



A Family of Extended Half-Distributions: Theory and Applications

Hassan S. Bakouch^a

^aDepartment of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt

Abstract. In this paper, we introduce a new family of distributions which extends several new half-distributions. Also, the family can be viewed as a general weighted family of distributions. Some general mathematical properties of the family are obtained, involving the quantile function, (conditional) moments, moment generating function, entropies, order statistics, record values and a bivariate extension of the family. Different estimation procedures for the family parameters are discussed. Some sub-models of the family that can accommodate various shapes for the hazard rate and density functions are given. Using two reliability data sets, the potentiality of proposed sub-models of the family is shown under the estimation procedures.

1. Introduction

During the last two decades or more, various approaches of generating new families of distributions are introduced for increasing chances of modeling practical data that come from a wide variety of disciplines. Among such families we can mention: The Marshal-Olkin-G (MO-G) by [11], the Exponential-G by [10], the beta generator (beta-G) family by [8], the gamma-G (type 1) by [15], the gamma-G (type 2) by [14], the Transformed-Transformer (T-X) by [2], the odd-Burr G by [1] and the Kumaraswamy odd Burr G by [13]. [3] introduced a new extended odd family of probability distributions with a study to a sub-model of the family. [5] proposed a new class of trigonometric distributions based on a cosine-sine transformation. Very recently, [6] introduced a new generator of distributions based on a polynomial-exponential transformation of an existing cumulative distribution function.

In this paper, we introduce a new family of distributions which extends several new families of half-distributions. Also, the family can be viewed as a general weighted family of distributions. Interpretations and motivations of the introduced family will be shown in the next lines of this section.

Let $\alpha > 0$ and $G(x)$ be a base cumulative distribution function (cdf) and $g(x)$ be the associated probability density function (pdf). We consider the new cdf

$$F(x) = \frac{\alpha}{\alpha + 1 - G(x)} G(x), \quad x \in \mathbb{R}, \quad \alpha > 0, \quad (1)$$

and its corresponding pdf is given by

$$f(x) = \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G(x))^2} g(x). \quad (2)$$

2010 *Mathematics Subject Classification.* 60E05; 62E15

Keywords. General family of distributions, Weighted distribution, Distribution on a unit-interval, Estimation methods, Reliability data.

Received: 06 May 2019; Revised: 04 August 2019; Accepted: 27 August 2019

Communicated by Biljana Popović

Email address: hassan.bakouch@science.tanta.edu.eg (Hassan S. Bakouch)

Denoting $S_F(x) = 1 - F(x)$ to be the survival function (sf) associated to $F(x)$ and $S_G(x)$ is the one associated to $G(x)$, then we have

$$S_F(x) = \frac{\alpha + 1}{\alpha + 1 - G(x)} S_G(x). \tag{3}$$

Let $h_F(x) = \frac{f(x)}{S(x)}$ denote the hazard rate function (hrf) associated to $F(x)$ and $h_G(x)$ be the one associated to $G(x)$, we have

$$h_F(x) = \frac{\alpha}{\alpha + 1 - G(x)} h_G(x). \tag{4}$$

Let us now present some special examples for $\alpha = 1$.

Example 1. Let $G(x)$ be the cdf associated to the exponential distribution of parameter $\lambda > 0$, then the cdf $F(x)$ given by (1), with $\alpha = 1$, is

$$F(x) = \frac{1 - e^{-\lambda x}}{1 + e^{-\lambda x}}, \quad x > 0,$$

and its corresponding pdf is

$$f(x) = \frac{2\lambda e^{-\lambda x}}{(1 + e^{-\lambda x})^2}, \quad x > 0.$$

In literature, it corresponds to a pdf of the so-called half-logistic distribution that having a major role in reliability and survival analysis. The half-logistic distribution can be viewed as a weighted exponential distribution with the weight function $W(x) = \frac{1}{(1+e^{-\lambda x})^2}$ with expectation $E[W(X)] = \frac{1}{2}$.

Example 2. Let $G(x)$ be the cdf associated to the Lomax distribution given by

$$G(x) = 1 - (1 + \lambda x)^{-\alpha}, \quad x > 0,$$

then the cdf $F(x)$ corresponding to (1), with $\alpha = 1$, is

$$F(x) = \frac{1 - (1 + \lambda x)^{-\alpha}}{1 + (1 + \lambda x)^{-\alpha}} = \frac{(1 + \lambda x)^\alpha - 1}{(1 + \lambda x)^\alpha + 1}, \quad x > 0.$$

A corresponding pdf is given by

$$f(x) = \frac{2\lambda\alpha (1 + \lambda x)^{-(\alpha+1)}}{\left(1 + (1 + \lambda x)^{-\alpha}\right)^2}, \quad x > 0. \tag{5}$$

Note that it is a weighted Lomax distribution with weight function $W(x) = \frac{1}{(1+(1+\lambda x)^{-\alpha})^2}$ and $E[W(X)] = \frac{1}{2}$, therefore, it may be a half of any literature distribution which is not known to us.

Remark 1. Based on the two examples above and equation (2), we can get new family of half-distributions to any common $g(x)$ and $G(x)$ using the weight function $W(x) = \frac{1}{(2-G(x))^2}$ and $E[W(X)] = \frac{1}{2}$. Concept of weighted distribution is of a major importance in practical situations as it is not usual that the data are not equally represented.

From the earlier discussions, we shall denote the proposed family given by (2) as family of extended half-distributions. Other important motivations of this family are as follows.

- The family defined by (2) can be viewed as a general weighted family of distributions to a common base $g(x)$ with weight $W(x) = \frac{\alpha}{(\alpha+1-G(x))^2}$ and $E[W(X)] = \frac{1}{\alpha+1}$.
- In reliability, a parallel system fails when all components fail and the lifetime of this system is determined by $Z_M = \max(X_1, \dots, X_M)$, where X_1, \dots, X_M are stochastically independent lifetime components. Assume that X_1, \dots, X_M follow a common base cdf $G(x)$ and the number of components M is random and following a modified geometric distribution, with parameter $\frac{\alpha}{\alpha+1}$, defined by

$$P(M = n) = \frac{\alpha}{(\alpha + 1)^n}, \quad n = 1, 2, \dots, \alpha > 0,$$

hence, the cdf of Z_M is the cdf of the proposed family given by (2).

