STATISTICA, anno LXXXII, n. 3, 2022

A MUTH-PARETO DISTRIBUTION: PROPERTIES, ESTIMATION, CHARACTERIZATIONS AND APPLICATIONS

Musaddiq Sirajo
Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
Mohammad Shakil [1]
Department of Mathematics, Miami Dade College, Hialeah Campus, Hialeah, Fl., USA
Mohammad Ahsanullah
Emeritus Professor, Sciences Rider University, Lawrenceville, NJ, USA

SUMMARY

In this paper, a new distribution of the Muth-generated family is introduced by considering the Pareto model as baseline with the goal of having increased flexibility and improved goodness of fit in terms of studying tail characteristics. Maximum likelihood estimated parameters of the distribution are found to be consistent and asymptotically unbiased. From a practical point of view, it is shown that the proposed distribution is more flexible than some common statistical distributions. In particular, the proposed model proves to fit well into unimodal data structures. Some mathematical properties are derived, and characterization is investigated by a truncated first moment where a product of the reverse hazard rate and another function of the truncated point is considered. Other characterizations by order statistics and upper record values based on the characterization by the first truncated moment are also established.

Keywords: Characterizations; Estimation; Muth-Pareto distribution; Order statistics; Truncated moment; Upper record values.

1. INTRODUCTION

Numerous distributions have been established, but still there are continuous demands for developing new distributions that either add flexibility or are good for fitting particular real world or naturally occurring phenomena. This has always been the motivation for numerous researchers pursuing and generating different distributions of varying flexibility. Some of the most popular techniques of generating families of distributions include the Beta-Generated family by Eugene *et al.* (2002), Transformed-Transformer

[1] Corresponding Author. E-mail: mshakil@mdc.edu

$(T-X)$ proposed by Alzaatreh *et al.* (2013), the Weibull-G class defined by Bourguignon *et al.* (2014) among several others. Using the $T-X$ technique (by letting the Muth distribution be the Transformed), the Muth-Generated $(M-G)$ was proposed as a new family of distributions by Almarashi and Elgarhy (2018), wherein some sub-models were defined, namely, Muth-Uniform, Muth-Lomax, Muth-Rayleigh, Muth-Exponential and Muth-Weibull, but only the Muth-Weibull was studied comprehensively. The Muth-Pareto Distribution (MPD) was introduced as part of an MSc dissertation by Musaddiq Sirajo (2020) to be a new member of the $M-G$ family, with the goal of enhancing the use of the seemingly neglected Muth distribution (see e.g. Leemis and McQueston, 2008, for a discussion) by way of generalization. The Muth distribution was first introduced by Muth (1977), but comprehensive studies of its properties, an extension and important applications are reported in Jodrá *et al.* (2015) and Jodrá *et al.* (2017).

For an arbitrary baseline cumulative distribution function (cdf) $G(x)$ and probability density function (pdf) $g(x)$, Almarashi and Elgarhy (2018) defined the $M-G$ family with the respective pdf and cdf

$$f(x) = e^{\frac{1}{\alpha}} g(x)[1-\alpha G(x)^{\alpha}]G(x)^{-2\alpha-1}\exp(-\tfrac{1}{\alpha}G(x)^{-\alpha}), \quad x \in R, \tag{1}$$

$$F(x) = e^{\frac{1}{\alpha}} G(x)^{-\alpha}\exp(-\tfrac{1}{\alpha}G(x)^{-\alpha}), \quad x \in R, \tag{2}$$

where $\alpha \in (0,1]$.

The choice of the Pareto distribution as a baseline in this paper is due to its flexibility and wide applicability in various areas of human endeavor, including the context of reliability theory wherein the Muth distribution is originally proposed. The pdf of the Pareto distribution is given (for $\lambda \leq x < \infty$ and $\lambda, \theta > 0$) by

$$g(x) = \frac{\theta\lambda^{\theta}}{x^{\theta+1}}, \tag{3}$$

where λ is a location parameter and θ is the shape parameter.

The corresponding cdf is given by

$$G(x) = 1 - \left(\frac{\lambda}{x}\right)^{\theta}. \tag{4}$$

It is hoped that by combining the Muth and the Pareto distributions into one, it could also lead to increased flexibility and improved goodness of fit in terms of studying tail characteristics.

The remainder of this paper is organized as follows. In Section 2, the MPD is defined, and some of its general mathematical properties are established, including parameter estimation using the method of maximum likelihood. Section 3 concerns some characterizations of the MPD based on a truncated first moment, order statistics and upper record values. In Section 4, a Monte Carlo simulation study was carried out to assess the maximum likelihood estimated parameters. An illustration on the basis of real data sets is provided in Section 5. Finally, Section 6 summarizes the results and concludes the paper.

2. The Muth-Pareto distribution

By substituting Eq.(3) and Eq.(4) into Eq.(1), the pdf of the MPD model is given (for $x \geq \lambda$) by

$$f(x) = \frac{\theta\lambda^{\theta}}{x^{\theta+1}} \left\{1-\alpha\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{\alpha}\right\}\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-2\alpha-1} \times \exp\left\{\frac{1}{\alpha}\left(1-\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right)\right\}, \quad (5)$$

with corresponding cdf

$$F(x) = \left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha} \exp\left\{\frac{1}{\alpha}\left(1-\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right)\right\}, \quad (6)$$

where $\lambda > 0$ is a threshold parameter determining the location of the MPD random variable, and $\alpha \in (0,1]$, $\theta > 0$, are positive shape parameters demonstrating the diverse shapes of the MPD.

The reliability function $R(x)$ and failure rate function $h(x)$ of MPD, respectively, are given by

$$R(x) = 1-\left(\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha} \exp\left\{\frac{1}{\alpha}\left(1-\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right)\right\}\right), \quad (7)$$

$$h(x) = \frac{\theta\lambda^{\theta}\left\{1-\alpha\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{\alpha}\right\}\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-2\alpha-1}\exp\left\{\frac{1}{\alpha}\left(1-\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right)\right\}}{x^{\theta+1}\left(1-\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\exp\left\{\frac{1}{\alpha}\left(1-\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right)\right\}\right)}. \quad (8)$$

Figures 1 and 2 show some plots of the MPD density function, distribution function, reliability function and failure rate function for different values of the parameters α, θ and λ. The MPD is unimodal, heavy-tailed and positively skewed, with a failure rate function that has a non-monotone unimodal shape. Therefore, if the empirical study suggests a non-monotone failure rate function which has a unimodal shape, then the MPD may be adopted for the analysis of such data sets.

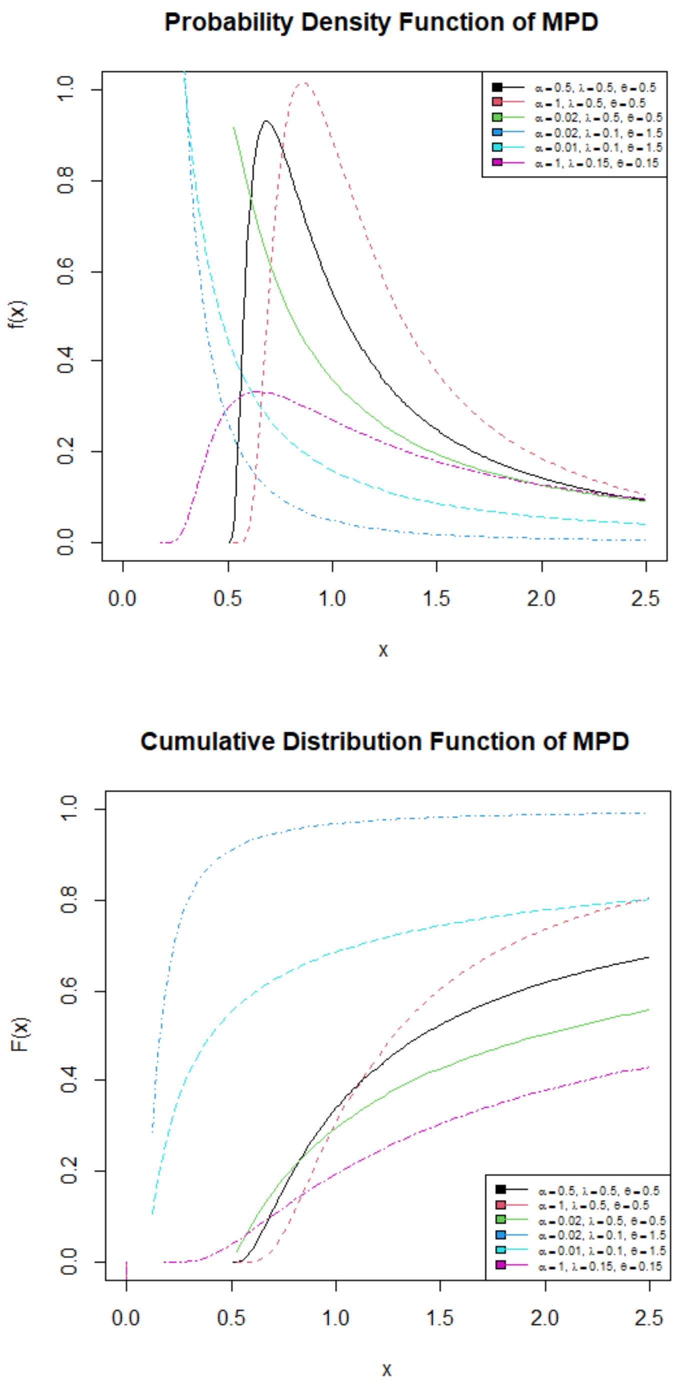

Figure 1 – The graphs of the MPD density and distribution functions for some parameter values.

