PROPERTIES AND DATA MODELLING APPLICATIONS OF THE KUMARASWAMY GENERALIZED MARSHALL-OLKIN-G FAMILY OF DISTRIBUTIONS

Subrata Chakraborty 1 and Laba Handique[2]

Department of Statistics,

Dibrugarh University,

Dibrugarh-786004, India

[1]*Email: subrata_stats@dibru.ac.in*

[2]*Email: handiquelaba@gmail.com*

Abstract: In this article we propose further extension of the generalized Marshall-Olkin-G (GMO - G) family of distribution. The density and survival functions are expressed as infinite mixture of the GMO - G distribution. Asymptotes, Rényi entropy, order statistics, probability weighted moments, moment generating function, quantile function, median, random sample generation and parameter estimation are investigated. Selected distributions from the proposed family are compared with those from four sub models of the family as well as with some other recently proposed models by considering real life data fitting applications. In all cases the distributions from the proposed family out on top.

Key words: Kw-G family, MO-G family, AIC, KS test, LR test.

---

*Corresponding author

1   Introduction

Extension of existing well-known distributions to enhance flexibility in modelling variety of data has attracted attention of researchers recently. Some of the notable new family of distributions proposed of late includes among others the Kumaraswamy Marshall-Olkin-G family by Alizadeh et al. (2015), beta generated Kumaraswamy-G family by Handique et al. (2017a), Marshall-Olkin Kumaraswamy-G family by Handique et al. (2017b), Generalized Marshall-Olkin Kumaraswamy-G family by Chakraborty and Handique (2017),Transmuted Topp-Leone-G family by Haitham et al., 2017, beta generated Kumaraswamy Marshall-Olkin-G family by Handique and Chakraborty (2017a), beta generalized Marshall-Olkin Kumaraswamy-G family by Handique and Chakraborty (2017b), a new generalized family by Tahir et al. (2018) and exponentiated generalized Marshall-Olkin family by Handique et al. (2018) among others.

In this paper we propose another family of continuous probability distributions by integrating the $GMO - G(\theta, \alpha)$ family of Jayakumar and Mathew (2008) in the $Kw - G(a, b)$ family of Cordeiro and de Castro (2011) and refer to it as the Kumaraswamy Generalized Marshall-Olkin-G ($KwGMO - G(a, b, \theta, \alpha)$) family. Various mathematical and statistical characteristics along with comparative real data modelling illustration are presented. The probability density function (pdf) of $KwGMO -$

$G(a, b, \theta, \alpha)$ family of distribution is given by

$$f^{KwGMOG}(t; a, b, \theta, \alpha)$$
$$= \frac{ab\theta\alpha^\theta g(t)\bar{G}(t)^{\theta-1}}{[1 - \bar{\alpha}\bar{G}(t)]^{\theta+1}} [1$$
$$- \{\frac{\alpha\bar{G}(t)}{1 - \alpha\bar{G}(t)}\}^\theta]^{a-1}[1 - [1 - \{\frac{\alpha\bar{G}(t)}{1 - \alpha\bar{G}(t)}\}^\theta]^a]^{b-1}$$
$$, 0 < t < \infty, 0 < \alpha < \infty, 0 < \theta < \infty, a > 0, b > 0 \tag{1}$$

The corresponding cumulative distribution function (cdf), survival function (sf) and hazard rate function (hrf) are respectively given by

$$cdf: F^{KwGMOG}(t, a, b, \theta, \alpha) = 1 - [1 - [1 - [\alpha\bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}]^\theta]^a]^b \tag{2}$$

$$sf: \bar{F}^{KwGMOG}(t, a, b, \theta, \alpha) = [1 - [1 - [\alpha\bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}]^\theta]^a]^b;$$

$$hrf: h^{KwGMOG}(t, a, b, \theta, \alpha) \tag{3}$$

$$= \frac{ab\theta\alpha^\theta g(t)\bar{G}(t)^{\theta-1}}{[1 - \bar{\alpha}\bar{G}(t)]^{\theta+1}} [1 - \{\frac{\alpha\bar{G}(t)}{1 - \alpha\bar{G}(t)}\}^\theta]^{a-1}[1 - [1 - \{\frac{\alpha\bar{G}(t)}{1 - \alpha\bar{G}(t)}\}^\theta]^a]^{-1}$$

Particular cases: For

(i) $\theta = 1, f^{KwGMOG}(t; a, b, \theta, \alpha) = f^{KwMOG}(t; a, b, \alpha)$ (Alizadeh $et\ al.$, 2015),

(ii) $\alpha = 1, f^{KwGMOG}(t; a, b, \theta, \alpha) = f^{GKwG}(t; a, b, \theta)$,

(iii) $\alpha = b = 1, f^{KwGMOG}(t; a, b, \theta, \alpha)$
    $= f^{GMOG}(t; \theta, \alpha)$ (Jayakumar and Mathew, 2008),

(iv) $\theta = \alpha = 1, f^{KwGMOG}(t; a, b, \theta, \alpha)$
    $= f^{KwG}(t; a, b)$ (Cordeiro and de Castro, 2011) and

(v) $\alpha = b = \theta = 1, f^{KwGMOG}(t; a, b, \theta, \alpha)$
    $= f^{MOG}(t; \alpha)$ (Marshall and Olkin, 1997).

1.1 Physical Basis of KwGMOG

For $i = 1, 2, \ldots, b$ where $b > 1$ is an integer, if $\{T_{i1}, T_{i2,\ldots}, T_{ia}\}$ where $a > 1$ is an integerbe 'b' sequences of i.i.d. $GMO - G(\theta, \alpha)$ random variables with cdf $1 - [\alpha\bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}]^\theta$ then $min_{1 \le i \le b}\{max(T_{i1}, T_{i2,\ldots}, T_{ia})\}$ is distributed as $KwGMO - G(a, b, \theta, \alpha)$.

Proof : The cdf of $Z_i = max(T_{i1}, T_{i2,\ldots}, T_{ia})$ is obtained as

$$P(Z_i \le t) = [1 - [\alpha\bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}]^\theta]^a$$

Hence the cdf of $T = min(Z_1, Z_2, \ldots, Z_b)$ is given by

$$P(T \le t) = 1 - [1 - [1 - [\alpha\bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}]^\theta]^a]^b.$$

It may be noted that the $GMO - G(\theta, \alpha)$ distribution arises as the distribution of the minimum of θ i.i.d. $MO - G(\alpha)$ random variables.