- Let $\frac{\alpha}{1+\alpha} = p \in (0, 1)$, hence equation (1) is reduced to the restricted complementary G-geometric (CGG) family proposed by [4] as the parameter p belongs to a unit interval.
- Sub-models of the family can accommodate various shapes for the hazard rate and density functions (see Section 4 of this paper) which reflect potentiality of the family for modeling many practical data. Also, a new bounded support distribution (which is rare in literature) is introduced as a sub-model of the family beside many of unbounded ones. Moreover, it is shown that the applied estimation procedures of those sub-models do not involve complexity in computations and working well under two reliability data sets.

The rest of the paper is structured as follows. In Section 2 we obtain some general mathematical properties of the family, such as quantile function, moments, conditional moments, moment generating function, entropies, order statistics, record values and a bivariate extension of the family. Some estimation procedures for the family parameters are discussed in Section 3, namely the maximum likelihood estimation, ordinary and weighted least square estimations, and Cramér-von Mises estimation. Section 4 mentions some sub-models of the family that can accommodate various shapes for the hazard rate and density functions. Section 5 shows the potentiality of proposed sub-models of the family for the earlier estimation procedures under two reliability data sets. Some concluding remarks and future work are summarized in Section 6.

2. Some mathematical properties of the family

In this section we get some general mathematical properties of the family, such as quantile function, (conditional) moments, moment generating function, entropies, order statistics, record values and a bivariate extension of the family.

2.1. Quantile function

Let $Q_F(x)$ be the quantile function associated to $F(x)$ defined by (1) and $Q_G(x)$ be the quantile function associated to $G(x)$. Then we have $F(Q_F(x)) = x$, which implies

$$\begin{aligned} \frac{\alpha}{\alpha + 1 - G(Q_F(x))} G(Q_F(x)) &= x, \\ G(Q_F(x)) &= \frac{\alpha + 1}{x + \alpha} x, \end{aligned}$$

hence

$$Q_F(x) = Q_G\left(\frac{\alpha + 1}{x + \alpha} x\right).$$

The considered distribution can be simulated by using $X = Q_F(U)$, where U is a random variable having the uniform distribution on $[0, 1]$.

2.2. Some useful expansions

Here, we provide some useful expansions of $F(x)$ and $f(x)$ in terms of the exponentiated sf $S_G(x)$. For any x such that $G(x) < 1$, using the geometric and binomial series, we have

$$\begin{aligned} F(x) &= \frac{\alpha}{\alpha + 1} G(x) \frac{1}{1 - \frac{1}{\alpha+1} G(x)} = \frac{\alpha}{\alpha + 1} G(x) \sum_{k=0}^{+\infty} \left(\frac{1}{\alpha + 1}\right)^k [G(x)]^k \\ &= \sum_{k=0}^{+\infty} \alpha \left(\frac{1}{\alpha + 1}\right)^{k+1} [G(x)]^{k+1} = \sum_{k=0}^{+\infty} \sum_{\ell=0}^{k+1} A_{k,\ell} [S_G(x)]^\ell, \end{aligned}$$

where

$$A_{k,\ell} = \alpha \binom{k+1}{\ell} (-1)^\ell \left(\frac{1}{\alpha + 1}\right)^{k+1}.$$

Similarly, we have

$$\begin{aligned} f(x) &= \frac{\alpha}{\alpha + 1} g(x) \frac{1}{\left(1 - \frac{1}{\alpha+1} G(x)\right)^2} = \frac{\alpha}{\alpha + 1} g(x) \sum_{k=1}^{+\infty} k \left(\frac{1}{\alpha + 1}\right)^{k-1} [G(x)]^{k-1} \\ &= \sum_{k=1}^{+\infty} \alpha \left(\frac{1}{\alpha + 1}\right)^k k [G(x)]^{k-1} g(x) = \sum_{k=1}^{+\infty} \sum_{\ell=0}^{k-1} B_{k,\ell} [S_G(x)]^\ell g(x), \end{aligned} \tag{6}$$

where

$$B_{k,\ell} = \alpha \binom{k-1}{\ell} (-1)^\ell \left(\frac{1}{\alpha + 1}\right)^k k.$$

2.3. (Conditional) Moments and moment generating function

Here and after, we consider a random variable X following a distribution characterized by the cdf $F(x)$ given by (1).

Many of the interesting features of a family of distributions can be determined by its (conditional) moments. Using the expansion (6), the r -th moment of X is attained as

$$E(X^r) = \int_{-\infty}^{+\infty} x^r f(x) dx = \sum_{k=1}^{+\infty} \sum_{\ell=0}^{k-1} B_{k,\ell} \int_{-\infty}^{+\infty} x^r [S_G(x)]^\ell g(x) dx.$$

The moment generating function of X is given as

$$M(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \sum_{k=1}^{+\infty} \sum_{\ell=0}^{k-1} B_{k,\ell} \int_{-\infty}^{+\infty} e^{tx} [S_G(x)]^\ell g(x) dx,$$

for t such that the integral exists.

Now, we get some conditional moments of X as follows.

First of all, thanks to (6), set

$$J_r(t) = \int_0^t x^r f(x) dx = \sum_{k=1}^{+\infty} \sum_{\ell=0}^{k-1} B_{k,\ell} \int_{-\infty}^t x^r [S_G(x)]^\ell g(x) dx.$$

Hence, the r -th conditional moments of X is given by, for any $t \in \mathbb{R}$,

$$E(X^r | X > t) = \frac{1}{S_F(t)} \int_t^{+\infty} x^r f(x) dx = \frac{1}{S_F(t)} [E(X^r) - J_r(t)].$$

Further, the r -th reversed moments of X having the expression

$$E(X^r | X \leq t) = \frac{1}{F(t)} J_r(t).$$

2.4. Some entropy functions

The entropy of a random variable X with pdf $f(x)$ is a measure of variation of the uncertainty of physical systems. In this section we obtain two popular entropies, namely Shannon entropy and Rényi entropy.