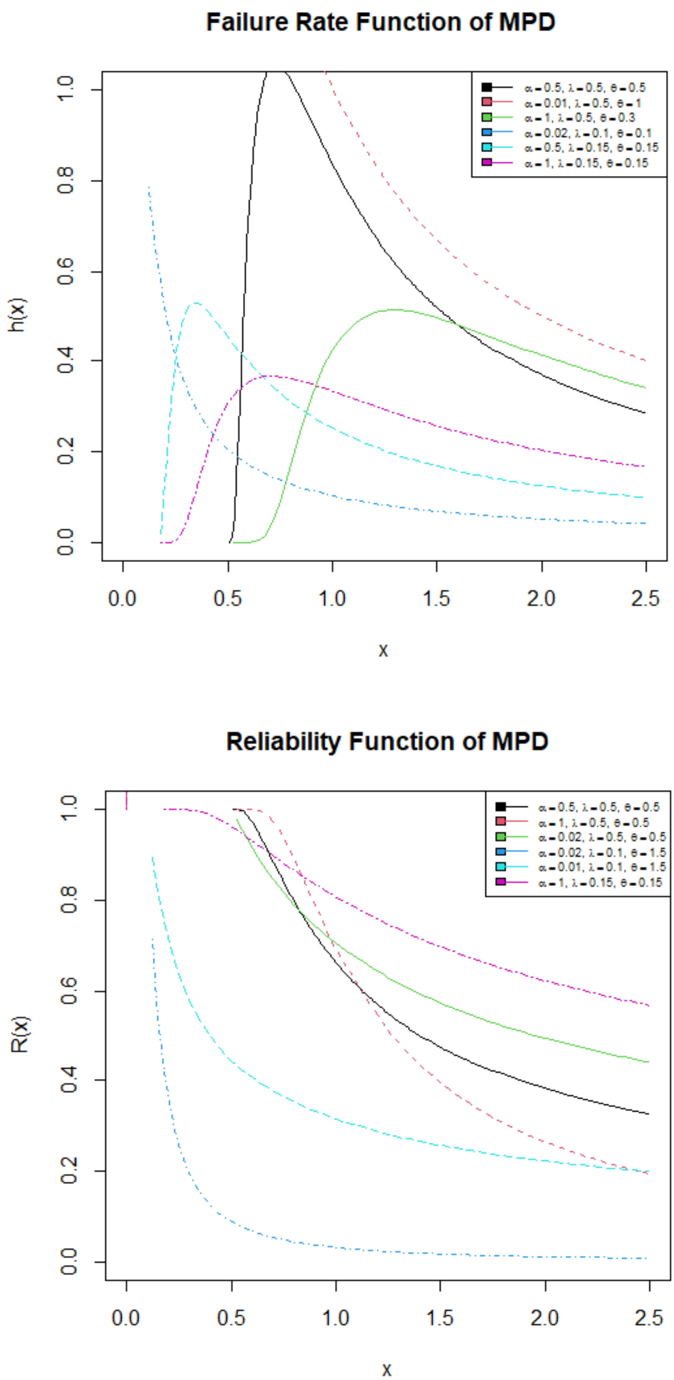

Figure 2 – The graphs of the MPD failure rate and relability functions for some parameter values.

2.1. Mixture representation

The MPD density function given in Eq.(5) can be expressed using power series as

$$f(x)=\sum_{i=0}^{\infty}\frac{e^{\frac{1}{\alpha}}(-1)^i}{\alpha^i i!}\frac{\theta\lambda^\theta}{x^{\theta+1}}\left\{\left[1-\left(\frac{\lambda}{x}\right)^\theta\right]^{-\alpha(i+2)-1}-\alpha\left[1-\left(\frac{\lambda}{x}\right)^\theta\right]^{-\alpha(i+1)-1}\right\}.$$

Using the generalized binomial series and after some algebra, the MPD density function can be expressed as

$$f(x)=\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{e^{\frac{1}{\alpha}}(-1)^{i+k}}{\alpha^i i!}\binom{j}{k}\left\{\binom{\alpha(i+2)+j}{j}-\alpha\binom{\alpha(i+1)+j}{j}\right\}$$
$$\times\frac{\theta\lambda^\theta}{x^{\theta+1}}\left[1-\left(\frac{\lambda}{x}\right)^\theta\right]^k,$$

or equivalently

$$f(x)=\sum_{k=0}^{\infty}\pi_k h_{\lambda,\theta,k}(x), \tag{9}$$

where

$$\pi_k=\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{e^{\frac{1}{\alpha}}(-1)^{i+k}}{\alpha^i i!}\binom{j}{k}\left\{\binom{\alpha(i+2)+j}{j}-\alpha\binom{\alpha(i+1)+j}{j}\right\}, \tag{10}$$

and

$$h_{\lambda,\theta,k}(x)=\frac{\theta\lambda^\theta}{x^{\theta+1}}\left[1-\left(\frac{\lambda}{x}\right)^\theta\right]^k.$$

Here, $h_{\lambda,\theta,k}(x)$ is the pdf of the exponentiated Pareto distribution (EPD) with parameters λ, θ and k. This means that the MPD density can be expressed as a linear combination of EPD densities. Consequently, the properties of MPD can be derived as linear combinations of those of the EPD. The cdf of MPD can also be obtained in similar fashion as

$$F(x)=\sum_{l=0}^{\infty}\pi_l h_{\lambda,\theta,l}(x), \tag{11}$$

where

$$\pi_l=\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{e^{\frac{1}{\alpha}}(-1)^{i+l}}{\alpha^i i!}\binom{j}{l}\binom{\alpha(i+1)+j-1}{j}. \tag{12}$$

2.2. Moments and moment generating function

The Pearson's rth order moment, denoted by $E(X^r)=\mu_r'$, of the MPD is defined as

$$E(X^r)=\mu_r'=\sum_{k=0}^{\infty}\pi_k\int_{\lambda}^{\infty}x^r h_{\lambda,\theta,k}(x)dx,$$

$$\mu_r'=\lambda^r\sum_{k=0}^{\infty}\pi_k B\left(\frac{\theta-r}{\theta},(k+1)\right), \tag{13}$$

where $B(p,q)$ is the well-known Beta function.

Specifically, for $r=1$, in Eq.(13) above, the mean of X is

$$\mu_1'=\lambda\sum_{k=0}^{\infty}\pi_k B\left(\frac{\theta-1}{\theta},(k+1)\right),$$

and the second moment is

$$\mu_2'=\lambda^2\sum_{k=0}^{\infty}\pi_k B\left(\frac{\theta-2}{\theta},(k+1)\right).$$

The coefficient of variation of X is obtained as

$$CV(X)=\frac{\left\{\sum\limits_{k=0}^{\infty}\pi_k B\left(\frac{\theta-2}{\theta},(k+1)\right)-\left[\sum\limits_{k=0}^{\infty}\pi_k B\left(\frac{\theta-1}{\theta},(k+1)\right)\right]^2\right\}^{\frac{1}{2}}}{\sum\limits_{k=0}^{\infty}\pi_k B\left(\frac{\theta-1}{\theta},(k+1)\right)}. \tag{14}$$

From Eq.(14) above, it can be seen that the coefficient of variation is free of the location parameter λ. Thus, the coefficient of variation of MPD random variable is expressed in terms of the shape parameters only.

The variance, measures of kurtosis and skewness may be calculated from the Pearson's rth moments by means of the existing familiar relations.

The moment generating function (mgf) of a random variable X is

$$M_x(t)=E(e^{tx})=\int_{-\infty}^{\infty}e^{tx}f(x)dx,$$

since

$$e^{tx}=1+tx+\frac{(tx)^2}{2!}+\frac{(tx)^3}{3!}+\ldots=\sum_{r=0}^{\infty}\left(\frac{t^r x^r}{r!}\right),$$

$$M_x(t)=\sum_{r=0}^{\infty}\frac{t^r}{r!}\int_{-\infty}^{\infty}x^r f(x)dx=\sum_{r=0}^{\infty}\frac{t^r}{r!}\mu'_r. \tag{15}$$

The mgf of MPD (for $x\geq\lambda$) follows from the definition in Equations (13) and (15):

$$M_x(t)=\sum_{r=0}^{\infty}\frac{t^r}{r!}\lambda^r\sum_{k=0}^{\infty}\pi_k B\left(\frac{\theta-r}{\theta},(k+1)\right).$$

2.3. Incomplete moments

The sth incomplete moment, say $I_s(t)$, of X can be expressed as

$$I_s(t)=\int_{-\infty}^{t}x^s f(x)dx.$$

After some algebra and using the lower incomplete Beta function and Eq.(9) in the expression above (for $s\leq\theta$), the sth incomplete moment is obtained as

$$I_s(t)=\lambda^s\sum_{k=0}^{\infty}\pi_k B_t\left(1-\frac{s}{\theta},(k+1)\right), \tag{16}$$

where $B_z(p,q)=\int_0^z y^{p-1}(1-y)^{q-1}dy$ is the incomplete Beta function. Thus, when $s=1$, the 1st incomplete moment of X is given by

$$I_1(t)=\lambda\sum_{k=0}^{\infty}\pi_k B_t\left(1-\frac{1}{\theta},(k+1)\right).$$

2.4. Quantile function

The MPD exhibits a variate generation property, as its quantile function can be obtained in a closed form in terms of the Lambert W-function. This function has two branches, namely the principal branch denoted by W_0 and the negative branch W_{-1} (Figure 3). For details on the Lambert W-function, see Corless *et al.* (1996).