The plotted pdf and hrf of the $KwGMO - G(a, b, \theta, \alpha)$ for some selected values of the parameters are presented in Figures 1 and 2, indicated lots of flexibility in shapes of the distributions of the family.
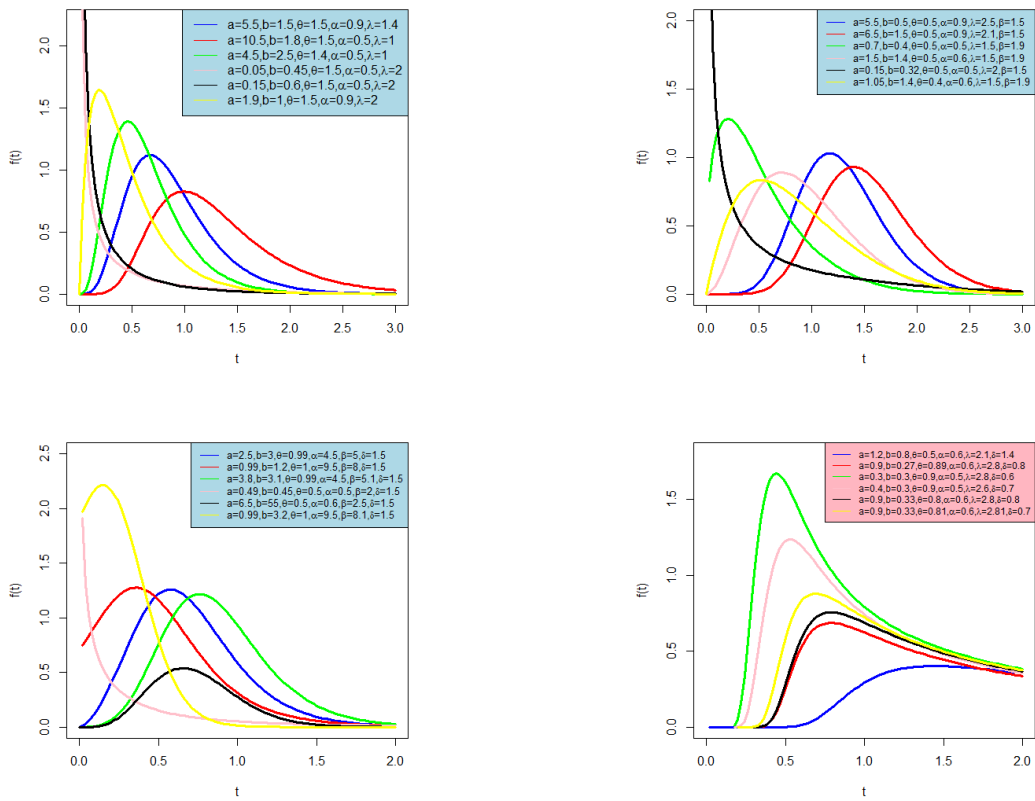
Fig. 1: Pdf plots top left: $KwGMO - E$, top right: $KwGMO - W$, bottom left: $KwGMO - L$ and bottom right: $KwGMO - Fr$ distributions

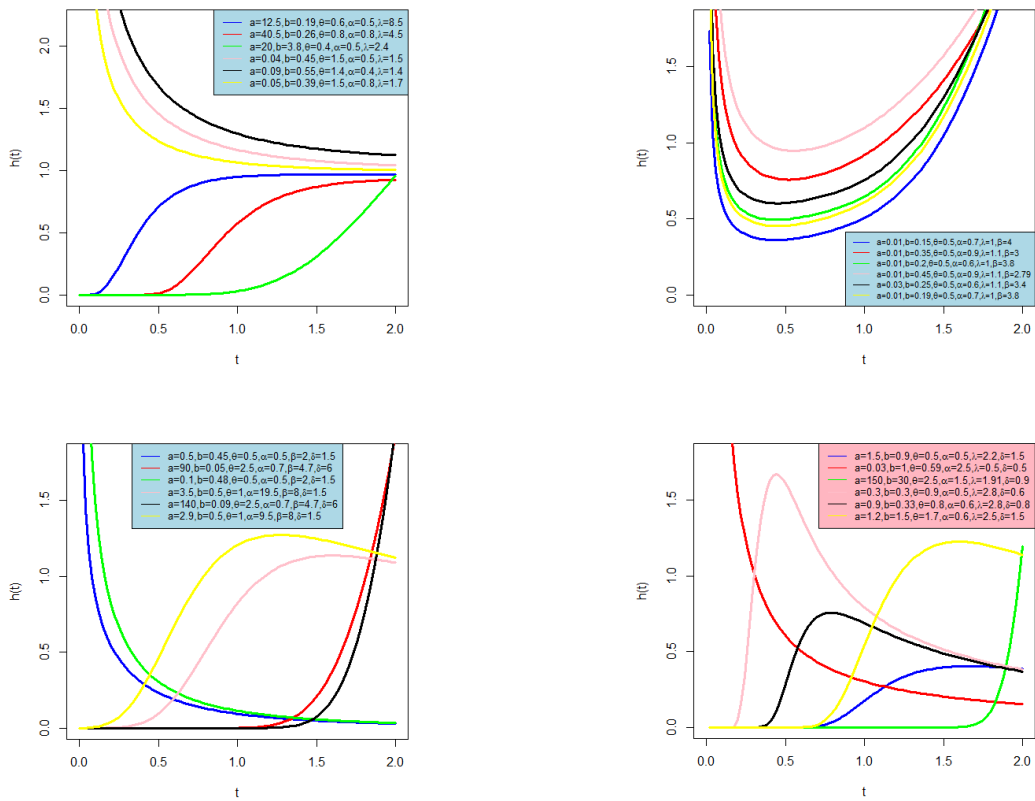
Fig. 2: Hrf plots top left: $KwGMO - E$, top right: $KwGMO - W$, bottom left: $KwGMO - L$ and bottom right: $KwGMO - Fr$ distributions.

The prime motivation behind the construction of this new family is to propose an extension of the $GMO - G(\theta, \alpha)$ distribution by adding two additional parameters which will cover some important distributions as special and related cases and also to bring in extra flexibility with respect to skewness, kurtosis, tail weight and length. Moreover distributions from this extended family should show significant improvement in data adjustment when compared to it sub models and other existing ones with respect to various model selection criteria, test of goodness-of-fits.

The rest of the paper is organized in four more Sections. In Section 2, we furnish mathematical properties of the new model. The maximum likelihood estimation is investigated in Section 3. Applications of real data are performed in Section 4. The paper ends with a conclusion in the last Section.

2 Mathematical & Statistical Properties

2.1 Linear Representation

By applying binomial expansion in equation (1), we obtain

$$f^{KwGMOG}(t; a, b, \theta, \alpha) = f^{GMO}(t; \theta, \alpha) \sum_{j=0}^{b-1} A_j \left[F^{GMO}(t; \theta, \alpha)\right]^{a(j+1)-1} \tag{4}$$

$$= \sum_{j=0}^{b-1} \frac{A_j}{a(j+1)} \frac{d}{dt} \left[F^{GMO}(t; \theta, \alpha)\right]^{a(j+1)}$$

$$= \sum_{j=0}^{b-1} A_j' \frac{d}{dt} \left[F^{GMO}(t; \theta, \alpha)\right]^{a(j+1)}, \tag{5}$$

where, $A_j' = \frac{(-1)^j}{a(j+1)} ab \binom{b-1}{j}$ and $A_j = A_j' a(j+1)$

We can also expand the pdf as

$$f^{KwGMOG}(t; a, b, \theta, \alpha) = f^{GMO}(t; \theta, \alpha) \sum_{k=0}^{a(j+1)-1} B_k \left[\bar{F}^{GMO}(t; \theta, \alpha)\right]^k \tag{6}$$

$$= -\sum_{k=0}^{a(j+1)-1} B_k' \frac{d}{dt} \left[\bar{F}^{GMO}(t; \theta, \alpha)\right]^{k+1} \tag{7}$$

$$= \sum_{k=0}^{a(j+1)-1} B_k' f^{GMO}(t; \theta, \alpha(k+1)) \tag{8}$$

where $B_k' = \frac{1}{k+1} \sum_{j=0}^{b-1} A_j (-1)^k \binom{a(j+1)-1}{k}$ and $B_k = B_k'(k+1)$.