For Shannon entropy, we have

$$H(f) = -E(\log[f(X)]) = - \int_{-\infty}^{+\infty} f(x) \log[f(x)] dx = - \log \left[\frac{\alpha}{\alpha + 1} \right] - E(\log[g(X)]) + \frac{2\alpha}{\alpha + 1} \int_{-\infty}^{+\infty} g(x) \frac{1}{\left(1 - \frac{1}{\alpha+1}G(x)\right)^2} \log \left[1 - \frac{1}{\alpha + 1}G(x) \right] dx.$$

The second term can be calculated according to the definition of $g(x)$ and the inversion formula. Denoting I to be the third and last term, using a direct primitive computation, we have

$$I = 2\alpha \left[\frac{1 + \log \left[1 - \frac{1}{\alpha+1}G(x) \right]}{1 - \frac{1}{\alpha+1}G(x)} \right]_{-\infty}^{+\infty} = 2 + 2(\alpha + 1) \log \left[1 - \frac{1}{\alpha + 1} \right].$$

Let us now focus our attention to the Rényi entropy defined by $J_R(\beta) = \frac{1}{1-\beta} \log \left(\int_{-\infty}^{+\infty} [f(x)]^\beta dx \right)$, where $\beta \neq 1$ and $\beta > 0$. For any x such that $G(x) < 1$, using the binomial series, we have

$$[f(x)]^\beta = \left(\frac{\alpha}{\alpha + 1} \right)^\beta [g(x)]^\beta \frac{1}{\left(1 - \frac{1}{\alpha+1}G(x)\right)^{2\beta}} = \left(\frac{\alpha}{\alpha + 1} \right)^\beta [g(x)]^\beta \sum_{k=0}^{+\infty} \binom{-2\beta}{k} \left(-\frac{1}{\alpha + 1} \right)^k [G(x)]^k.$$

Assuming that $\int_{-\infty}^{+\infty} [g(x)]^\beta [G(x)]^k dx$ exists, we have

$$\int_{-\infty}^{+\infty} [f(x)]^\beta dx = \sum_{k=0}^{\infty} C_k \int_{-\infty}^{+\infty} [g(x)]^\beta [G(x)]^k dx,$$

where

$$C_k = \left(\frac{\alpha}{\alpha + 1} \right)^\beta \binom{-2\beta}{k} \left(-\frac{1}{\alpha + 1} \right)^k.$$

Therefore, we get

$$J_R(\beta) = \frac{1}{1-\beta} \log \left(\int_{-\infty}^{+\infty} [f(x)]^\beta dx \right) = \frac{1}{1-\beta} \log \left[\sum_{k=0}^{\infty} C_k \int_{-\infty}^{+\infty} [g(x)]^\beta [G(x)]^k dx \right].$$

2.5. Order statistics and record values

Order statistics and record values have a major role in statistics, in general, and in reliability and life testing, in particular. Let X_1, X_2, \dots, X_n be n i.i.d. random variables having the pdf $f(x)$ defined by (2). Let us consider its order statistics to be $X_{1:n}, X_{2:n}, \dots, X_{n:n}$, then the pdf of the i -th order statistic $X_{i:n}$ is given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} \left[\frac{\alpha}{\alpha + 1 - G(x)} G(x) \right]^{i-1} \left[\frac{\alpha + 1}{\alpha + 1 - G(x)} S_G(x) \right]^{n-i} \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G(x))^2} g(x) = \left[\frac{\alpha^i(\alpha + 1)^{n-i+1}}{(\alpha + 1 - G(x))^{n+1}} \right] \frac{n!}{(i-1)!(n-i)!} [G(x)]^{i-1} [S_G(x)]^{n-i} g(x).$$

For n i.i.d. random variables Y_1, Y_2, \dots, Y_n having the pdf $g(x)$ and the pdf of the i -th order statistic $Y_{i:n}$, denoted by $g_{i:n}(x)$, we can write

$$f_{i:n}(x) = \left[\frac{\alpha^i(\alpha + 1)^{n-i+1}}{(\alpha + 1 - G(x))^{n+1}} \right] g_{i:n}(x).$$

From this expression, for any $r > 0$, the r -th moment of $X_{i:n}$ is given by

$$\mathbb{E}(X_{i:n}^r) = \mathbb{E}\left(Y_{i:n}^r \left(\frac{\alpha^i(\alpha + 1)^{n-i+1}}{(\alpha + 1 - G(Y_{i:n}))^{n+1}}\right)\right).$$

Now, move to the record values of the proposed family. Let X_1, X_2, \dots , be a sequence of i.i.d. random variables having the cdf $F(x)$ given by (1). We define a sequence of record times $U(n)$ as follows: $U(1) = 1$, $U(n) = \min\{j; j > U(n - 1), X_j > X_{U(n-1)}\}$ for $n \geq 2$. We define the i -th upper record value by $R_i = X_{U(i)}$, with $R_1 = X_1$. Using the pdf $f(x)$ defined by (2) and the sf $S_F(x)$ having (3), the pdf of R_i is given by

$$f_{R_i}(x) = \frac{[-\log(S_F(x))]^{i-1}}{(i - 1)!} f(x) = \frac{1}{(i - 1)!} \left[-\log\left(\frac{\alpha + 1}{\alpha + 1 - G(x)} S_G(x)\right)\right]^{i-1} \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G(x))^2} g(x).$$

Using the pdf $f(x)$ defined by (2) and the hrf $h_F(x)$ having (4), the joint pdf of (R_1, \dots, R_n) is given as

$$f_{(R_1, \dots, R_n)}(x_1, \dots, x_n) = f(x_n) \prod_{k=1}^{n-1} h_F(x_k) = \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G(x_n))^2} g(x_n) \prod_{k=1}^{n-1} \frac{\alpha}{\alpha + 1 - G(x_k)} h_G(x_k),$$

for $x_1 < \dots < x_n$.