THEOREM 1. *For any fixed $\lambda,\theta>0$ and $0<\alpha\leq 1$, the quantile function of the MPD random variable is*

$$Q(u)=\frac{\lambda}{\left[1-\left\{-\alpha W_{-1}\left(-\frac{u}{\alpha\exp(1/\alpha)}\right)\right\}^{-1/\alpha}\right]^{1/\theta}},\quad 0<u<1$$

where W_{-1} represents the negative branch of the Lambert W-function.

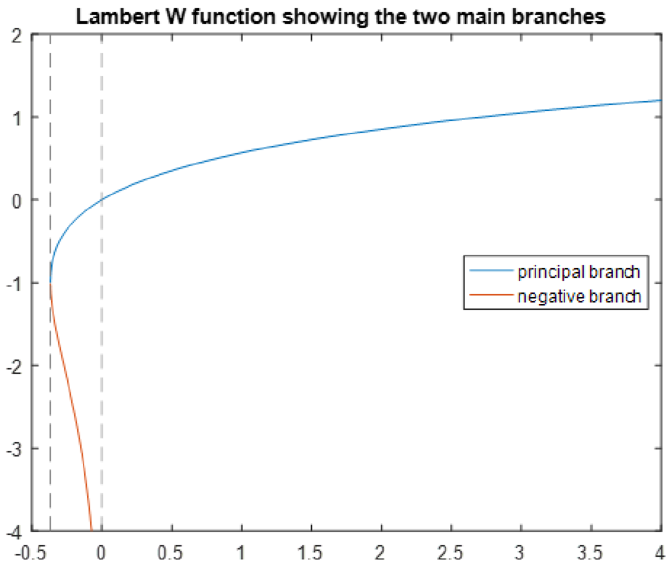

Figure 3 – Two main branches of the Lambert W-function

PROOF. Consider any selected $\lambda, \theta > 0$ and $0 < \alpha \leq 1$, as well as $u \in (0,1)$. The problem involves solving for x in $F(x) = u$ (taking $x > \lambda$) as in the following:

$$F(x) = \left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha} \exp\left\{\frac{1}{\alpha}\left(1 - \left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right)\right\} = u, \tag{17}$$

then multiply both sides of Eq.(17) by $-\frac{1}{\alpha}\exp\left(-\frac{1}{\alpha}\right)$ to obtain

$$-\frac{1}{\alpha}\left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha} \exp\left\{-\frac{1}{\alpha}\left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right\} = -\frac{u}{\alpha}\exp\left(-\frac{1}{\alpha}\right),$$

then $-\frac{1}{\alpha}\left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}$ is the W-function of the real argument $-\frac{u}{\alpha}\exp\left(-\frac{1}{\alpha}\right)$, so that

$$W\left[-\frac{u}{\alpha \exp(\frac{1}{\alpha})}\right] = -\frac{1}{\alpha}\left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}. \tag{18}$$

Again, for any fixed $\lambda, \theta > 0$, $0 < \alpha \leq 1$, $u \in (0,1)$ and $x > \lambda$, it is clear that

$$-\frac{1}{\alpha}\left[1 - \left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha} \leq -1,$$

satisfying the condition of the negative branch of the W-function. It can as well be further checked that

$$-\frac{1}{\alpha}\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\exp\left\{-\frac{1}{\alpha}\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha}\right\}=-\frac{u}{\alpha}\exp\left(-\frac{1}{\alpha}\right)\in\left(-\frac{1}{e},0\right).$$

Therefore, by taking into consideration the properties of the negative branch of the Lambert W-function, Eq.(18) becomes

$$W_{-1}\left[-\frac{u}{\alpha\exp(\frac{1}{\alpha})}\right]=-\frac{1}{\alpha}\left[1-\left(\frac{\lambda}{x}\right)^{\theta}\right]^{-\alpha},$$

which by further simplification leads to the result. □

2.5. Order statistics

Let $x_1, x_2, ..., x_n$ be a random sample that comes from a continuous independent and identical distribution with cdf $F(x)$, and $X_{1:n} < X_{2:n} < ... < X_{n:n}$ to be the analogous order statistics. The density of the ith order statistic of the Muth-Generated family is defined by Almarashi and Elgarhy (2018) as

$$f_{i:n}(x)=\frac{f(x)}{B(i,n-i+1)}\sum_{v=0}^{n-i}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\pi_k \mathrm{P}_{l,v}F(x)^{k+l},$$

where

$$\mathrm{P}_{l,v}=(-1)^{v}\binom{n-i}{v}\pi_l.$$

Following this definition, the density of the ith order statistic of the MPD is

$$f_{i:n}(x)=\frac{1}{B(i,n-i+1)}\sum_{v=0}^{n-i}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\pi_k \mathrm{P}_{l,v}h_{\lambda,\theta,(k+l)}(x),$$

and π_k as well as π_l are as defined in Eq.(10) and Eq.(12) respectively. Consequently, the rth moment of ith order statistic of the MPD is defined as

$$E(X_{i:n}^{r})=\frac{\lambda^{r}}{B(i,n-i+1)}\sum_{v=0}^{n-i}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\pi_k \mathrm{P}_{l,v}B\left(1-\frac{r}{\theta},(k+l+1)\right).$$

2.6. *Parameter estimation*

The technique of maximum likelihood estimation (MLE) is employed for the estimation of the MPD parameters. If $x_1, x_2, ..., x_n$ is a random sample of size n from the Muth-Generated family with $\boldsymbol{\Phi} = (\alpha, \varphi)^T$ as vector of parameters, then the log-likelihood function may be stated as

$$l = \frac{n}{\alpha} + \sum_{i=1}^{n} \ln[g(x_i)] - (2\alpha+1)\sum_{i=1}^{n} \ln[G(x_i)] + \sum_{i=1}^{n} \ln[1-\alpha G(x_i)^{\alpha}] - \frac{1}{\alpha}\sum_{i=1}^{n}[G(x_i)^{-\alpha}].$$

It follows from above that the log-likelihood function of the MPD with parameter vector $\mathrm{T} = (\alpha, \lambda, \theta)^T$ is

$$l = \frac{n}{\alpha} + \sum_{i=1}^{n} \ln\theta - \sum_{i=1}^{n} \ln x_i + 3\alpha\theta \sum_{i=1}^{n} \ln\left(\frac{\lambda}{x_i}\right) + 2\theta \sum_{i=1}^{n} \ln\left(\frac{\lambda}{x_i}\right) - \sum_{i=1}^{n} \ln\alpha - \frac{1}{\alpha}\sum\left[1-\left(\frac{\lambda}{x_i}\right)^{\theta}\right]^{-\alpha}. \quad (19)$$

Assuming λ to be known (since $x > \lambda$), by setting $\hat{\lambda} = X_{1:n}$, the lowest order statistic of the sample, the parameter vector is reduced to

$$\mathrm{T} = (\alpha, \theta)^T = \left(\frac{\partial l}{\partial \alpha}, \frac{\partial l}{\partial \theta}\right).$$

The elements of this vector are then derived as

$$\frac{\partial l}{\partial \alpha} = -\frac{n}{\alpha^2} + 3n\theta \ln\lambda - 3\theta \sum_{i=1}^{n} \ln x_i - \frac{n}{\alpha} + \sum_{i=1}^{n} \frac{\ln p_i}{\alpha p_i^{\alpha}} + \sum_{i=1}^{n} \frac{1}{\alpha^2 p_i^{\alpha}}, \quad (20)$$

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta} + (3\alpha+2)n\ln\lambda - (3\alpha-2)\sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} \frac{z_i}{p_i^{\alpha+1}}, \quad (21)$$

where

$$p_i = 1 - \left(\frac{\lambda}{x_i}\right),$$

and

$$z_i = \left(\frac{\lambda}{x_i}\right)^{\theta} \ln\left(\frac{\lambda}{x_i}\right).$$

The estimates of α and θ may be computed iteratively by equating these elements to zero and solving for them numerically using standard techniques like the Newton-Raphson (N-R) algorithm. Alternatively, any two-dimensional optimization technique may be used to maximize l directly, and being a simple two-dimensional optimization problem, obtaining initial guesses is also not difficult.

the MPD is a regular family, therefore, the following asymptotic result holds: as $n \to \infty$, $\sqrt{n\left(\hat{\alpha}-\alpha,\, \hat{\theta}-\theta\right)}$ converges to a bivariate normal distribution with mean vector $\mathbf{0}$ and the variance-covariance matrix $\mathbf{I}_2^{-1}$, where $\mathbf{I}_2 = \left(I_{ij}\right)$ is the Fisher information matrix. The elements of $\mathbf{I}_2 = \left(I_{ij}\right)$ are given by

$$I_{11} = \frac{\partial^2 l}{\partial \alpha^2} = \frac{2n}{\alpha^3} + \frac{n}{\alpha^2} - \sum_{i=1}^{n} \frac{(\ln p_i)(1+\ln p_i{}^{\alpha})\, p_i^{\alpha}}{\left(\alpha p_i^{\alpha}\right)^2} - \alpha \sum_{i=1}^{n} \frac{\left(\ln p_i^{\alpha}+2\right) p_i^{\alpha}}{\left(\alpha^2 p_i^{\alpha}\right)^2},$$

$$I_{12} = \frac{\partial^2 l}{\partial \alpha \partial \theta} = 3\left(\ln \lambda^n - \sum_{i=1}^{n} \ln x_i\right) + \sum_{i=1}^{n} \frac{z_i (\ln p_i)\, p_i^{\alpha}}{p_i^{2\alpha+1}},$$

$$I_{22} = \frac{\partial^2 l}{\partial \theta^2} = -\frac{n}{\theta^2} - \sum_{i=1}^{n} \frac{z_i \ln(\lambda/x_i)}{p_i^{\alpha+1}}.$$

Using the above result, the asymptotic variance of the unknown parameters can be easily obtained and confidence intervals constructed.