Similar expansion for the sf of $KwGMO - G(a, b, \theta, \alpha)$ is given by

$$\bar{F}^{KwGMOG}(t; a, b, \theta, \alpha) = \sum_{l=0}^{b} C_l \left[F^{GMO}(t; \theta, \alpha)\right]^{al}, where \ C_l = \binom{b}{l}(-1)^l. \tag{9}$$

and

$$\bar{F}^{KwGMOG}(t; a, b, \theta, \alpha) = \sum_{m=0}^{al} \delta_m [\bar{F}^{GMO}(t; \theta, \alpha m)], \tag{10}$$

where $\delta_m = \sum_{l=0}^{b} C_l \binom{al}{m}(-1)^m$.

2.2 Asymptotes

Asymptotes of the proposed family are discussed in the following two propositions.
Proposition 1. The asymptotes of equations (6), (7) and (8) as $t \to 0$ are given by

$$f(t) \sim [\{\theta a b g(t)\}/\alpha][1 - \{\alpha \bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}\}^\theta]^{a-1} \qquad as\ t \to 0;$$
$$F(t) \sim 1 - [1 - \{1 - \bar{G}(t)^\theta\}^a]^b \qquad as\ t \to 0;$$
$$h(t) \sim [\{\theta a b g(t)\}/\alpha][1 - \{\alpha \bar{G}(t)/\{1 - \bar{\alpha}\bar{G}(t)\}\}^\theta]^{a-1} \qquad as\ t \to 0.$$

Proposition 2. The asymptotes of equations (6), (7) and (8) as $t \to \infty$ are given by

$$f(t) \sim a b \theta \alpha^\theta g(t) \bar{G}(t)^{\theta-1}[1 - [1 - \{\alpha \bar{G}(t)\}^\theta]^a]^{b-1} \qquad as\ t \to \infty;$$
$$1 - F(t) \sim [1 - [1 - \{\alpha \bar{G}(t)\}^\theta]^a]^b \qquad as\ t \to \infty;$$
$$h(t) \sim a b \theta \alpha^\theta g(t) \bar{G}(t)^{\theta-1}[1 - [1 - \{\alpha \bar{G}(t)\}^\theta]^a]^{-1} \qquad as\ t \to \infty.$$

2.3 Rényi entropy

The Rényi entropy is defined by $I_R(\delta) = (1 - \delta)^{-1} \log(\int_{-\infty}^{\infty} f(t)^\delta\ dt)$, where $\delta > 0$ and $\delta \neq 1$. Using binomial expansion in equation (6) we can write

$$f^{KwGMOG}(t; a, b, \theta, \alpha)^\delta = f^{GMO}(t; \theta, \alpha)^\delta \sum_{i=0}^{\delta(b-1)} Z_i\ [F^{GMO}(t; \theta, \alpha)]^{a(i+\delta)-\delta}.$$

Thus, $I_R(\delta) = (1 - \delta)^{-1} log \left(\int_0^\infty f^{GMO}(t; \theta, \alpha)^\delta \sum_{i=0}^{\delta(b-1)} Z_i\ [F^{GMO}(t; \theta, \alpha)]^{a(i+\delta)-\delta}\ dt \right)$

$$= (1 - \delta)^{-1} log \left(\sum_{i=0}^{\delta(b-1)} Z_i \int_0^\infty f^{GMO}(t, \theta, \alpha)^\delta)[F^{GMO}(t; \theta, \alpha)]^{a(i+\delta)-\delta} dt \right)$$

Where $Z_i = \{ab\}^\delta \binom{\delta(b-1)}{i}(-1)^i$.

2.4 Order Statistics

Given a random sample $T_1, T_2, \ldots, T_n$ from any $KwGMO - G(a, b, \theta, \alpha)$ distribution let $T_{i:n}$ be the ith order statistics. Then pdf of $T_{i:n}$ is given by

$$f_{i:n}(t) = \frac{n!}{(i-1)!\,(n-1)!} f^{KwGMOG}(t; a, b, \theta, \alpha)$$
$$[1 - \bar{F}^{KwGMOG}(t; a, b, \theta, \alpha)]^{i-1} \bar{F}^{KwGMOG}(t)^{n-i}$$

$$= \frac{n!}{(i-1)!\,(n-1)!} f^{KwGMOG}(t; a, b, \theta, \alpha) \sum_{k=0}^{i-1}(-1)^k \binom{i-1}{k} \bar{F}^{KwGMOG}(t; a, b, \theta, \alpha)^{n+k-i}.$$

Using the results of section 2.1 the pdf of $T_{i:n}$ can be written as

$$f_{i:n}(t) = f^{GMO}(t; \theta, \alpha) \sum_{m=0}^{al} \sum_{p=0}^{a(j+1)-1} \eta_{m,p}\ [\bar{F}^{GMO}(t; \theta, \alpha)]^{m(n+k+i)+p} \qquad (11)$$

$$= \sum_{m=0}^{al} \sum_{p=0}^{a(j+1)-1} \eta'_{m,p}\ f^{GMO}\big(t; \theta, \alpha(m(n+k+i)+p+1)\big), \qquad (12)$$

where, $\eta_{m,p} = n\delta_m B_p \binom{n-1}{i-1} \sum_{k=0}^{i-1} \binom{i-1}{k}(-1)^k$ and $\eta'_{m,p} = \eta_{m,p}/[m(n+k+i)+p+1]$ and B_P, δ_m defined in Section 2.1.

2.5 Probability weighted moments

Greenwood *et al.* (1979) first proposed probability weighted moments (PWMs), as the expectations of certain functions of a random variable whose mean exists. The $(p, q, r)^{th}$ PWM of T is defined by $\Gamma_{p,q,r} = \int_{-\infty}^{\infty} t^p [F(t)]^q [1 - F(t)]^r f(t) dt.$ From equations (4) and (6) the sth moment of T can be written either as

$$E(T^s) = \int_0^{\infty} t^s f^{GMO}(t; \theta, \alpha) \sum_{j=0}^{b-1} A_j [F^{GMO}(t; \theta, \alpha)]^{a(j+1)-1} dt$$

$$= \sum_{j=0}^{b-1} A_j \int_0^{\infty} t^s [1 - [\alpha\bar{G}(t)/1 - \bar{\alpha}\bar{G}(t)]^{\theta}]^{a(j+1)-1} [\theta \alpha^{\theta} g(t)\bar{G}(t)^{\theta-1}/[1 - \bar{\alpha}\bar{G}(t)]^{\theta+1}] dt$$

$$= \sum_{j=0}^{b-1} A_j \, \Gamma_{s,a(j+1)-1,0}$$

and $(T^s) \sum_{k=0}^{a(j+1)-1} B_K \, \Gamma_{S,0,K}$, respectively.

Where

$\Gamma_{p,q,r} =$

$$\int_{-\infty}^{\infty} t^p \{1 - [\alpha\bar{G}(t)/1 - \bar{\alpha}\bar{G}(t)]^{\theta}\}^q \{[\alpha\bar{G}(t)/1 - \bar{\alpha}\bar{G}(t)]^{\theta}\}^r [\theta\alpha^{\theta} g(t)\bar{G}(t)^{\theta-1}/[1 - \bar{\alpha}\bar{G}(t)]^{\theta+1}] dt$$

is the PWM of $GMO - G(\theta, \alpha)$ distribution.