2.6. Bivariate extension of the family

We now propose a bivariate version of the proposed family. Let (X, Y) be a bivariate random variable with the joint cdf

$$F_{X,Y}(x, y) = \frac{\alpha}{\alpha + 1 - G(x, y)} G(x, y), \quad (x, y) \in \mathbb{R}^2,$$

where $G(x, y)$ is denoted as a bivariate base cdf. Let $G_1(x)$ and $G_2(y)$ be the corresponding marginal base cdfs, $g_1(x)$ and $g_2(y)$ be the corresponding marginal pdfs and $g(x, y)$ be the corresponding bivariate pdf. Then the marginal cdfs of X and Y are given by

$$F_X(x) = \frac{\alpha}{\alpha + 1 - G_1(x)} G_1(x)$$

and

$$F_Y(y) = \frac{\alpha}{\alpha + 1 - G_2(y)} G_2(y).$$

Whereas, the marginal pdfs of (X, Y) are given by

$$f_X(x) = \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G_1(x))^2} g_1(x)$$

and

$$f_Y(y) = \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G_2(y))^2} g_2(y).$$

Moreover, the pdf of (X, Y) is given by

$$f(x, y) = \frac{\alpha(\alpha + 1)}{(\alpha + 1 - G(x, y))^2} \theta(x, y),$$

where

$$\theta(x, y) = g(x, y) + \frac{2}{\alpha + 1 - G(x, y)} \frac{\partial G(x, y)}{\partial x} \frac{\partial G(x, y)}{\partial y}.$$

3. Estimation procedures for the family parameters

In this section we give some estimation procedures for the family parameters, as many of the family features depend on the parameters, namely the maximum likelihood estimation, ordinary and weighted least square estimations, and Cramér-von Mises estimation. Those procedures are compared numerically by using some goodness of fit statistics for two practical data sets.

3.1. Maximum likelihood estimators

Let X_1, X_2, \dots, X_n be a random sample from the family characterized by (2) with parameter vector Θ and x_1, x_2, \dots, x_n are the corresponding observed values, then the likelihood function is given by

$$L(\Theta) = \prod_{i=1}^n f(x_i) = \frac{\alpha^n (\alpha + 1)^n}{\prod_{i=1}^n (\alpha + 1 - G(x_i))^2} \prod_{i=1}^n g(x_i).$$

The log-likelihood function can be expressed as

$$\ell(\Theta) = \log(L(\theta)) = n \log(\alpha) + n \log(\alpha + 1) - 2 \sum_{i=1}^n \log[\alpha + 1 - G(x_i)] + \sum_{i=1}^n \log(g(x_i)).$$

The nonlinear log-likelihood equations given by $\frac{\partial \ell(\Theta)}{\partial \Theta} = 0$ can be solved numerically for obtaining the maximum likelihood estimators (MLEs). Issue of unimodalness of the log-likelihood function (unique solution in the parameters) of some sub-models of the family is discussed visually under two practical data sets in Section 5.

3.2. Ordinary least-square estimators

Let x_1, x_2, \dots, x_n be an ordered sample of the random sample of size n from the family characterized by the cdf $F(x)$ defined by (1). Then, we have $E(F(X_{i:n})) = \frac{i}{n+1}$ and the least square estimators (LSE) can be obtained by minimizing

$$M(\Theta) = \sum_{i=1}^n \{F(x_i) - E(F(X_{i:n}))\}^2 = \sum_{i=1}^n \left\{ \frac{\alpha}{\alpha + 1 - G(x_i)} G(x_i) - \frac{i}{n+1} \right\}^2,$$

with respect to the unknown parameters of the family, and hence the LSEs are solutions of the nonlinear equations $\frac{\partial M(\Theta)}{\partial \Theta} = 0$.

3.3. Weighted least-square estimators

Based on the LSEs, the weighted least square estimators (WLSE) can be obtained by minimizing the expression

$$W(\Theta) = \sum_{i=1}^n w_i \left\{ \frac{\alpha}{\alpha + 1 - G(x_i)} G(x_i) - \frac{i}{n+1} \right\}^2,$$

with respect to the unknown parameters of the proposed family, where

$$w_i = \frac{1}{\text{Var}(F(X_{i:n}))} = \frac{(n+1)^2(n+2)}{i(n-i+1)}.$$

Hence, solving the nonlinear equations $\frac{\partial W(\Theta)}{\partial \Theta} = 0$ gives the WLSEs of the family parameters.

3.4. Cramér-von Mises estimators

As a type of minimum distance estimators, let us consider the Cramér-von Mises (CVM) estimation method. Assume that x_1, x_2, \dots, x_n are the observed values from the proposed family, in increasing order. Hence, the CVM estimators can be obtained by minimizing

$$C(\Theta) = \frac{1}{12n} + \sum_{i=1}^n \left\{ \frac{\alpha}{\alpha + 1 - G(x_i)} G(x_i) - \frac{2i - 1}{2n} \right\}^2$$

with respect to the unknown parameters of the proposed family.

4. Some sub-models of the family

In this section, we mention some sub-models of the family that reflect the potentiality of the family for modeling practical data, as such models can accommodate various shapes for the hazard rate and density functions, see the displayed plots in this section. Moreover, the estimation procedures are compared under those models.

4.1. Extended half-Burr XII Distribution

The Burr XII (BXII) distribution has, respectively, the pdf and cdf : $g(x) = \frac{ckx^{c-1}}{(x^c+1)^{k+1}}$ and $G(x) = 1 - (1 + x^c)^{-k}$, $k, c, x, > 0$, hence the cdf and pdf of the extended half-Burr XII (EHBXII) distribution are given by

$$F(x) = \frac{\alpha(1 - (x^c + 1)^{-k})}{\alpha + (x^c + 1)^{-k}}, \quad x, \alpha, c, k > 0$$

and

$$f(x) = \frac{\alpha(\alpha + 1)ckx^{c-1}}{(x^c + 1)^{k+1} (\alpha + (x^c + 1)^{-k})^2},$$

respectively. Using these functions, we immediately obtain the expression of the hrf. Figure 1 shows some plots of pdfs and hrfs for the EHBXII distribution with arbitrary choices for the parameters (c, k, α).