3. Some characterizations

Various methods of characterization of a probability distribution have been proposed by different researchers in recent years. For a comprehensive account of these techniques, the interested readers are referred to Nagaraja (2006), Ahsanullah *et al.* (2014), Ahsanullah (2017), among others. A characterization of a particular probability distribution states that it is the only distribution that satisfies some specified conditions. In recent years, there has been a great interest in the characterizations of probability distributions by truncated moments. For example, the development of the general theory of the characterizations of probability distributions by truncated moment began with the work of Galambos and Kotz (2006). Other notable contributions are Kotz and Shanbhag (1980), Glänzel *et al.* (1984), and Glänzel (1987). Most of these characterizations are based on a simple relationship between two different moments truncated from the left at the same point; these may serve as the basis for goodness-of-fit tests, or testing the efficiency of a particular test of hypothesis and the power of a particular parameter estimation technique (Glänzel, 1987; Glanzel, 1990; Volkova and Nikitin, 2015). For an excellent survey of goodness-of-fit and symmetry tests based on the characterization properties of distributions, the interested readers are referred to two recent nice papers of Nikitin (2017) and Miloševic (2017), respectively, and references therein.

3.1. *Characterization by truncated first moment*

In this section, a characterization of the MPD by truncated first moment is presented by considering a product of the reverse hazard rate and another function of the truncated point. Lemmas 2 and 3 are proved that will be useful for proving the main characterization results. For details, readers are referred to Ahsanullah *et al.* (2015), or Ahsanullah (2017), or Shakil *et al.* (2018). The main characterization results are provided in Theorems 4 and 5.

LEMMA 2. *Suppose that X is an absolutely continuous (with respect to Lebesgue measure) random variable with cdf $F(x)$ and pdf $f(x)$. Let $F(-\infty) = 0$ and $F(x) > 0$ for all real $x > 0$. It is assumed that $E(X)$ exists and that $f'(x)$ exist for all x, $-\infty < x < \infty$. Then if $E(X|X \leq x) = g(x)\eta(x)$, where $g(x)$ is a differentiable function of x, $-\infty < x < \infty$, and $\eta(x) = \frac{f(x)}{F(x)}$, then $f(x) = ce^{\int \frac{X-g'(x)}{g(x)}dx}$, where c is determined by the condition that $\frac{1}{c} = \int_{-\infty}^{\infty} f(x)dx$.*

PROOF.

$$\int_{-\infty}^{x} uf(u)du = g(x)f(x),$$

which after differentiating with respect to x gives

$$xf(x) = g'(x)f(x) + g(x)f'(x),$$

and on simplification becomes

$$\frac{f'(x)}{f(x)} = \frac{x - g\prime(x)}{g(x)}.$$

On integrating the above Equation, the result follows. □

LEMMA 3. *Suppose that X is an absolutely continuous (with respect to Lebesgue measure) random variable with cdf $F(x)$ and pdf $f(x)$. Let $F(-\infty) = 0$ and $F(x) > 0$ for all real $x > 0$. It is assumed that $E(x)$ exists and that f' exist for all $x, -\infty < x < \infty$. Then if $E(X \mid X \geq x) = h(x)r(x)$, where $h(x)$ is a continuous differentiable function of x, $-\infty < x < \infty$, and $r(x) = \frac{f(x)}{1-F(x)}$, then $f(x) = ce^{\int -\frac{x+h'(x)}{h(x)}dx}$, where c is a constant determined by the condition $\frac{1}{c} = \int_{-\infty}^{\infty} f(x)dx$.*

PROOF. The proof is similar to Lemma 2. □

THEOREM 4. *Suppose that the random variable X has an absolutely continuous (with respect to Lebesgue measure) cdf $F(x)$ and pdf $f(x)$. Let $F(0) = 0$, $F(x) > 0 \;\forall\; x > 0$. It is assumed that $f'(x)$ exists for all $x \in (0, \infty)$, and $E(X) < \infty$. Let $\eta(x) = \frac{f(x)}{F(x)}$, then X has the Muth-Pareto distribution with the probability density function defined in Eq. (5) if and only if $E(X|X \leq x) = g(x)\eta(x)$, where*

$$g(x) = \frac{\lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty} \pi_k \frac{\theta\lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^k}, \tag{22}$$

and π_k *is as given by Eq.(10),* $x \geq \lambda$, $\lambda, \alpha, \theta > 0$, *and* $B_z(p,q) = \int_0^z y^{p-1}(1-y)^{q-1} dy$ *denotes the incomplete Beta function.*

PROOF. Suppose that the random variable X has the Muth-Pareto distribution with the pdf given by Eq.(5), which may alternatively be written using Eq.(9) as

$$f(x) = \sum_{k=0}^{\infty} \pi_k \frac{\theta \lambda^{\theta}}{x^{\theta+1}}(1-(\tfrac{\lambda}{x})^{\theta})^k,$$

where π_k is as given by Eq.(10), $x \geq \lambda$ and $\lambda, \alpha, \theta > 0$.

Suppose that

$$E(X|X \leq x) = g(x)\eta(x) = g(x)\frac{f(x)}{F(x)},$$

since

$$E(X|X \leq x) = \frac{\int_\lambda^x u f(u) du}{F(x)},$$

then

$$g(x) f(x) = \int_\lambda^x u f(u) du$$

$$= \sum_{k=0}^{\infty} \pi_k \int_\lambda^x u \frac{\theta \lambda^{\theta}}{u^{\theta+1}}(1-(\tfrac{\lambda}{u})^{\theta})^k du$$

$$= \lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\tfrac{1}{\theta}, k+1),$$

where $B_z(p,q) = \int_0^z y^{p-1}(1-y)^{q-1} dy$ is the incomplete Beta function.

Thus,

$$g(x) = \frac{\lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty} \pi_k \frac{\theta \lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^k},$$

and

$$g'(x) = x - \frac{\lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty} \pi_k \frac{\theta \lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^k} m(x), \tag{23}$$

where

$$m(x) = -\tfrac{1}{x}(\alpha+1) + \tfrac{1}{x^2}\theta \alpha^2 \lambda \left(\tfrac{1}{x}\lambda\right)^{\theta-1} \frac{\left(1-\left(\frac{1}{x}\lambda\right)^{\theta}\right)^{\alpha-1}}{\alpha\left(1-\left(\frac{1}{x}\lambda\right)^{\theta}\right)^{\alpha}-1}$$

$$= -\frac{1}{x^2}\theta\lambda(2\alpha-1)\,\frac{\left(\frac{1}{x}\lambda\right)^{\theta-1}}{\left(\frac{1}{x}\lambda\right)^{\theta}-1} - \frac{1}{x^2}\theta\lambda\left(\frac{1}{x}\lambda\right)^{\theta-1}\left(1-\left(\frac{1}{x}\lambda\right)^{\theta}\right)^{\alpha-1}.$$

Simplifying Eq.(23) leads to

$$g'(x) = x - g(x)\, m(x),$$

or equivalently,

$$\frac{x-g'(x)}{g(x)} = m(x). \tag{24}$$

Applying Lemma 2 to Eq.(24) gives

$$\frac{f'(x)}{f(x)} = m(x),$$

which on integrating with respect to x becomes

$$f(x) = \frac{c}{x^{\alpha+1}}(1-\alpha(1-(\tfrac{\lambda}{x})^{\theta})^{\alpha})(1-(\tfrac{\lambda}{x})^{\theta})^{2\alpha-1}\exp(\tfrac{1}{\alpha}(1-(1-(\tfrac{\lambda}{x})^{\theta})^{-\alpha}),$$

where

$$\frac{1}{c} = \int_{-\infty}^{\infty} f(x)dx = \frac{1}{\theta\lambda^{\theta}}.$$

This completes the proof of the "only if" condition of Theorem 4.

□

THEOREM 5. *Suppose that the random variable X has an absolutely continuous (with respect to Lebesgue measure) cdf $F(x)$ and pdf $f(x)$. Let $F(0)=0$, $F(x)>0\ \forall\ x>0$. It is assumed that $f'(x)$ exists for all $x\in(0,\infty)$, and $E(X)<\infty$. Then $E\left(X\,|X\geq x\right) = h(x)\,r(x)$, where $r(x) = \frac{f(x)}{1-F(x)}$ and*

$$h(x) = \frac{E(X) - \lambda\sum_{k=0}^{\infty}\pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty}\pi_k\,\frac{\theta\lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^{k}}$$

if and only if the pdf $f(x)$ of X is as given in Eq.(5).