Thus the moments of the $KwGMO - G(a, b, \theta, \alpha)$ can be expressed in terms of the PWMs of $GMO - G(\theta, \alpha)$. Similarly, we can derive s^{th} moment of the ith order statistic $T_{i:n}$,in a random sample of size n from $KwGMO - G(a, b, \theta, \alpha)$ on using equation (11) as $E(T_{i,n}^S) = \sum_{m=0}^{al} \sum_{p=0}^{a(j+1)-1} \eta_{m,p} \, \Gamma_{s,0,m(n+k+i)+p}$,where A_j, B_k and $\eta_{m,p}$ are defined in section 2.1 and 2.2 respectively.

2.6 Moment generating function (mgf)

The mgf of $KwGMO - G(a, b, \theta, \alpha)$ family can be easily expressed in terms of those of the exponentiated $GMO - G(\theta, \alpha)$ distribution using the results of section 2.1. For example using equation (7) it can be seen that

$$M_T(s) = E[e^{sT}] = \int_{-\infty}^{\infty} e^{st} f(t) dt = -\int_0^{\infty} e^{st} \sum_{k=0}^{a(j+1)-1} B_k' \frac{d}{dt} [\bar{F}^{GMO}(t; \theta, \alpha)]^{k+1} dt$$

$$= -\sum_{k=0}^{a(j+1)-1} B_k' \int_0^{\infty} e^{st} \frac{d}{dt} [\bar{F}^{GMO}(t; \theta, \alpha)]^{k+1} dt = \sum_{k=0}^{a(j+1)-1} B_k \, M_X(s).$$

where $M_X(s)$ is the mgf of a $GMO - G(\theta, \alpha)$ distribution.

2.7 Quantile function, median and random sample generation

The pth quantile t_p for $KwGMO - G(a, b, \theta, \alpha)$ can be easily obtained by solving the equation $F^{KwGMOG}(t) = p$ as
$$t_p = G^{-1}\big(1 - \{[1 - \{1 - \{1 - p\}^{1/b}\}^{1/a}]^{1/\theta}\}/\{\alpha + \bar{\alpha}[1 - \{1 - \{1 - p\}^{1/b}\}^{1/a}]^{1/\theta}\}\big).$$
A random number 't' from $KwGMO - G(a, b, \theta, \alpha)$ via an uniform random number 'u' can be generated by using the formula
$$t_p = G^{-1}(1 - \{[1 - \{1 - \{1 - p\}^{1/b}\}^{1/a}]^{1/\theta}\}/ \tag{13}$$
$$\{\alpha + \bar{\alpha}[1 - \{1 - \{1 - p\}^{1/b}\}^{1/a}]^{1/\theta}\}).$$

Median:

$$t = G^{-1}\big(1 - \{[1 - \{1 - \{1 - 0.5\}^{1/b}\}^{1/a}]^{1/\theta}\}/\{\alpha + \bar{\alpha}[1 - \{1 - \{1 - 0.5\}^{1/b}\}^{1/a}]^{1/\theta}\}\big).$$

For example, when we consider the exponential distribution having pdf and cdf as $g(t:\lambda) = \lambda e^{-\lambda t}, t > 0 \; \lambda > 0$, and $G(t:\lambda) = 1 - e^{-\lambda t}$, as our G, the pth quantile, t_p is given by

$$t_p = -\frac{1}{\lambda}\log[1 - [1 - \{[1 - \{1 - p\}^{1/b}\}^{1/a}]^{1/\theta}\}/$$
$$\{\alpha + \bar{\alpha}[1 - \{1 - \{1 - p\}^{1/b}\}^{1/a}]^{1/\theta}\}]].$$

It may be noted that for given uniform random number 'u' corresponding random numbers 't' from $KwGMO - G(a, b, \theta, \alpha)$ can be easily obtained using equation (13).

3 Maximum likelihood estimation

Let $t = (t_1, t_{2,...,}, t_n)^T$ be a random sample of size n from $KwGMO - G(a, b, \theta, \alpha)$ with parameter vector $\boldsymbol{\rho} = (a, b, \theta, \alpha, \boldsymbol{\beta}^T)^T$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_q)^T$ corresponds to the parameter vector of the baseline distribution G. Then the log-likelihood function for $\boldsymbol{\rho}$ is given by

$$\ell = \ell(\boldsymbol{\rho}) = n \log(ab) + n \log\theta + n\theta \log\alpha \tag{14}$$

$$+ \sum_{i=1}^{n} \log[g(t_i, \boldsymbol{\beta}) + (\theta - 1)] \sum_{i=1}^{n} \log[\bar{G}(t_i, \boldsymbol{\beta})]$$

$$- (\theta + 1) \sum_{i+1}^{n} \log[1 - \bar{\alpha}\bar{G}(t_i, \boldsymbol{\beta})]$$

$$+ (a - 1) \sum_{i=1}^{n} \log[1 - [\alpha\bar{G}(t_i, \boldsymbol{\beta})/\{1 - \bar{\alpha}\bar{G}(t_i, \boldsymbol{\beta})\}]^{\theta}]$$

$$+ (b - 1) \sum_{i=1}^{n} \log[1 - [1 - [\alpha\bar{G}(t_i, \boldsymbol{\beta})/\{1 - \bar{\alpha}\bar{G}(t_i, \boldsymbol{\beta})\}]^{\theta}]^{a}].$$

The mle are obtained by maximizing the log-likelihood function numerically by using Optim function of R.

The asymptotic variance-covariance matrix of the mles of parameters can obtained by inverting the Fisher information matrix $I(\boldsymbol{\rho})$ which can be derived using the second partial derivatives of the log-likelihood function with respect to each parameter. The $(i\,j)^{th}$ elements of $I_n(\boldsymbol{\rho})$ are given by

$$I_{i\,j} = -E[\partial^2 l(\boldsymbol{\rho})/\partial \rho_i \partial \rho_j], \qquad i,j = 1,2,...,4 + q.$$

In practice one can estimate $I_n(\boldsymbol{\rho})$ by the observed Fisher's information matrix $\hat{I}_n(\hat{\boldsymbol{\rho}}) = (\hat{I}_{i\,j})$ defined as:

$$\hat{I}_{i\,j} \approx (-\partial^2 l(\boldsymbol{\rho})/\partial \rho_i \partial \rho_j)_{\eta = \hat{\eta},} \; i,j = 1,2,...,4 + q$$

Using the general theory of mle under some regularity conditions on the parameters as $n \to \infty$ the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ is $N_k(o, V_n)$ where $V_n = (v_{jj}) = I_n^{-1}(\boldsymbol{\rho})$. The asymptotic behaviour remains valid if V_n is replaced by $\hat{V}_n = \hat{I}^{-1}(\hat{\boldsymbol{\rho}})$. Using this result large sample standard errors of jth parameter ρ_j is given by $\sqrt{\hat{v}_{jj}}$.

4 Real life applications

Here we consider fitting of two real data sets to show that the distributions from the proposed $KwGMO - G(a, b, \theta, \alpha)$ family can provide better model than the corresponding distributions from the sub models $MO - G, Kw - G, GMO - G, KwMO - G$ by taking exponential (E) and Frechet (Fr) distribution as G. We also compare the proposed family with the recently introduced $GMOKw -$

$G(a, b, \theta, \alpha)$ (Chakraborty and Handique, 2017) family by taking exponential and Frechet as the base line ^{G} distribution.