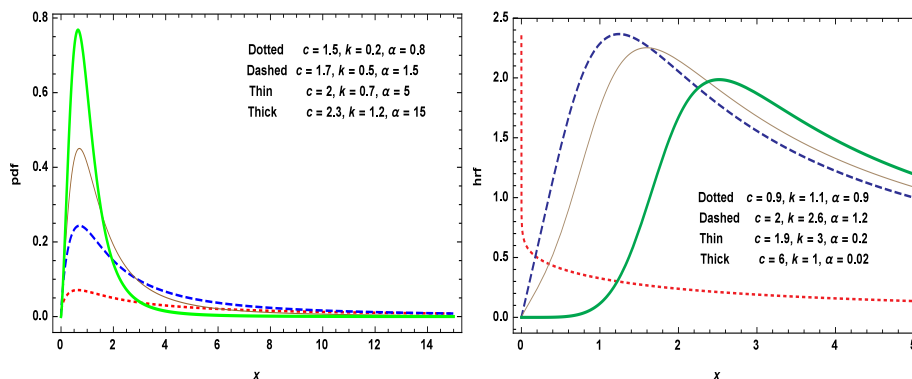


Figure 1: Plots of the pdf and hrf of the EHBXII distribution for arbitrary parameter choices.

4.2. Extended half-Weibull Distribution

Since, the Weibull distribution having the pdf and cdf given, respectively, by $g(x) = \frac{k(\frac{x}{\lambda})^{k-1} e^{-(\frac{x}{\lambda})^k}}{\lambda}$ and $G(x) = 1 - e^{-(\frac{x}{\lambda})^k}$, $k, \lambda, x, > 0$, then the cdf and pdf of extended half-Weibull (EHW) distribution are given by

$$F(x) = \frac{\alpha \left(1 - e^{-(\frac{x}{\lambda})^k}\right)}{\alpha + e^{-(\frac{x}{\lambda})^k}}, \quad x, \alpha, \lambda, k > 0$$

and

$$f(x) = \frac{\alpha(\alpha + 1)k \left(\frac{x}{\lambda}\right)^{k-1} e^{-(\frac{x}{\lambda})^k}}{\lambda \left(\alpha + e^{-(\frac{x}{\lambda})^k}\right)^2},$$

respectively. Making use of the functions above, the expression of the hrf follows immediately. Figure 2 shows some plots of pdfs and hrfs for the EHW distribution with different choices for the parameters (λ, k, α) .

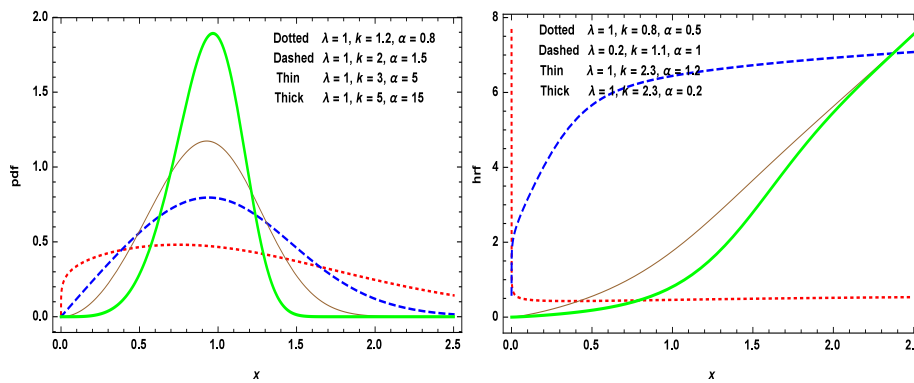


Figure 2: Plots of the pdf and hrf of the extended half-Weibull distribution for arbitrary parameter choices.

4.3. Extended half-Power Lindley Distribution

Since, the power Lindley (PL) distribution ([9]) having the pdf and cdf given, respectively, by $g(x) = \frac{\beta^2 \lambda (x^\lambda + 1) x^{\lambda-1} e^{-\beta x^\lambda}}{\beta + 1}$ and $G(x) = 1 - \left(\frac{\beta x^\lambda}{\beta + 1} + 1\right) e^{-\beta x^\lambda}$, $\beta, \lambda, x, > 0$, hence the cdf and pdf of the extended half-power Lindley (EHPL) distribution are given by

$$F(x) = \frac{\alpha \left(1 - \left(\frac{\beta x^\lambda}{\beta + 1} + 1\right) e^{-\beta x^\lambda}\right)}{\alpha + e^{-\beta x^\lambda} \left(\frac{\beta x^\lambda}{\beta + 1} + 1\right)}, \quad x, \alpha, \lambda, \beta > 0$$

and

$$f(x) = \frac{\alpha(\alpha + 1)\beta^2 \lambda (x^\lambda + 1) x^{\lambda-1} e^{-\beta x^\lambda}}{(\beta + 1) \left(\alpha + e^{-\beta x^\lambda} \left(\frac{\beta x^\lambda}{\beta + 1} + 1\right)\right)^2},$$

respectively. Using the two preceding functions, the hrf can be easily found. Figure 3 shows some plots of pdfs and hrfs for the EHPL distribution with different choices for the parameters (λ, β, α) .

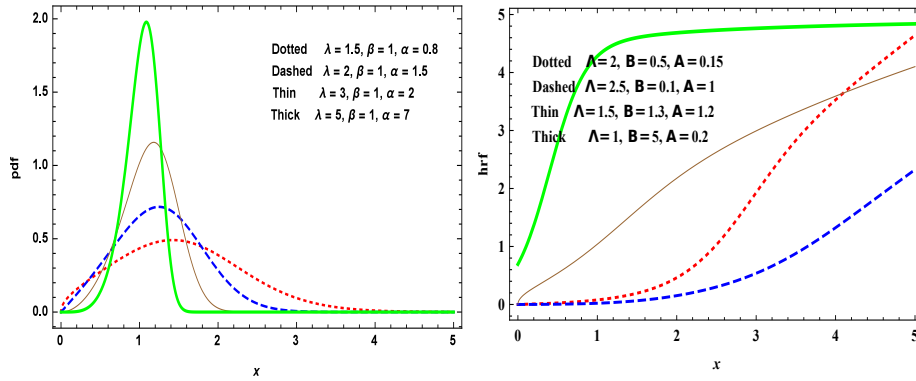


Figure 3: Plots of the pdf and hrf of the extended half-power Lindley distribution for arbitrary parameter choices.

