8 displays the index plot of the modified deviance residuals and its Q-Q plot against to $N(0,1)$ quantiles for Stanford heart transplant data set. Based on Figure 8, we conclude that none of observed values appears as possible outliers. Therefore, the fitted model is appropriate for these data set.

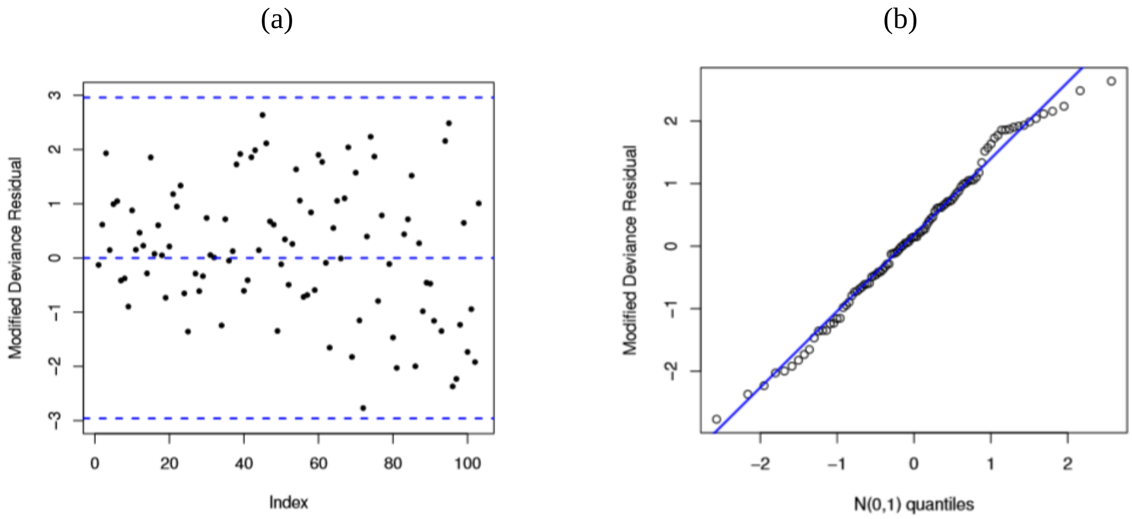

Figure 8:(a) Index plot of the modified deviance residual and (b) Q-Q plot for modified deviance residual.

8. Concluding remarks

A new log-location regression model with applications to voltage and Stanford heart transplant data sets is presented and studied. The martingale and modified deviance residuals to detect outliers and evaluate the model assumptions are defined. The new model can be very useful in analysing real data and provides more realistic fits than other regression models like the log odd log-logistic generalized half-normal, the log beta generalized half-normal, the log generalized half-normal, the log-Topp-Leone odd log-logistic-Weibull and the log-Weibull moels. Characterizations based on truncated moments as well as in terms of the reverse hazard function are presented. Based on the index plot of the modified deviance residual and the Q-Q plot for modified deviance residual we note that OLLFr model is more appropriate for voltage and Stanford heart transplant data sets than all the existing regression models.The maximum likelihood method is discussed to estimate the model parameters by means of a graphical Monte Carlo simulation study.The flexibility of the new model illustrated by means of the four real data sets.

References

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

[2] Birnbaum, Z. W. and Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. Journal of Applied Probability, 6(2), 328-347

[3] Bjerkedal, T. (1960). Acquisition of Resistance in Guinea Pies infected with Different Doses of Virulent Tubercle Bacilli. American Journal of Hygiene, 72(1), 130-48.

[4] Brito, E., Cordeiro, G. M., Yousof, H. M., Alizadeh, M. and Silva, G. O. (2017). The Topp-Leone odd log-logistic family of distributions. Journal of Statistical Computation and Simulation, 1-19.

[5] 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 Univ Press.

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

[7] Fréchet, M. (1927). Sur la loi de probabilite de lecart maximum. Ann. de la Soc. polonaisede Math, 6, 93–116.

[8] 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.

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

[10] Gleaton, J.U. and Lynch, J.D. (2006). Properties of generalized log-logistic families of lifetime distributions. Journal of Probability and Statistical Science, 4, 51-64.

[11] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une serie aleatoire.Annals of Mathematics, Volume 44.

[12] Harlow, D. G. (2002). Applications of the Fréchet distribution function ,Int.J. Mater.Prod. Technol, 17, 482–495.

[13] Kotz, S. and Johnson, N. L. (1992). Breakthroughs in Statistics: Foundations and basic theory. Springer,Volume 1.

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

[15] 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.

[16] Lawless, J. F. (2003). Statistical models and methods for lifetime data,Wiley Series in Probability And Statistics, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition.

[17] Mansoor, M., Tahir, M. H., Alzaatreh, A., Cordeiro, G. M., Zubair, M. and Ghazali, S.S. (2016). An extended Fréchet distribution: properties and applications. Journal of Data Science, 14(1), 167-188.

[18] Mead, M. E. and Abd-Eltawab, A. R. (2014). A note on Kumaraswamy Fréchet distribution. Australian Journal of Basic and Applied Sciences, 8(15), 294-300.

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

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

[21] Pescim, R. R., Ortega, E. M., Cordeiro, G. M., & Alizadeh, M. (2017). A new log-location Regression model: estimation, influence diagnostics and residual analysis. Journal of Applied Statistics, 44(2),233-252.

[22] Pescim, R.R., Ortega, E.M.M., Cordeiro, G.M., Demetrio, C.G.B. and Hamedani, G.G. (2013). The Log-Beta Generalized Half-Normal Regression Model. Journal of Statistical Theory and Applications,12, 330-347.

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

[24] Von Mises, R. (1936). La distribution de la plus grande de n valeurs. Rev. math. Union interbalcanique,Volume 1.

[25] Von Mises, R. (1964). Selected papers of Richard von Mises, American mathematical society,Volume 1.

Appendix A

Theorem 1.

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a given probability space and let $H = [a, b]$ be an interval for some $d < b$ $(a = -\infty,\ b = \infty$ might as well be allowed) . Let : $\Omega \rightarrow \mathrm{H}$ 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

$$E = [q_2(X)|X \geq x] = E[q_1(X)|X \geq x]\eta(x), x \in H,$$

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

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

where the function s is a solution of the differential equation $s' = \frac{\eta' q_1}{\eta q_1 - q_2}$ $and\ C$ 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 $\eta_n (n \in \mathbb{N})$ satisfy the conditions of Theorem 1 and let $q_{1n} \rightarrow q_1$, $q_{2n} \rightarrow q_2$ 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 η_n converges to , where

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

This stability theorem makes sure that the convergence of distribution functions is reflected by corresponding convergence of the functions q_1 , q_2 and η, respectively. It guarantees, for instance, the ’convergence’ of characterization of the Wald distribution to that of the Lévy-Smirnov distribution if $\alpha \rightarrow \infty$.

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, η should be as simple as possible. Since the function triplet is not uniquely determined it is often possible to choose $\xi\eta$ 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