We have considered four model selection criteria the AIC, BIC, CAIC and HQIC and the Kolmogorov-Smirnov (K-S) statistics for goodness of fit to compare the fitted models. Asymptotic standard errors and confidence intervals of the mles of the parameters for each competing models are also computed.

The $KwGMO - G(a, b, \theta, \alpha)$ distribution reduces to $MO - G(\alpha)$ when $a = b = \theta = 1$, to $Kw - G(a, b)$ if $\theta = \alpha = 1$, to $GMO - G(\theta, \alpha)$ for $a = b = 1$and to $KwMO - G(a, b, \alpha)$ for $\theta = 1$, hence we have conducted the likelihood ratio (LR) test for nested models to test following hypotheses:

    (i) $H_0: a = b = \theta = 1$, that is the sample is from $MO - G(\alpha)$.
                   $H_1: a \neq b \neq \theta \neq 1$, that is the sample is from $KwGMO - G(a, b, \theta, \alpha)$.
    (ii) $H_0: \theta = \alpha = 1$, that is the sample is from $Kw - G(a, b)$.
                   $H_1: \theta \neq 1, \alpha \neq 1$, that is the sample is from $KwGMO - G(a, b, \theta, \alpha)$.
    (iii) $H_0: \alpha = b = 1$, that is the sample is from $GMO - G(\theta, \alpha)$
                   $H_1: \alpha \neq b \neq 1$, that is the sample is from $KwGMO - G(a, b, \theta, \alpha)$.
    (iv) $H_0: \theta = 1$, that is the sample is from $KwGMO - G(a, b, \alpha)$.
                   $H_1: \theta \neq 1$, that is the sample is from $KwGMO - G(a, b, \theta, \alpha)$.

The LR test statistic is given by $LR = -2ln(L(\widehat{\rho}^*; x)/L(\widehat{\rho}; x))$, where $\widehat{\rho}^*$ is the restricted ML estimates under the null hypothesis H_0 and $\boldsymbol{\rho}$ is the unrestricted ML estimates under the alternative hypothesis H_1. Under the null hypothesis H_0 the LR criterion follows Chi-square distribution with degrees of freedom (df) $(df_{alt} - df_{null})$.The null hypothesis is can not be accepted for p-value less than 0.05.

Fitted densities and the fitted cdf's are presented in the form of a Histograms and Ogives of the data in Figures 4, 5, 6 and 7. These plots reveal that the proposed distributions provide a good fit to these data.

Data Set I: Consists of 100 observations of breaking stress of carbon fibres (in Gba) given by Nichols and Padgett (2006).

Data Set II: Survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli, observed and reported by Bjerkedal (1960), also used by Shibu and Irshad(2016).

TTT plots and Descriptive Statistics for the data sets:

The total time on test (TTT) plot (see Aarset, 1987) is a technique to extract the information about the shape of the hazard function. A straight diagonal line indicates constant hazard for the data set, where as a convex (concave) shape implies decreasing (increasing) hazard. The TTT plots for the data sets in Fig. 3 indicate that the both data sets have increasing hazard rate. We have also presented the descriptive statistics of the data sets in Table 1

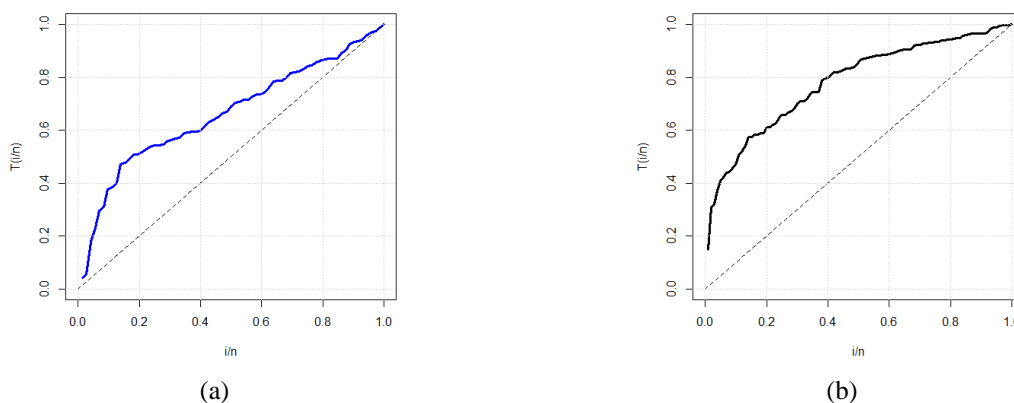   (a)   (b)
Fig. 3: TTT-plots for (a) Data set I and (b) Data set II

Table 1: Descriptive Statistics for data sets I and II

Data Sets	n	Min.	Mean	Median	s.d.	Skewness	Kurtosis	1st Qu.	3rd Qu.	Max.
I	100	0.390	2.621	2.700	1.014	0.362	0.043	1.840	3.220	5.560
II	72	0.080	1.837	1.560	1.215	1.718	3.955	1.080	2.303	7.000

Table 2 (a): MLEs, standard errors, confidence interval (in parentheses) for the data set I

Models	$\hat{a}$	$\hat{b}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\lambda}$
$MO - E(\alpha, \lambda)$	-	-	-	75.665 (33.464) (10.07,141.25)	1.535 (0.153) (1.24,1.83)
$Kw - E(a, b, \lambda)$	3.439 (0.552) (2.36,4.52)	48.150 (14.213) (20.29,76.01)	-	-	0.129 (0.103) (0,0.33)
$GMO - E(\theta, \alpha, \lambda)$	-	-	6.936 (1.258) (4.47,9.40)	122.034 (49.541) (24.933,219.13)	0.967 (0.128) (0.71,1.22)
$KwMO - E(a, b, \alpha, \lambda)$	2.647 (2.066) (0,6.69)	4.571 (14.461) (0,32.91)	-	3.899 (10.290) (0,24.07)	0.591 (1.103) (0,275)
KwGMO $- \mathbf{E(a, b, \theta, \alpha, \lambda)}$	2.342 (1.137) (0.11,4.57)	2.077 (0.449) (1.19,2.96)	1.813 (1.207) (0,4.19)	11.829 (2.631) (7.20,16.46)	0.829 (0.523) (0,1.85)
$GMOKw - E(a, b, \theta, \alpha, \lambda)$	3.524 (2.167) (0,7.77)	3.363 (12.168) (0,27.21)	1.614 (1.215) (0,3.99)	6.891 (2.893) (1.22,12.56)	0.549 (1.127) (0,293)

Table 2 (b): AIC, BIC, CAIC, HQIC, K-S (p-value) and LR (p-value) values for the data set I