PROOF. Given

$$f(x) = \sum_{k=0}^{\infty}\pi_k\,\frac{\theta\lambda^{\theta}}{x^{\theta+1}}(1-(\tfrac{\lambda}{x})^{\theta})^{k},$$

suppose that

$$E\left(X\,|X \geq x\right) = h(x)\,r(x) = \frac{h(x)f(x)}{1-F(x)},$$

since

$$E\left(X\,|X \geq x\right) = \frac{\sum_{x}^{\infty} u\,f(u)\,du}{1-F\left(x\right)},$$

then

$$h(x)f(x) = \int_{x}^{\infty} u\,f(u)\,du,$$

or equivalently,

$$h(x)f(x) = \int_{\lambda}^{\infty} u\,f(u)\,du - \int_{\lambda}^{x} u\,f(u)\,du,$$

which on applying Eq.(13) gives

$$h(x)f(x) = E\left(X\right) - \lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\tfrac{1}{\theta}, k+1).$$

Dividing both sides by $f(x)$ defined in Eq.(22) leads to

$$h\left(x\right) = \frac{E\left(X\right) - \lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty} \pi_k \frac{\theta\lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^{k}}.$$

Then,

$$h'\left(x\right) = -x - \frac{E\left(X\right) - \lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty} \pi_k \frac{\theta\lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^{k}}\, m(x), \tag{25}$$

where

$$m\left(x\right) = -\frac{1}{x}\left(\alpha+1\right) + \frac{1}{x^2}\theta\alpha^2\lambda\left(\frac{1}{x}\lambda\right)^{\theta-1} \frac{\left(1-\left(\frac{1}{x}\lambda\right)^{\theta}\right)^{\alpha-1}}{\alpha\left(1-\left(\frac{1}{x}\lambda\right)^{\theta}\right)^{\alpha}-1}$$

$$= -\frac{1}{x^2}\theta\lambda(2\alpha-1)\frac{\left(\frac{1}{x}\lambda\right)^{\theta-1}}{\left(\frac{1}{x}\lambda\right)^{\theta}-1} - \frac{1}{x^2}\theta\lambda\left(\frac{1}{x}\lambda\right)^{\theta-1}\left(1-\left(\frac{1}{x}\lambda\right)^{\theta}\right)^{\alpha-1}.$$

It is obvious from Eq.(25) that

$$h'(x) = -x - h\left(x\right) m\left(x\right),$$

so that

$$\frac{-x - h'(x)}{h(x)} = m(x). \tag{26}$$

Applying Lemma 3 to Eq.(26) gives

$$\frac{f'(x)}{f(x)} = m(x),$$

which on integration with respect to x gives

$$f(x) = \frac{c}{x^{\alpha+1}}(1-\alpha(1-(\tfrac{\lambda}{x})^{\theta})^{\alpha})(1-(\tfrac{\lambda}{x})^{\theta})^{2\alpha-1}\exp(\tfrac{1}{\alpha}(1-(1-(\tfrac{\lambda}{x})^{\theta})^{-\alpha}),$$

where

$$\frac{1}{c} = \int_{-\infty}^{\infty} f(x)dx = \frac{1}{\theta\lambda^{\theta}}.$$

This completes the proof of the "if" condition of Theorem 5.

□

3.2. *Characterization by order statistics*

Let $X_1, X_2, ..., X_n$ be n independent copies of the random variable X with absolutely continuous distribution function $F(x)$ and pdf $f(x)$, and let $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ be the corresponding order statistics. Then, $X_{j,n}|X_{k,n} = x$, for $1 \leq k < j \leq n$, is distributed as the $(j-k)$th order statistics from $(n-k)$ independent observations from the random variable V having the pdf $f_V(v|x)$, where $f_V(v|x) = \frac{f(v)}{1-F(x)}$, $0 \leq v < x$, and $X_{i,n}|X_{k,n} = x$, $1 \leq i < k \leq n$ is distributed as the ith order statistics from k independent observations from the random variable W having the pdf $f_W(w|x)$, where $f_W(w|x) = \frac{f(w)}{F(x)}$, $w < x$; see Ahsanullah *et al.* (2013) or Arnold *et al.* (2005) for details.

THEOREM 6. *Suppose that the random variable X has an absolutely continuous (with respect to Lebesgue measure) cdf $F(x)$ and pdf $f(x)$. It is assumed that $F(0) = 0$, $F(x) > 0$, $\forall\ x > 0$, $f'(x)$ exists for all $x \in (0, \infty)$, and $E(X) < \infty$. Let*

$$S_{k-1} = \frac{1}{k-1}\left(X_{1,n} + X_{2,n} + ... + X_{k-1,n}\right),$$

then $E(S_{k-1}|X_{k,n} = x) = g(x)\tau(x)$, where $\tau(x) = \frac{f(x)}{F(x)}$ and $g(x)$ as defined in Eq.(22), if and only if X has the Muth-Pareto distribution with the pdf defined by Eq.(5).

PROOF. It is known that

$$E(S_{k-1}|X_{k,n}=x)=E(X|X\leq x),$$

see Ahsanullah *et al.* (2013) and David and Nagaraja (2003). Hence, by Theorem 4, the result follows. □

THEOREM 7. *Suppose that the random variable X has an absolutely continuous (with respect to Lebesgue measure) cdf $F(x)$ and pdf $f(x)$. It is assumed that $F(0)=0$, $F(x)>0$, $\forall\ x>0$, $f'(x)$ exists for all $x\in(0,\infty)$, and $E(X)<\infty$. Let*

$$T_{k,n}=\frac{1}{n-k}\left(X_{k+1,n}+X_{k+2,n}+...+X_{n,n}\right),$$

then

$$E(T_{k,n}|X_{k,n}=x)=h(x)\,\frac{f(x)}{1-F(x)},$$

where

$$h(x)=\frac{E(X)-\lambda\sum_{k=0}^{\infty}\pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta},k+1)}{\sum_{k=0}^{\infty}\pi_k\frac{\theta\lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^k},$$

if and only if X has the Muth-Pareto distribution with the pdf defined by Eq.(5).

PROOF. Since $E(T_{k,n}|X_{k,n}=x)=E(X|X\geq x)$ (Ahsanullah *et al.*, 2013; David and Nagaraja, 2003), the result follows from Theorem 5. □

3.3. *Characterization by upper record values*

Let $X_1,X_2,...$ be a sequence of independent and identically distributed absolutely continuous random variables with distribution function $F(x)$ and pdf $f(x)$. If $Y_n=\max(X_1,X_2,...,X_n)$ for $n\geq 1$ and $Y_j>Y_{j-1}$, $j>1$, then X_j is called an upper record value of $\{X_n,\ n\geq 1\}$. The indices at which the upper records occur are given by the record times $\left\{U(n)>\min\left(j|j>U(n+1),X_j>X_{U(n-1)},\ n>1\right)\right\}$ and $U(1)=1$. For details on record values, see Ahsanullah (1995).

THEOREM 8. *Suppose that the random variable X has an absolutely continuous (with respect to Lebesgue measure) cdf $F(x)$ and pdf $f(x)$. It is assumed that $F(0)=0$, $F(x)>0$, $\forall\ x>0$, $f'(x)$ exists for all $x\in(0,\infty)$, and $E(X)<\infty$. Let the nth upper record value be denoted by $X(n)=X_{U(n)}$. Then*

$$E(X(n+1)|X(n)=x)=h(x)\,\frac{f(x)}{1-F(x)},$$

where

$$h(x) = \frac{E(X) - \lambda \sum_{k=0}^{\infty} \pi_k B_{1-(\frac{\lambda}{x})^{\theta}}(1-\frac{1}{\theta}, k+1)}{\sum_{k=0}^{\infty} \pi_k \frac{\theta \lambda^{\theta}}{x^{\theta+1}}(1-(\frac{\lambda}{x})^{\theta})^k},$$

if and only if X has the Muth-Pareto distribution with the pdf given by Eq.(5).

PROOF. It is known from Ahsanullah *et al.* (2013), and Nevzorov (2001) that $E(X(n+1)|X(n)=x) = E(X|X \geq x)$. Then, the result follows from Theorem 5. □

4. SIMULATION

Simulation results are presented to measure the performance of the proposed MLE procedure in estimating the parameters of the MPD.

4.1. *Simulated Numerical Procedure*

The problem is that of maximizing the log-likelihood function of the MPD so as to obtain MLEs of the parameters. This is an optimization problem that may be stated as

$$\begin{aligned} &\max \ l(\alpha, \lambda, \theta), \\ &\text{s.t. } 0 < \alpha \leq 1, \\ &\quad \lambda, \theta > 0, \end{aligned} \tag{27}$$

where $l(\alpha, \lambda, \theta)$ is the log-likelihood function defined in Eq.(19). The problem was solved using the default method of *optim* package of the R language, which is an implementation of Nelder and Mead (1965), which explicitly uses the log-likelihood function values and is robust. The method also works reasonably well for both differentiable and non-differentiable functions. The performance of the estimates is evaluated based on bias and root mean square error (RMSE) using the formulation demonstrated below for estimates of α_j, $j = 1, 2, ..., N$.

- Mean: $\bar{\alpha} = \frac{1}{N} \sum_{j=1}^{N} \hat{\alpha}_j$
- Variance: $\text{Var}(\hat{\alpha}) = \frac{1}{N} \sum_{j=1}^{N} (\hat{\alpha}_j - \bar{\alpha})^2$
- Bias: $\text{Bias}(\hat{\alpha}) = \frac{1}{N} \sum_{j=1}^{N} (\hat{\alpha}_j - \alpha) = (\bar{\alpha} - \alpha)$
- RMSE: $\text{RMSE}(\hat{\alpha}) = \sqrt{\text{Var}(\hat{\alpha}) + (\text{Bias}(\hat{\alpha}))^2}$

Equivalent quantities were also calculated for the parameters λ and θ. The simulations were repeated $N = 10,000$ times each with sample size $n = 1000, 500, 100, 25$ and parameter values $\mathbf{I}: \alpha = 1, \ \lambda = 0.15, \ \theta = 0.15$ and $\mathbf{II}: \alpha = 0.5, \ \lambda = 0.5, \ \theta = 0.5$.

4.2. Numerical Results of the Simulation

The results in Table 1 indicate that the maximum likelihood estimates are consistently stable and asymptotically unbiased, since the RMSE and bias decrease as the sample size increases.

5. Application of the Muth-Pareto distribution

Two real data sets previously analyzed by researchers in the literature are considered.

5.1. Application 1: Failure time dataset

The first real-life data set consists of the pooled data on the times of successive failures of the air conditioning system of each member of a fleet of Boeing 720 jet airplanes (Table 2). These data were studied by Proschan (1963), Tahir *et al.* (2018), among others.