4.4. Extended half-Kumaraswamy Distribution

In this subsection, we exhibit a new distribution defined on the unit-interval $(0, 1)$, which is useful for modeling proportional practical data.

The Kumaraswamy (Kw) distribution has the pdf and cdf given, respectively, as $g(x) = abx^{a-1} (1 - x^a)^{b-1}$ and $G(x) = 1 - (1 - x^a)^b$, $x \in (0, 1)$, $a, b > 0$, then the cdf and pdf of extended half-Kumaraswamy (EHKw) distribution are given by

$$F(x) = \frac{\alpha(1 - (1 - x^a)^b)}{(1 - x^a)^b + \alpha}, \quad x \in (0, 1), \alpha, a, b > 0$$

and

$$f(x) = \frac{\alpha(\alpha + 1)abx^{a-1} (1 - x^a)^{b-1}}{((1 - x^a)^b + \alpha)^2},$$

respectively. The hrf of the EHKw distribution follows using the two equations above. Some plots of the pdfs and hrfs for the EHKw distribution are displayed by Figure 4 for some arbitrary choices of the parameters (a, b, α) .

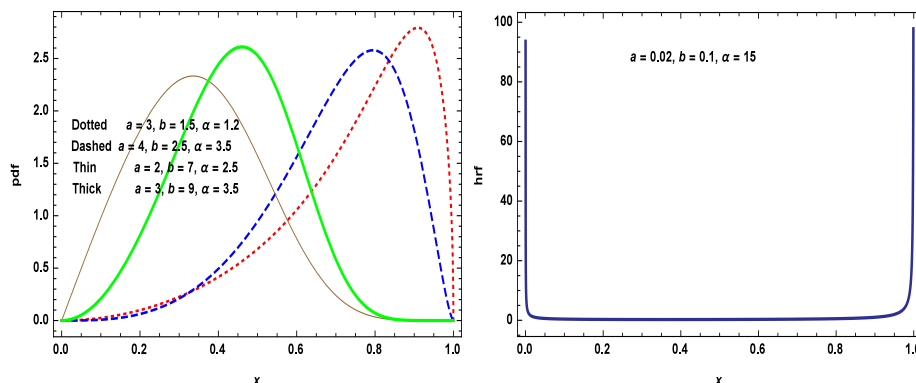


Figure 4: Plots of the pdf and hrf of the extended half-Kumaraswamy distribution for arbitrary parameter choices.

Table 1: Descriptive statistics of the data sets

| Data set 1 | | | | | | | |
|------------|--------|--------|--------|----------|----------|--------|------|
| n | Mean | Median | SD | Skewness | Kurtosis | M1 | M2 |
| 50 | 0.1632 | 0.16 | 0.0811 | 0.0723 | 2.2166 | 0.0653 | 0.07 |
| Data set 2 | | | | | | | |
| 20 | 10.65 | 11.5 | 4.2831 | -0.3588 | 2.4969 | 3.385 | 2.5 |

SD = Standard Deviation, M1 = Mean deviation about the mean,
M2 = Mean deviation about the median

5. Illustrative practical data with analysis

In this section, the potentiality of the proposed family is highlighted by comparing four sub-models of this family, namely EHBXII, EHW, EHPL and EHKw to the corresponding base distributions (BXII, W, PL and Kw) under all the preceding estimation procedures: MLE, LSE, WLSE and CVME, for two reliability data sets. Descriptive statistics of both data sets are summarized in Table 1 which indicates that the first one is under-dispersed with positive skewness while the second data set is over-dispersed with negative skewness. Moreover, the mean deviation about the mean and median is small for the first data set, and it is large for the second data set.

The description of the considered data sets is as follows.

Data set 1. The first data set refers to the 50 observations on Burr (in the millimeter unit) with hole diameter 12 mm and sheet thickness 3.15 mm reported by [7]. The data values are: 0.04, 0.02, 0.06, 0.12, 0.14, 0.08, 0.22, 0.12, 0.08, 0.26, 0.24, 0.04, 0.14, 0.16, 0.08, 0.26, 0.32, 0.28, 0.14, 0.16, 0.24, 0.22, 0.12, 0.18, 0.24, 0.32, 0.16, 0.14, 0.08, 0.16, 0.24, 0.16, 0.32, 0.18, 0.24, 0.22, 0.16, 0.12, 0.24, 0.06, 0.02, 0.18, 0.22, 0.14, 0.06, 0.04, 0.14, 0.26, 0.18, 0.16.

Data set 2. The second data set represents the failure times of 20 components. The data set can be found in [12, Page 245] and its values are: 2, 3, 6, 6, 7, 9, 9, 10, 10, 11, 12, 12, 12, 13, 13, 13, 15, 16, 16, 18.

We check the adequacy of the fitted models, under all the estimation procedures, via the statistics of Anderson-Darling (A^*) and the Cramér-von Mises (W^*). They allow to determine how closely a specific distribution fits the associated empirical distribution for a given data set. The smaller statistics give the better fit. The obtained results are presented in Tables 2 and 3 for both data sets. Based on values of A^* and W^* in the preceding tables, we conclude that the considered sub-models of the family are highly competitor to the compared distributions under all the estimation procedures discussed in Section 3. Also, it can be noted that the EHKw distribution has a superior performance for the first data set, while the EHW distribution represents the best fit for the second data set, among the compared distributions. Moreover, all the estimation procedures do not involve complexity in computations and working well for both data sets, although the WLSE procedure is recommended more for the first data set. The results on estimation procedures are displayed visually by Figure 5 where the estimated densities of the superior distribution are superimposed on the histogram for both data sets based on the MLE, LSE, WLSE and CVME methods.