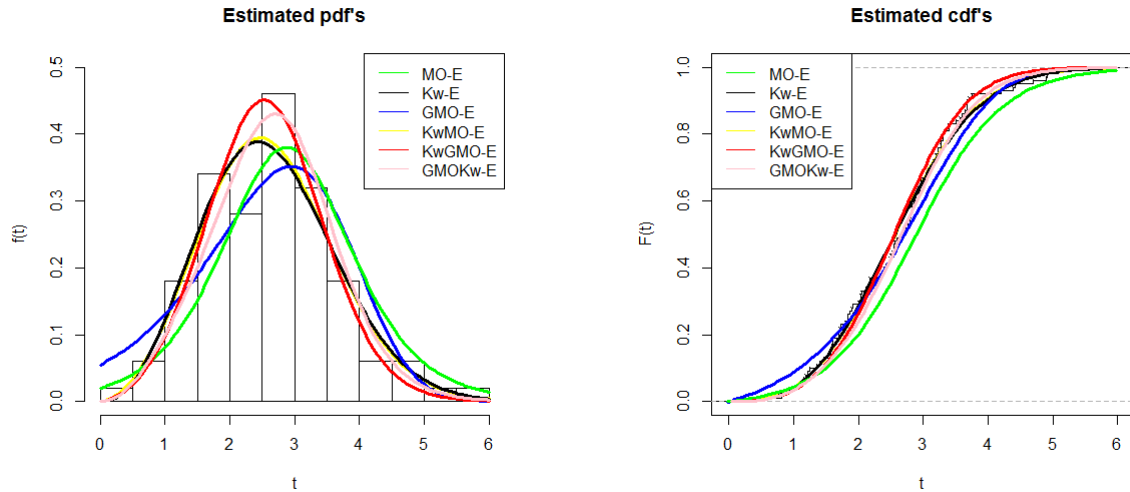| Models	(-1_{max})	AIC	BIC	CAIC	HQIC	K-S (p-value)	LR (p-value)
$MO - E(\alpha, \lambda)$	147.52	299.04	304.26	299.16	301.16	0.13 (0.11)	20.58 (0.0001)
$Kw - E(\alpha, b, \lambda)$	145.32	296.64	304.47	296.89	299.82	0.07 (0.73)	16.18 (0.0003)
$GMO - E(\theta, \alpha, \lambda)$	146.90	299.80	307.63	300.05	302.98	0.09 (0.39)	19.34 (0.04)
$KwMO - E(a, b, \alpha, \lambda)$	141.25	290.50	300.94	290.92	294.74	0.10 (0.28)	8.04 (0.004)
KwGMO $- \mathbf{E(a, b, \theta, \alpha, \lambda)}$	**137.23**	**284.46**	**297.51**	**285.09**	**289.76**	0.07 **(0.76)**	-
$GMOKw - E(a, b, \theta, \alpha, \lambda)$	138.13	286.26	299.31	286.89	291.56	0.08 (0.65)	-

Fig. 4: Plots of the observed histogram and estimated pdf's on the left and observed ogive and estimated cdf's for the $MO - E, Kw - E, GMO - E, KwMO - E, kwGMO - E$ and $GMOKw - E$ distributions for data set I.

Table 3 (a):   MLEs, standard errors, confidence interval (in parentheses) for the data set I

Models	$\hat{a}$	$\hat{b}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\delta}$
$MO - Fr(\alpha, \lambda, \delta)$	-	-	-	4185.00 (5976.04) (0,15898.3)	4.118 (0.342) (3.44,4.79)	0.328 (0.163) (0.008,0.65)
$Kw - Fr(a, b, \lambda, \delta)$	55.362 (1.429) (52.56,58.16)	140.774 (34.243) (73.66,207.89)	-	-	0.518 (0.213) (0.10,0.94)	0.027 (0.008) (0.01,0.04)
$GMO - Fr(\theta, \alpha, \lambda, \delta)$	-	-	249.805 (67.241) (118.01,381.59)	0.558 (0.039) (0.48,0.634)	0.431 (0.021) (0.39,0.47)	186.990 (37.682) (113.13,260.85)
$KwMO - Fr(a, b, \alpha, \lambda, \delta)$	9.243 (2.958) (3.45,15.04)	19.203 (14.763) (0,48.14)	-	17.222 (7.881) (1.78,32.67)	0.953 (0.172) (0.62,1.29)	(0.051) (0.026) (0.03,010)
KwGMO **$- Fr(a, b, \theta, \alpha, \lambda, \delta)$**	4.153 (2.813) (0,9.67)	13.863 (7.651) (0,28.86)	52.487 (21.241) (10.85,94.11)	7.314 (1.869) (3.65,10.98)	0.399 (0.093) (0.22,0.58)	24.312 (15.780) (0,55.24)
$GMOKw - Fr(a, b, \theta, \alpha, \lambda, \delta)$	4.132 (1.731)	43.694 (9.364)	3.55.563 (122.431)	0.111 (0.013)	0.232 (0.087)	256.127 (69.321)

Table 3 (b): AIC, BIC, CAIC, HQIC, K-S (p-value) and LR (p -value) values for the data set I

Models	$(-l_{max})$	AIC	BIC	CAIC	HQIC	K-S (p-value)	LR (p-value)
$MO - Fr(\alpha, \lambda, \delta)$	146.28	298.56	306.39	298.81	301.74	0.09 (0.32)	14.02 (0.003)
$Kw - Fr(a, b, \lambda, \delta)$	144.69	297.38	307.82	297.80	301.62	0.08 (0.51)	10.84 (0.004)
$GMO - Fr(\theta, \alpha, \lambda, \delta)$	144.87	297.74	308.18	298.16	301.98	0.10 (0.34)	11.20 (0.004)
$KwMO - Fr(a, b, \alpha, \lambda, \delta)$	144.83	299.66	312.71	300.29	304.96	0.10 (0.35)	11.12 (0.0008)
KwGMO $- Fr(a, b, \theta, \alpha, \lambda, \delta)$	**139.27**	**290.54**	**306.20**	**291.44**	**296.90**	0.08 **(0.54)**	-
$GMOKw - Fr(a, b, \theta, \alpha, \lambda, \delta)$	141.22	294.44	310.10	295.34	300.80	0.09 (0.45)	-
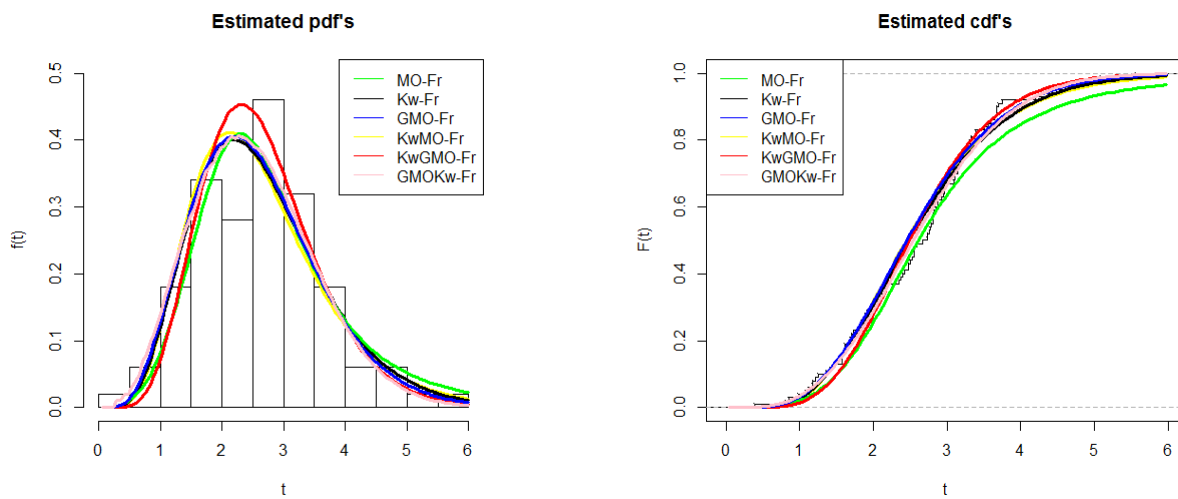
Fig. 5: Plots of the observed histogram and estimated pdf's on the left and observed ogive and estimated cdf's for the MO − Fr, Kw − Fr, GMO − Fr, KwMO − Fr, kwGMO − Fr and GMOKw − Fr distributions for data set I.