To gain some idea about possible candidate distributions that can effectively fit these data, we used the *fitdistrplus* package of the R language (R Core Team, 2021), and obtained a Cullen and Frey (C&F) graph, which compares distributions in terms of kurtosis and squared skewness, and which may also help in selecting which distribution(s) to fit among potential candidates. The C&F plot is shown in Figure 4. The big blue dot represents the data, and it stays very far away from the points representing the normal, uniform and logistic distributions. The data also cannot follow a Beta distribution, since the values are outside the Beta range $(0,1)$. This same point also stays away from the line representing the lognormal distribution, but is a little closer to the line representing the Gamma distribution and the point representing the exponential distribution. Thus, the Gamma and exponential distributions are likely to fit these data well.

Proschan (1963) fitted an exponential distribution, whereas Tahir *et al.* (2018) fitted the Inverted Nadarajah-Haghighi (INH) distribution to these data. The MPD is hereby fitted, and the results are compared with those provided by the Gamma, exponential, INH, two-parameter Weibull (2P-Weibull) and Muth-Weibull (MW) distributions. The choice of MW is paramount since both the MPD and MW are sub-models of the $M-G$ family. The pdf associated with the competitive models are respectively given by:

$$f_G(x\,;\lambda,\theta)=\frac{\lambda^{\theta}x^{\theta-1}e^{-\lambda x}}{\Gamma(\theta)},\quad \text{for } x,\lambda,\theta>0, \tag{28}$$

TABLE 1
Simulated maximum likelihood estimates with associated variance, bias and RMSE.

α	λ	θ	n	Parameters	Mean	Variance	Bias	RMSE
				α	1.2348	0.3835	0.2348	0.6623
1	0.15	0.15	25	λ	0.1623	0.0324	0.0123	0.1804
				θ	0.1627	0.0065	0.0127	0.0815
				α	1.1245	0.0954	0.1245	0.3331
1	0.15	0.15	100	λ	0.1341	0.0065	-0.0159	0.0824
				θ	0.1446	0.0014	-0.0054	0.0381
				α	1.0217	0.0054	0.0217	0.0763
1	0.15	0.15	500	λ	0.1446	0.0007	-0.0054	0.0267
				θ	0.1477	0.0001	-0.0023	0.0122
				α	1.0092	0.0012	0.0092	0.0358
1	0.15	0.15	1000	λ	0.1476	0.0002	-0.0024	0.0144
				θ	0.1491	0.0000	-0.0009	0.0067
				α	0.2922	0.1773	-0.2078	0.4695
0.5	0.5	0.5	25	λ	0.5510	0.0094	0.0510	0.1097
				θ	0.6432	0.0439	0.1432	0.2538
				α	0.4043	0.0585	-0.0957	0.2602
0.5	0.5	0.5	100	λ	0.5183	0.0023	0.0183	0.0512
				θ	0.5514	0.0120	0.0514	0.1212
				α	0.4783	0.0094	-0.0217	0.0995
0.5	0.5	0.5	500	λ	0.5044	0.0004	0.0044	0.0199
				θ	0.5105	0.0016	0.0105	0.0416
				α	0.4887	0.0040	-0.0113	0.0643
0.5	0.5	0.5	1000	λ	0.5024	0.0002	0.0024	0.0131
				θ	0.5062	0.0008	0.0062	0.0287

TABLE 2
Data Set 1: Failure time data.

194	413	90	74	55	23	97	50	359	50	130	487	102
15	14	10	57	320	261	51	44	9	254	493	18	209
41	58	60	48	56	87	11	102	12	5	100	14	29
37	186	29	104	7	4	72	270	283	7	57	33	100
61	502	220	62	141	22	603	35	98	54	181	65	49
12	239	14	18	39	3	12	5	32	9	14	70	47
120	142	3	104	85	67	169	24	21	246	47	68	15
2	91	59	447	56	29	176	225	77	197	438	43	134
184	20	386	182	71	80	188	230	152	36	79	59	33
246	1	79	3	27	201	84	27	21	16	88	130	14
118	44	15	42	106	46	230	59	153	104	20	206	5
66	34	29	26	35	5	82	5	61	31	118	326	12
54	36	34	18	25	120	31	22	18	156	11	216	139
67	310	3	46	210	57	76	14	111	97	62	26	71
39	30	7	44	11	63	23	22	23	14	18	13	34
62	11	191	14	16	18	130	90	163	208	1	24	70
16	101	52	208	95								

$$f_E(x\,;\lambda) = \lambda e^{-\lambda x}, \quad \text{for } 0 \le x < \infty, \tag{29}$$

$$f_{\text{INH}}(x\,;\alpha,\lambda) = \alpha\lambda x^{-2}\left(1-\lambda x^{-1}\right)^{\alpha-1} \times \exp\left\{1-\left(1-\lambda x^{-1}\right)^{\alpha}\right\}, \quad \text{for } x,\alpha,\lambda > 0, \tag{30}$$

$$f_{2W}(x;\beta,\lambda) = \beta\lambda x^{\beta-1} e^{-\lambda x^{\beta}}, \quad \text{for } x,\beta,\lambda > 0, \tag{31}$$

$$f_{\text{MW}}(x\,;\alpha,\beta,\gamma) = \exp(1/\alpha)\gamma\beta x^{\gamma-1}\exp(-\beta x^{\gamma})\left(1-\alpha(1-\mathrm{e}^{-\beta x^{\gamma}})^{\alpha}\right) \times (1-\mathrm{e}^{-\beta x^{\gamma}})^{-2\alpha-1}\exp\left(-\tfrac{1}{\alpha}(1-\mathrm{e}^{-\beta x^{\gamma}})^{-\alpha}\right), \quad \text{for } x,\beta,\gamma > 0. \tag{32}$$

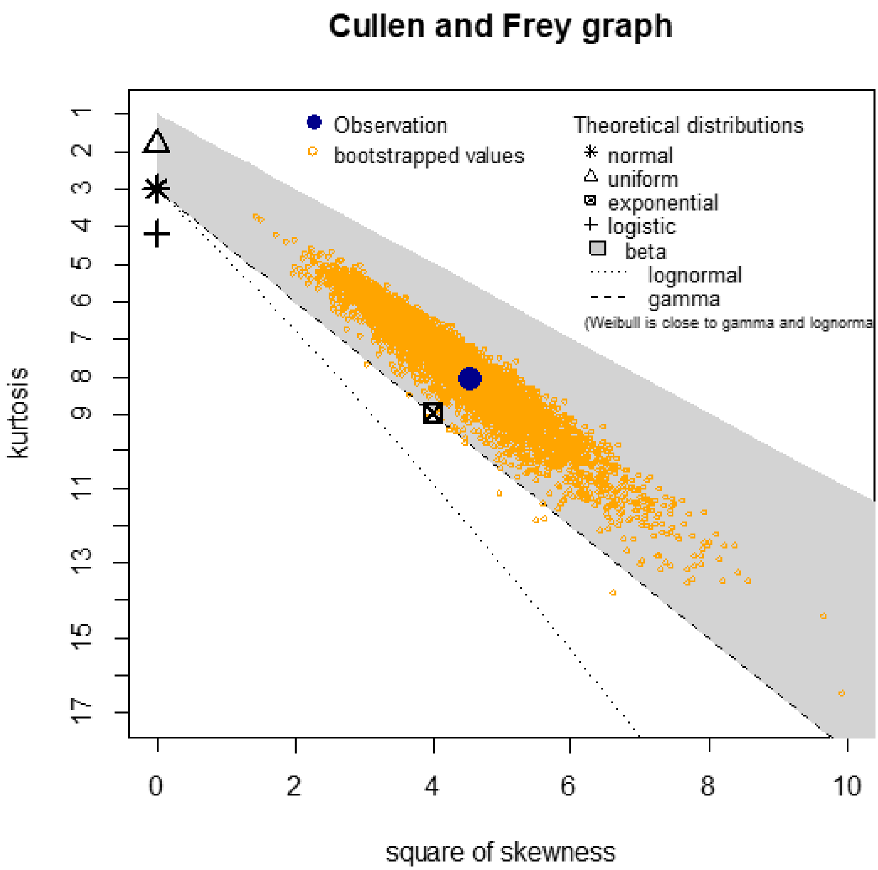

Figure 4 – C&F graph for Data Set 1.

The maximum likelihood estimated parameters of these distributions together with the resulting log-likelihood function value (`loglik`) are reported in Table 3 along with their associated AIC and BIC for model comparison purposes, where lower values indicate a better fit. The MPD demonstrates a better performance. Other goodness-of-fit statistics, namely, Kolmogorov-Smirnov (K-S), Cramer-Von Mises (CVM) and Anderson-Darling (A-D) were calculated and presented in Table 4. The density plots are shown in Figure 5.

TABLE 3
MLEs and goodness-of-fit criteria for Data Set 1.

Model	MLEs	loglik	AIC	BIC
MPD	$\hat{\alpha}=1.3\text{x}10^{-5}$, $\hat{\lambda}=0.2504$ $\hat{\theta}=3.0990$	**-102.77**	**211.54**	**212.54**
MW	$\hat{\alpha}=3.2\text{x}10^{-3}$, $\hat{\beta}=0.2458$ $\hat{\gamma}=0.5040$	-114.79	235.57	236.58
INH	$\hat{\lambda}=0.5001$, $\hat{\alpha}=40.5770$	-115.13	234.26	234.93
Exponential	$\hat{\lambda}=0.01073643$	-116.98	235.95	236.29
Gamma	$\hat{\lambda}=0.920976329$, $\hat{\theta}=0.009890729$	-116.03	236.06	236.73
2P-Weibull	$\hat{\lambda}=89.5326303$, $\hat{\beta}=0.9244685$	-115.96	235.92	236.59

TABLE 4
Goodness-of-fit statistics for Data Set 1.