Now, we get the confidence intervals for parameters of the EHKw and EHW distributions under the MLEs for both data sets. The variance-covariance matrix of the MLEs of the EHKw distribution for data set 1 is given as

$$\begin{pmatrix} 0.291921 & 6.197111 & 0.370904 \\ 6.197112 & 156.535541 & 6.588642 \\ 0.370904 & 6.588641 & 0.562065 \end{pmatrix},$$

The variance-covariance matrix of the MLEs of the EHW distribution for data set 2 is

$$\begin{pmatrix} 18.875512 & 3.832981 & 1.745377 \\ 3.832986 & 0.832217 & 0.341986 \\ 1.745364 & 0.341986 & 0.172114 \end{pmatrix}.$$

Table 2: Comparison of fit of distributions (SE in parentheses) for Data set 1

| | | MLEs | | |
|----------------------------------|------------------------|--------------------------|------------------------|----------------------|
| Distributions | Estimates | | A* | W* |
| BXII(c, k) | 2.154673 (0.240272) | 39.124621 (15.014217) | | 0.747401 0.109825 |
| Weibull(λ, k) | 0.183743 (0.012843) | 2.119518 (0.246304) | | 0.719622 0.107633 |
| PL(λ, α) | 2.118267 (0.246809) | 37.147585 (14.489925) | | 0.718659 0.107563 |
| Kw(a, b) | 2.077396 (0.254849) | 33.137430 (13.921596) | | 0.688407 0.105252 |
| EHBXII(c, k, α) | 1.708924 (0.478488) | 33.154991 (12.186667) | 0.468754 (0.790817) | 0.592883 0.086292 |
| EHW(λ, k, α) | 0.129401 (0.053890) | 1.655082 (0.499366) | 0.460299 (0.779068) | 0.562626 0.083877 |
| EHPL(λ, β, α) | 1.652886 (0.500476) | 30.340066 (12.231523) | 0.459622 (0.777884) | 0.561481 0.083780 |
| EHkw(a, b, α) | 1.577041 (0.540297) | 25.179850 (12.511403) | 0.438597 (0.749710) | 0.525271 0.080833 |
| | | LSEs | | |
| Distributions | Estimates | | A* | W* |
| BXII(c, k) | 2.042318 | 29.929841 | | 0.678865 0.089502 |
| Weibull(λ, k) | 0.191141 | 2.008349 | | 0.654853 0.088067 |
| PL(λ, β) | 2.007059 | 28.639681 | | 0.654077 0.088026 |
| Kw(a, b) | 1.968748 | 25.410345 | | 0.628029 0.086481 |
| EHBXII(c, k, α) | 1.268607 | 22.259999 | 0.165237 | 0.529712 0.077199 |
| EHW(λ, k, α) | 0.081518 | 1.160859 | 0.142028 | 0.507214 0.075377 |
| EHPL(λ, β, α) | 1.155251 | 19.123569 | 0.140864 | 0.506426 0.075309 |
| EHkw(a, b, α) | 0.925545 | 12.597580 | 0.089717 | 0.482292 0.073169 |
| | | WLSEs | | |
| Distributions | Estimates | | A* | W* |
| BXII(c, k) | 2.088417 | 33.397569 | | 0.663788 0.091168 |
| Weibull(λ, k) | 0.188039 | 2.050470 | | 0.638934 0.089622 |
| PL(λ, β) | 2.049122 | 31.651971 | | 0.638152 0.089579 |
| Kw(a, b) | 2.005663 | 27.929922 | | 0.611548 0.087934 |
| EHBXII(c, k, α) | 1.371475 | 24.763653 | 0.207161 | 0.507336 0.077168 |
| EHW(λ, k, α) | 0.092639 | 1.274044 | 0.184110 | 0.483469 0.074996 |
| EHPL(λ, β, α) | 1.269374 | 21.497754 | 0.182989 | 0.482619 0.074914 |
| EHkw(a, b, α) | 1.092496 | 15.243087 | 0.137888 | 0.455857 0.072300 |
| | | CVMs | | |
| Distributions | Estimates | | A* | W* |
| BXII(c, k) | 2.103836 | 33.349682 | | 0.686005 0.088146 |
| Weibull(λ, k) | 0.190390 | 2.072067 | | 0.661431 0.086699 |
| PL(λ, β) | 2.070984 | 31.986544 | | 0.660673 0.086656 |
| Kw(a, b) | 2.035469 | 28.674555 | | 0.634133 0.085095 |
| EHBXII(c, k, α) | 1.323646 | 23.865052 | 0.176938 | 0.513039 0.075880 |
| EHW(λ, k, α) | 0.086384 | 1.220344 | 0.153920 | 0.488488 0.074048 |
| EHPL(λ, β, α) | 1.215197 | 20.629453 | 0.152774 | 0.487609 0.073979 |
| EHkw(a, b, α) | 1.009006 | 14.115457 | 0.104985 | 0.460062 0.071810 |