Table 4 (a): MLEs, standard errors, confidence interval (in parentheses) for the data set II

Models	$\hat{a}$	$\hat{b}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\lambda}$
$MO - E(\alpha, \lambda)$	-	-	-	8.778 (3.555) (1.81,15.75)	1.379 (0.193) (1.00,1.76)
$Kw - E(a, b, \lambda)$	3.304 (1.106) (1.14,5.47)	1.100 (0.764) (0,2.59)	-	-	1.037 (0.614) (0,2.24)
$GMO - E(\theta, \alpha, \lambda)$	-	-	0.179 (0.070) (0.04,0.32)	47.635 (44.901) (0,135.64)	4.465 (1.327) (1.86,7.07)
$KwMO - E(a, b, \alpha, \lambda)$	3.478 (1.891) (0,7.18)	3.306 (1.438) (0.48,6.12)	-	0.374 (0.127) (0.13,0.62)	0.295 (0.861) (0,1.98)
KwGMO $- E(a, b, \theta, \alpha, \lambda)$	1.276 (0.411) (0.47,2.08)	1.686 (0.734) (0.25,3.12)	0.164 (0.018) (0.12,0.19)	24.304 (15.072) (0,53.85)	4.308 (1.318) (1.72,6.89)
$GMOkw - E(a, b, \theta, \alpha, \lambda)$	2.855 (0.504) (1.86,3.84)	1.386 (6.534) (0,14.19)	1.616 (1.086) (0,3.74)	0.005 (0.003) (0,0.01)	0.072 (0.151) (0,0.37)

Table 4 (b): AIC, BIC, CAIC, HQIC, K-S (p-value) and LR (p -value) values for the data set II

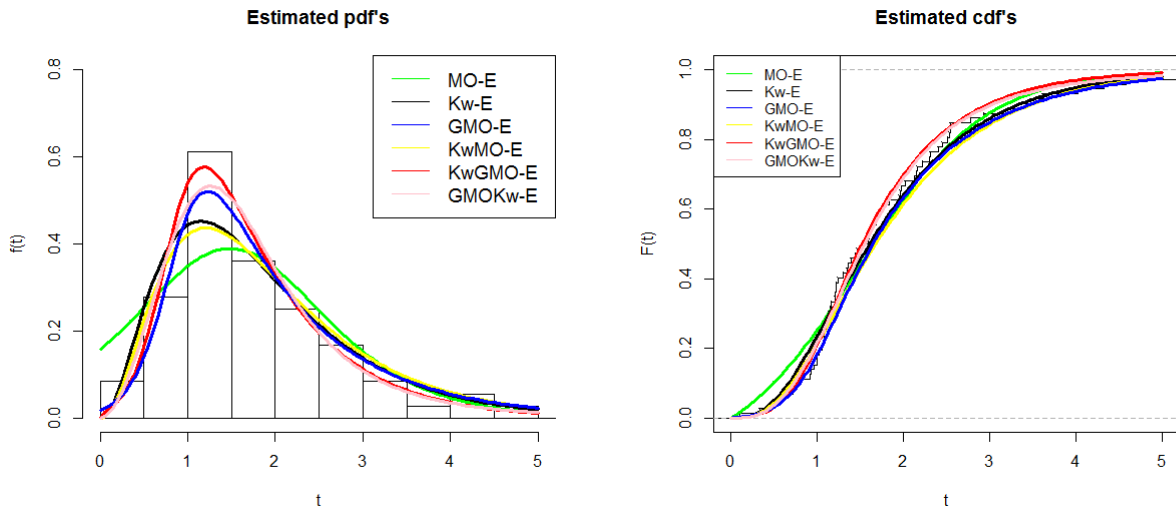| Models	$(-l_{max})$	AIC	BIC	CAIC	HQIC	K-S (p-value)	LR (p-value)
$MO - E(\alpha, \lambda)$	103.18	201.36	214.90	210.53	212.16	0.10 (0.43)	14.32 (0.002)
$Kw - E(a, b, \lambda)$	101.71	209.42	216.23	209.77	212.12	0.08 (0.50)	11.38 (0.003)
$GMO - E(\theta, \alpha, \lambda)$	102.27	210.54	217.38	201.89	213.24	0.09 (0.51)	12.50 (0.002)
KwMO $- E(a, b, \alpha, \lambda)$	100.51	209.02	218.10	209.62	212.63	0.07 (0.79)	8.98 (0.003)
KwGMO $- \mathbf{E(a, b, \theta, \alpha, \lambda)}$	**96.02**	**202.04**	**213.44**	**202.94**	**206.54**	0.07 **(0.81)**	-
GMOKw $- E(a, b, \theta, \alpha, \lambda)$	96.62	203.24	214.64	204.14	207.74	0.07 (0.78)	-

Fig. 6: Plots of the observed histogram and estimated pdf's on left and observed ogive and estimated cdf's for the $MO - E, Kw - E, GMO - E, KwMO - E, kwGMO - E$ and $GMOKw - E$ distributions for data set II.

Table 5 (a): MLEs, standard errors, confidence interval (in parentheses) for the data set II

Models	$\hat{a}$	$\hat{b}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\delta}$
$MO - Fr(\alpha, \lambda, \delta)$	-	-	-	7.231 (2.231)	2.884 (0.280)	0.871 (0.149)
$Kw - Fr(a, b, \lambda, \delta)$	1.503 (0.249 (1.01,1.99)	333.439 (113.651) (110.68,556.19)	-	-	0.291 (0.125) (0.05,0.54)	206.131 (67.237) (74.35,337.91)
GMO $- Fr(\theta, \alpha, \lambda, \delta)$	-	-	248.241 (367.908) (0,969.34)	2.486 (2.291) (0,6.98)	0.352 (0.035) (0.28,0.42)	150.838 (70.899) (11.86,289,80)
KwMO $- Fr(a, b, \alpha, \lambda, \delta)$	1.768 (0.462) (0.86,2.67)	36.031 (11.326) (13.83,58.23)	-	0.298 (0.088) (0.12,0.47)	0.328 (0.103) (0.13,0.53)	68.668 (62.923) (0,191.99)
KwGMO $- \mathbf{Fr(a, b, \theta, \alpha, \lambda, \delta)}$	21.333 (21.310) (0,63.10)	28.254 (30.266) (0,87.58)	40.079 (15.673) (9.36,70.79)	7.227 (1.329) (4.62,9.83)	0.127 (0.056) (0.02,0.24)	37.001 (40.878) (0,117.12)
GMOKw $- Fr(a, b, \theta, \alpha, \lambda, \delta)$	1.823 (0.192) (1.44,2.19)	10.662 (8.912) (0,28.13)	24.571 (21.836) (0,67.37)	4.018 (1.552) (0.98,7.06)	0.358 (0.073) (0.21,0.50)	21.007 (7.362) (6.57,35.43)

Table 5 (b): AIC, BIC, CAIC, HQIC, K-S (p-value) and LR (p -value) values for the data set II