Statistic	2P- Weibull	Gamma	Exponential	INH	MW	MPD
K-S	0.0519	0.0623	0.0726	0.0496	0.04873	0.0354
CVM	0.1270	0.1977	0.3241	0.3055	0.3106	0.1124
A-D	0.8232	1.1189	1.6919	0.9957	1.087	0.7341

5.2. *Application 2: Taxes revenue dataset*

The second data set (Table 5) corresponds to the data used by Nassar and Nada (2011) in their study of the Beta generalized Pareto (BGP) distribution as well as Mead (2014) in his study of the generalized Beta exponentiated Pareto (GBEP) distribution, representing "The monthly actual taxes revenue in Egypt from January 2006 to November 2010 (in 1000 millions of Egyptian pounds)".

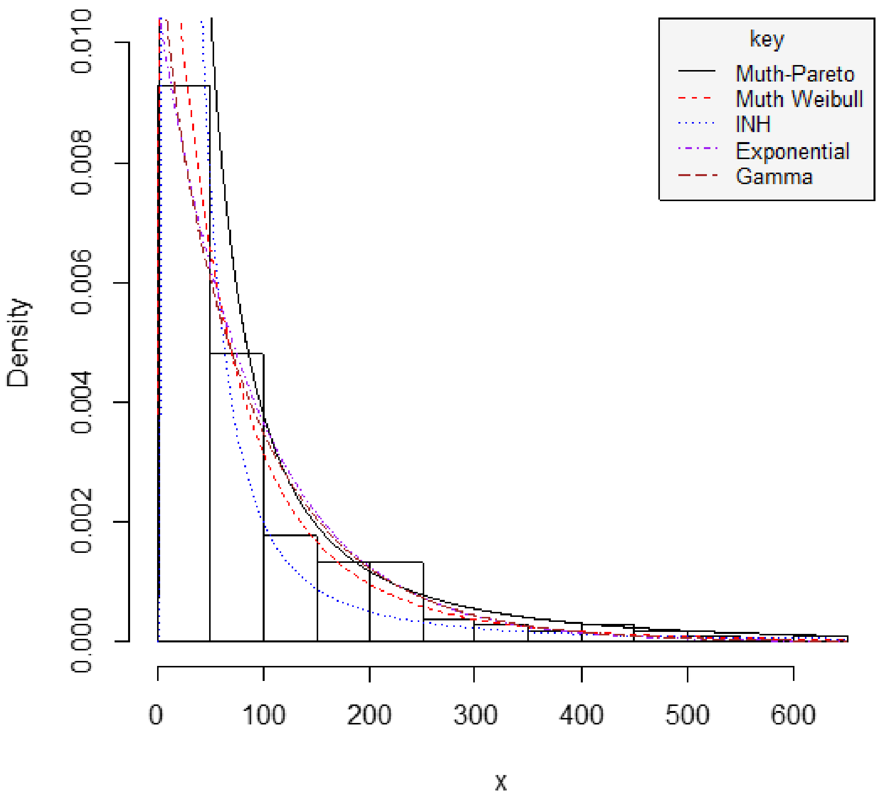

Figure 5 – Fitted densities for Data Set 1.

TABLE 5
Data Set 2: Egyptian taxes revenue data.

5.9	20.4	14.9	16.2	17.2	7.8	6.1	9.2	10.2	9.6	13.3	8.5	21.6
18.5	5.1	6.7	17.0	8.6	9.7	39.2	35.7	15.7	9.7	10.0	4.1	36.0
8.5	8.0	9.2	26.2	21.9	16.7	21.3	35.4	14.3	8.5	10.6	19.1	20.5
7.1	7.7	18.1	16.5	11.9	7.0	8.6	12.5	10.3	11.2	6.1	8.4	11.0
11.6	11.9	5.2	6.8	8.9	7.1	10.8						

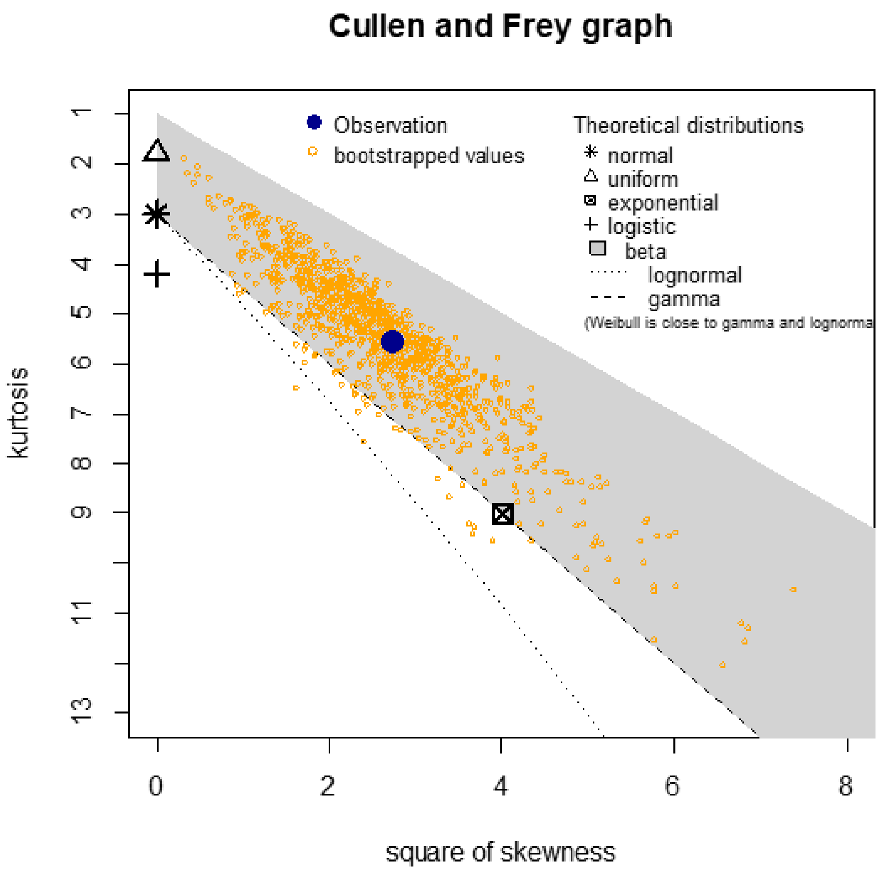

Figure 6 – C&F graph for Data Set 2.

The C&F graph for the second data set (Figure 6) suggests that the Gamma distribution could give a good fit. So, the outcomes of fitting the MPD and MW to the second data set were again compared to the results of fitting 2P-Weibull, Gamma, BGP and GBEP distributions, as well as the results of fitting other probability distributions of the same baseline as the MPD such as Kumaraswamy Pareto Distribution (KPD) by Bourguignon *et al.* (2012) and the Exponential Pareto Distribution (EPD) by Kareema and Boshi (2013), with respective density functions given by:

$$f_{BGP}(x;\alpha,\beta,\gamma,a,b)=\frac{\gamma}{B(a,b)}\left[1-\left(\frac{\alpha}{x}\right)^{\beta}\right]^{\gamma a-1}\left[1-\left(1-\left(\frac{\theta}{x}\right)^{\beta}\right)^{\gamma}\right]^{b-1} \times\frac{\beta}{\alpha}\left(\frac{\alpha}{x}\right)^{\beta+1}$$
$$\text{where } \alpha,\beta,\gamma,a,b>0 \text{ and } x\geq\alpha>0, \quad (33)$$

$$f_{GBEP}(x;\xi)=\frac{c\lambda k d^{k}x^{-(k+1)}}{B(a,b)}\left[1-\left(\frac{d}{x}\right)^{k}\right]^{\lambda ac-1}\left[1-\left(1-\left(\frac{d}{x}\right)^{k}\right)^{\lambda c}\right]^{b-1}$$
$$\text{for } x,a,b,c,d,\lambda,k>0, \quad (34)$$

$$f_{KP}(x;\alpha,\lambda,\theta,k)=\frac{\alpha\lambda k\theta^{k}}{x^{k+1}}\left[1-\left(\frac{\theta}{x}\right)^{k}\right]^{\alpha-1}\left[1-\left(1-\left(\frac{\theta}{x}\right)^{k}\right)^{\alpha}\right]^{\lambda-1}$$
$$\text{for } x,\alpha,\lambda,\theta,k>0, \quad (35)$$

$$f_{EP}(x;\lambda,\theta,k)=\frac{\lambda k}{\theta}\left(\frac{x}{\theta}\right)^{k-1}e^{-\lambda\left(\frac{x}{\theta}\right)^{k}}, \quad \text{for } x,\lambda,\theta,k>0. \quad (36)$$

The maximum likelihood estimated parameters along with the loglik, AIC and BIC are reported in Table 6. It is important to state that in the process of estimating the parameters of GBEP distribution, the value of d was set at $\hat{d}=0.1$ as is found in the work of Mead (2014).

In this case also, the MPD seems to be the best fitting model among the compared distributions, having the lowest values of AIC and BIC. The goodness-of-fit statistics for the best six models shown in Table 7 further support the flexibility of the MPD. The estimated densities of the compared distributions for the second data set are given in Figure 7.

TABLE 6
MLEs and goodness-of-fit criteria for Data Set 2.