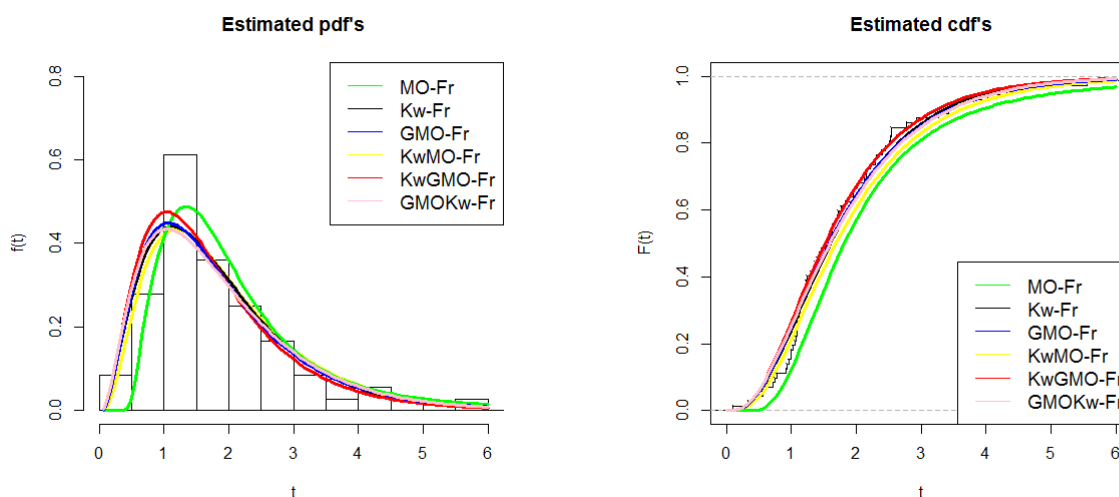| Models	$(-l_{max})$	AIC	BIC	CAIC	HQIC	K-S (p-value)	LR (p-value)
$MO - Fr(\alpha, \lambda, \delta)$	103.73	213.46	220.30	213.81	216.16	0.13 (0.15)	13.34 (0.004)
$Kw - Fr(a, b, \lambda, \delta)$	101.83	211.66	220.78	212.25	215.26	0.09 (0.65)	9.54 (0.008)
$GMO - Fr(\theta, \alpha, \lambda, \delta)$	101.49	210.98	220.10	211.57	214.58	0.09 (0.56)	8.86 (0.01)
$KwMO - Fr(a, b, \alpha, \lambda, \delta)$	100.63	211.26	222.64	212.16	215.79	0.08 (0.70)	7.14 (0.007)
KwGMO $- \mathbf{Fr(a, b, \theta, \alpha, \lambda, \delta)}$	**97.06**	**206.12**	**219.80**	**207.02**	**211.52**	0.07 **(0.79)**	-
$GMOKw - Fr(a, b, \theta, \alpha, \lambda, \delta)$	99.84	211.68	225.36	212.58	217.08	0.09 (0.51)	-

Fig. 7: Plots of the observed histogram and estimated pdf's on the left and observed Ogive and estimated cdf's for the $MO - Fr, Kw - Fr, GMO - Fr, KwMO - Fr, KwGMO - Fr$ and $GMOKw - Fr$ and $GMOKw - Fr$ distributions for data set II

In the Tables 2(a & b), 3(a & b), 4(a & b) and 5(a & b), the MLEs with standard errors of the parameters for all the fitted models along with their AIC, BIC, CAIC, HQIC and KS, LR statistic with *p-value* from the fitting results of the data sets I and II are presented respectively.

From the findings presented in the Tables 2(b), 3(b), 4(b) and 5(b) on the basis of the lowest value different criteria like AIC, BIC, CAIC, HQIC and highest *p-value* of the KS statistics the $KwGMO - G$ is found to be a better model than its sub families and recently introduced $GMOKw - G$ family for all the data sets considered here. As expected the LR test rejects the four sub models in favour of the $KwGMO - G$ family. A comparison of the closeness of the fitted densities with the observed histogram and fitted cdf's with the observed ogive of the data sets I and II are presented in the Figures 4, 5, 6 and 7 respectively. These plots cleary indicate that the distributions from proposed family provide comparatively closer fit to both the data sets.

6 Conclusion

A new extension of the Generalized Marshall-Olkin family of distributions is introduced and some of its important properties are studied. The maximum likelihood method for estimating the parameters is discussed. As many as four applications of real life data fitting shows good results in favour of the distributions from the proposed family when compared to some of its sub models and also distributions belonging to generalized Marshall-Olkin Kumaraswamy family of distributions. Therefore it is expected that the proposed family will be competitive and hence a useful contribution to the existing literature of continuous distributions.

References

[1] Aarset, M.V. (1987). How to identify a bathtub hazard rate. IEEE Transactions on Reliability, 36: 106 - 108.

[2] Alizadeh, M., Tahir, M.H., Cordeiro, G.M., Zubair, M. and Hamedani, G.G. (2015). The Kumaraswamy Marshal-Olkin-G family of distributions. Journal of the Egyptian Mathematical Society, 23: 546 - 557.

[3] Bjerkedal, T. (1960). Acquisition of resistance in Guinea pigs infected with different doses of virulent tubercle bacilli. American Journal of Hygiene, 72: 130 - 148.

[4] Chakraborty, S. and Handique, L. (2017). The generalized Marshall-Olkin-Kumaraswamy-G family of distributions. Journal of Data Science, 15: 391 - 422.

[5] Cordeiro, G.M. and De Castro, M. (2011). A new family of generalized distributions. J. Stat. Comput. Simul., 81: 883 - 893.

[6] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. Water Resour. Res., 15: 1049 - 1054.

[7] Handique, L. and Chakraborty, S. (2017a). A new beta generated Kumaraswamy Marshall-Olkin-G family of distributions with applications. Malaysian Journal of Science, 36: 157 - 174.

[8] Handique, L. and Chakraborty, S. (2017b). The Beta generated Marshall-Olkin Kumaraswamy-G family of distributions with applications. Int. J. Agricult. Stat. Sci., 13: 721 - 733.

[9] Handique, L., Chakraborty, S. and Ali, M.M. (2017a). The Beta-Kumaraswamy-G family of distributions. Pakistan Journal of Statistics, 33: 467 - 490.

[10] Handique, L., Chakraborty, S. and Hamedani, G.G. (2017b). The Marshall-Olkin-Kumaraswamy-G family of distributions. Journal of Statistical Theory and Applications, 16: 427 - 447.

[11] Handique, L., Chakraborty, S. and Thiago, A.N. (2018). The exponentiated generalized Marshall-Olkin family of distributions: Its properties and applications. *Annals of Data Science, 1-21* .

[12] Jayakumar, K. and Mathew, T. (2008). On a generalization to Marshall-Olkin scheme and its application to Burr type XII distribution. Stat. Pap., 49: 421 - 439.

[13] Marshall, A. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. Biometrika, 84: 641 - 652.

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

[15] Shibu, D.S. and Irshad, M.R. (2016). Extended new generalized Lindley distribution. Statistica, anno LXXVI, 1: 41 - 56.

[16] Tahir, M.H., Cordeiro, G.M., Mansoor, M., Alzaatreh, A., and Zubair, M. (2018). A new generalized family of distributions from bounded support. Journal of Data Science, 16: 251 - 276.

[17] Yousof, H.M., Alizadeh, M., Jahanshahi, S.M.A., Ramiresd, T.G, Ghoshe, I. and Hamedani, G.G (2017). The transmuted Topp-Leone G family of distributions: Theory, characterizations and applications. Journal of Data Science, 15: 723 - 740.