Model	MLEs	loglik	AIC	BIC
MPD	$\hat{\alpha}=7.317\text{x}10^{-5}$, $\hat{\lambda}=3.563$ $\hat{\theta}=0.4276$	**-110.07**	**226.13**	**232.36**
MW	$\hat{\alpha}=0.00007$, $\hat{\beta}=0.3078$ $\hat{\gamma}=0.8530$	-190.74	387.47	393.70
BGP	$\hat{\alpha}=4.1$, $\hat{\beta}=0.509$, $\hat{\gamma}=2.860$, $\hat{a}=0.744$, $\hat{b}=5.891$	-195.52	401.03	409.34
GBEP	$\hat{a}=50.173$, $\hat{b}=1.612$, $\hat{c}=0.276$, $\hat{\lambda}=0.201$, $\hat{k}=1.092$	-196.52	403.04	413.43
KPD	$\hat{\alpha}=0.9373$, $\hat{\lambda}=0.9500$ $\hat{\theta}=0.0016$, $\hat{k}=0.0498$	-187.77	383.53	381.43
EP	$\hat{\lambda}=0.0991$, $\hat{\theta}=4.3552$ $\hat{k}=1.8401$	-197.29	400.58	406.81
Gamma	$\hat{\lambda}=3.6787784$, $\hat{\theta}=0.2727684$	-193.08	390.16	394.32
2P-Weibull	$\hat{\lambda}=15.306416$, $\hat{\beta}=1.840212$	-197.29	398.58	402.74

TABLE 7
Goodness-of-fit statistics for Data Set 2.

Statistic	Weibull	Gamma	EP	KPD	MW	MPD
K-S	0.1432	0.1336	0.3035	0.1243	0.3504	0.1183
CVM	0.2804	0.2028	1.3120	0.2001	1.4230	0.1832
A-D	1.8406	1.2301	6.9218	1.3985	7.0365	1.2013

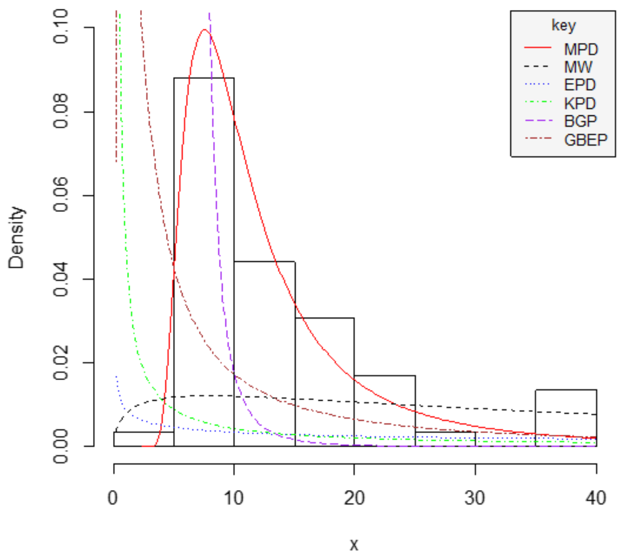

Figure 7 – Fitted densities for Data Set 2.

6. Conclusion

In this research paper, a Muth-Pareto distribution (MPD) is introduced with a failure rate function that has a non-monotone unimodal shape. Some of its mathematical properties including moments, quantile function, characterization and moments of order statistics are derived. The unknown parameters of the distribution are estimated using the method of maximum likelihood, and the performance of the estimator is investigated through simulation experiments. It was found that the maximum likelihood estimator behaves consistently in terms of the root mean squared error when the sample size gets large, and it looks asymptotically unbiased. The applicability of the proposed distribution is shown by means of two real-life applications. It was found that the MPD may be a better choice than its competing distributions for modeling positively skewed data sets having a non-monotone failure rate function.

REFERENCES

M. AHSANULLAH (1995). *Record Statistics*. Nova Science Publishers, New York, USA.

M. AHSANULLAH (2017). *Characterizations of Univariate Continuous Distributions*. Atlantis-Press, Paris, France.

M. AHSANULLAH, B. KIBRIA, M. SHAKIL (2014). *Normal and Student ´s t-Distributions and their Applications*. Atlantis-Press, Paris, France.

M. AHSANULLAH, V. NEVZOROV, M. SHAKIL (2013). *An Introduction to Order Statistics*. Atlantis-Press, Paris, France.

M. AHSANULLAH, M. SHAKIL, B. KIBRIA (2015). *Characterizations of folded Student's t-distribution*. Journal of Statistical Distributions and Applications, 2, no. 1, p. 15.

A. ALMARASHI, M. ELGARHY (2018). *A new muth-generated family of distributions with applications*. Journal of Nonlinear Sciences and Applications, 11, no. 1, pp. 1171–1184.

A. ALZAATREH, C. LEE, F. FAMOYE (2013). *A new method for generating families of continuous distributions*. Metron, 71, no. 1,2, p. 63–79.

B. ARNOLD, BALAKRISHNAN, H. NAGARAJA (2005). *First Course in Order Statistics*. Wiley, New York, USA.

M. BOURGUIGNON, R. S. B, L. ZEA, G. CORDEIRO (2012). *The kumaraswamy pareto distribution*. Statistical Methods, 42, pp. 1–20.

M. BOURGUIGNON, R. SILVA, G. CORDEIRO (2014). *The Weibull-G family of probability distributions*. Journal of Data Science, 12, no. 1, p. 53–68.

R. CORLESS, G. GONNET, D. HARE, D. JEFFREY, D. KNUTH (1996). *On the lambert w function*. Advances in Computational Mathematics, 5, no. 4, pp. 329–359.

H. DAVID, H. NAGARAJA (2003). *Order Statistics, Third Edition*. Wiley, New York, USA.

N. EUGENE, C. LEE, F. FAMOYE (2002). *Beta-normal distribution and its applications*. Communications in Statistics - Theory and Methods, 31, no. 1, p. 497–512.

J. GALAMBOS, S. KOTZ (2006). *Characterizations of Probability Distributions.: A Unified Approach with an Emphasis on Exponential and Related Models.*, vol. 675. Springer-Verlag, Berlin.

W. GLÄNZEL (1987). *A characterization theorem based on truncated moments and its application to some distribution families*. In P. BAUER, F. KONECNY, W. WERTZ (eds.), *Mathematical Statistics and Probability Theory*, Springer Netherlands, Dordrecht, pp. 75–84.

W. GLANZEL (1990). *Some consequences of a characterization theorem based on truncated moments*. Statistics, 21, no. 4, pp. 613–618.

W. GLÄNZEL, A. TELES, A. SCHUBERT (1984). *Characterization by truncated moments and its application to pearson-type distributions*. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 66, no. 2, pp. 173–183.

P. JODRÁ, H. GOMEZ, M. JIMÉNEZ-GAMERO, M. ALBA-FERNÁNDEZ (2017). *The power muth distribution*. Mathematical Modelling and Analysis, 22, no. 2, pp. 186–201.

P. JODRÁ, M. D. JIMÉNEZ GAMERO, M. V. ALBA FERNÁNDEZ (2015). *On the muth distribution*. Mathematical Modelling and Analysis, 20, no. 3, pp. 291–310.

A. KAREEMA, M. BOSHI (2013). *Exponential pareto distribution*. Mathemetical Theory and Modeling, 3, no. 5, pp. 135–146.

S. KOTZ, D. SHANBHAG (1980). *Some new approaches to probability distributions*. Advances in Applied Probability, 12, no. 4, pp. 903–921.

L. LEEMIS, J. MCQUESTON (2008). *Univariate distribution relationships*. The American Statistician, 62, no. 1, pp. 45–53.

M. MEAD (2014). *Extended pareto distribution*. Pakistan Journal of Statistics and Operations Research, 10, no. 3, pp. 313–329.

B. MILOŠEVIC (2017). *Some recent characterization based goodness of fit tests*. In T. WIKLUND (ed.), *the Proceedings of the 20th European Young Statisticians Meeting*. pp. 67–73.

MUSADDIQ SIRAJO (2020). *Generalization of Muth Distribution with It's Properties and Applications*. An MSc Dissertation of Department of Statistics, Ahmadu Bello University, Zaria-Nigeria.

E. J. MUTH (1977). *Reliability models with positive memory derived from the mean residual life function*. In C. P. TSOKOS, I. N. SHIMI (eds.), *Theory and applications of reliability*, Academic Press, pp. 410–434.

H. NAGARAJA (2006). *Characterizations of probability distributions*. In H. PHAM (ed.), *Springer Handbook of Engineering Statistics*, Springer, London, pp. 79–95.

M. NASSAR, N. NADA (2011). *The beta generalized pareto distribution*. Journal of statistics: Advances in Theory and Applications, 6, no. 1, pp. 1–17.

J. NELDER, R. MEAD (1965). *A simplex method for function minimization*. Computer Journal, 7, pp. 308–313.

V. B. NEVZOROV (2001). *Records: mathematical theory*. American Mathematical Society.

Y. Y. NIKITIN (2017). *Tests based on characterizations, and their efficiencies: a survey*. Acta et Commentationes Universitatis Tartuensis de Mathematica, 21, no. 1, pp. 3–24.

F. PROSCHAN (1963). *Theoretical explanation of observed decreasing failure rate*. Technometrics, 5, no. 1, pp. 375–383.

R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

M. SHAKIL, M. AHSANULLAH, B. G. KIBRIA (2018). *On the characterizations of chen's two-parameter exponential power life-testing distribution*. Journal of Statistical Theory and Applications, 17, no. 3, pp. 393–407.

M. H. TAHIR, G. M. CORDEIRO, S. ALI, S. DEY, A. MANZOOR (2018). *The inverted nadarajah–haghighi distribution: estimation methods and applications*. Journal of Statistical Computation and Simulation, 88, no. 14, pp. 2775–2798.

K. Y. VOLKOVA, Y. Y. NIKITIN (2015). *Exponentiality tests based on ahsanullah's characterization and their efficiency*. Journal of Mathematical Sciences, 204, no. 1, pp. 42–54.