Table 3: Comparison of fit of distributions for Data set 2

| MLEs | | | | | |
|----------------------------------|-------------------------|------------------------|------------------------|----------|----------|
| Distributions | Estimates | | | A* | W* |
| BXII(c, k) | 24.657543 (2692.603) | 0.018019 (1.967765) | | 5.797948 | 1.188298 |
| Weibull(λ, k) | 11.925657 (0.989351) | 2.812939 (0.524042) | | 0.418682 | 0.058246 |
| PL(λ, β) | 1.796760 (0.302042) | 0.025325 (0.019590) | | 0.541847 | 0.080353 |
| Kw(a, b) | – (–) | – (–) | – | – | – |
| EHBXII(c, k, α) | 4.213412 (5.355854) | 0.839190 (1.071301) | 0.000258 (0.000438) | 0.785025 | 0.090819 |
| EHW(λ, k, α) | 7.620245 (4.344589) | 1.905639 (0.912259) | 0.189626 (0.414867) | 0.267251 | 0.033943 |
| EHPL(λ, β, α) | 1.293935 (0.465332) | 0.176706 (0.269802) | 0.108617 (0.216541) | 0.290068 | 0.036230 |
| EHkw(a, b, α) | – (–) | – (–) | – (–) | – | – |
| LSEs | | | | | |
| Distributions | Estimates | | | A* | W* |
| BXII(c, k) | 28.309519 | 0.011539 | | 4.772040 | 0.934443 |
| Weibull(λ, k) | 12.417532 | 2.754084 | | 0.395595 | 0.043343 |
| PL(λ, β) | 1.839046 | 0.020804 | | 0.513021 | 0.050059 |
| Kw(a, b) | – | – | – | – | – |
| EHBXII(c, k, α) | 48.921365 | 0.081435 | 0.000079 | 0.873770 | 0.072056 |
| EHW(λ, k, α) | 6.113052 | 1.502687 | 0.110673 | 0.260659 | 0.035696 |
| EHPL(λ, β, α) | 1.101113 | 0.297046 | 0.077560 | 0.284118 | 0.037649 |
| EHkw(a, b, α) | – | – | – | – | – |
| WLSEs | | | | | |
| Distributions | Estimates | | | A* | W* |
| BXII(c, k) | 16.678209 | 0.022004 | | 4.91151 | 0.968509 |
| Weibull(λ, k) | 12.310095 | 2.721369 | | 0.380541 | 0.044386 |
| PL(λ, β) | 1.840101 | 0.021202 | | 0.491824 | 0.050620 |
| Kw(a, b) | – | – | – | – | – |
| EHBXII(c, k, α) | 28.556451 | 0.146676 | 0.000051 | 0.893632 | 0.071966 |
| EHW(λ, k, α) | 5.827897 | 1.454284 | 0.101438 | 0.257153 | 0.036465 |
| EHPL(λ, β, α) | 1.076496 | 0.323793 | 0.071072 | 0.276577 | 0.038212 |
| EHkw(a, b, α) | – | – | – | – | – |
| CVMEs | | | | | |
| Distributions | Estimates | | | A* | W* |
| BXII(c, k) | 28.535507 | 0.011487 | | 4.772192 | 0.934414 |
| Weibull(λ, k) | 12.331378 | 2.984354 | | 0.421755 | 0.040064 |
| PL(λ, β) | 1.988436 | 0.014534 | | 0.568295 | 0.046762 |
| Kw(a, b) | – | – | – | – | – |
| EHBXII(c, k, α) | 441.063161 | 0.009860 | 0.000032 | 0.983033 | 0.068889 |
| EHW(λ, k, α) | 6.674764 | 1.678951 | 0.124111 | 0.250348 | 0.032392 |
| EHPL(λ, β, α) | 1.205356 | 0.228599 | 0.084270 | 0.277204 | 0.034343 |
| EHkw(a, b, α) | – | – | – | – | – |

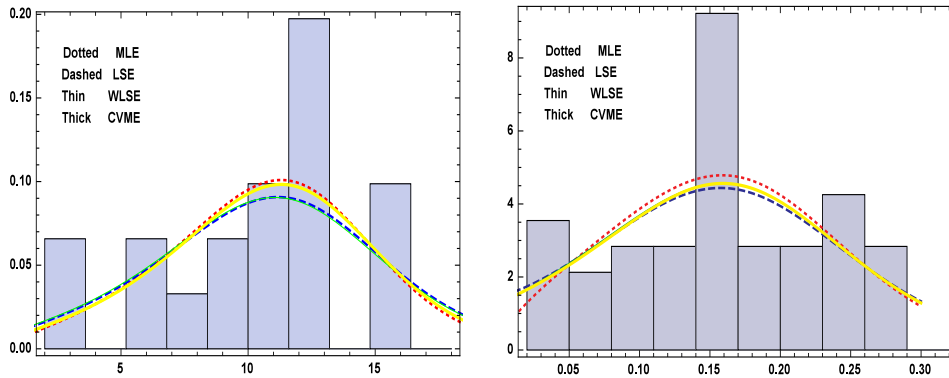


Figure 5: Plots of the pdf superimposed on the histogram for both data sets

Table 4: 95% Confidence intervals

| Data set 1 | EHKw | | |
|---------------------|----------------------|--------------|--|
| a | b | α | |
| [0.518059 2.636023] | [0.651568 49.708132] | [0 1.908028] | |
| Data set 2 | EHW | | |
| λ | k | α | |
| [0 16.135639] | [0.117612 3.693666] | [0 1.002765] | |

Note that the diagonal entries of the variance-covariance matrices for both the data sets are the variances of the MLEs of the parameters and other entries represent the covariances between each two estimates. The two-sided $(1 - \delta)100\%$ confidence interval (CI) for the vector parameter θ of a distribution is $\hat{\theta} \pm z_{\delta/2} \sqrt{Var(\hat{\theta})}$, where $Var(\hat{\theta})$ is a variance of the estimator $\hat{\theta}$ which given by the diagonal entries and z_{δ} is the 100δ percentile of a standard normal distribution.

Using those matrices and the CI, the 95% confidence intervals for the parameters of the EHKw and EHW distributions are summarized in Table 4 for both data sets.

For the MLE, issue of unique solution for the log-likelihood equations of parameters of the EHKw and EHW distributions are justified using the profile log-likelihood functions in Figures 6 and 7 under both data sets.

6. Concluding remarks

In this paper, we proposed a new general family of distributions which extends various half-distributions. A reliability consideration of the family is outlined. Some general mathematical and statistical properties of the family are obtained, such as the moment generating function, entropies, order statistics, record values, a bivariate extension, different estimation procedures. Some sub-models of the family (unbounded and bounded unit-interval continuous distributions) are considered and it is shown that they can accommodate many shapes for the hazard rate and density functions which reflect potentiality of the family for modeling several practical data. Potentiality of the mentioned sub-models for the considered estimation procedures under two reliability data sets is investigated. Finally, sub-models of the family are recommended to analyze the reliability data and hope this can be extended to other applied areas.

A future research will focus on an extensive study to one sub-model of the family on different estimation methods, including the ones discussed in this paper, moments and Bayesian estimation. Hence, we can easily evaluate the performance of the estimation methods using a simulation study as suggested by the referee.

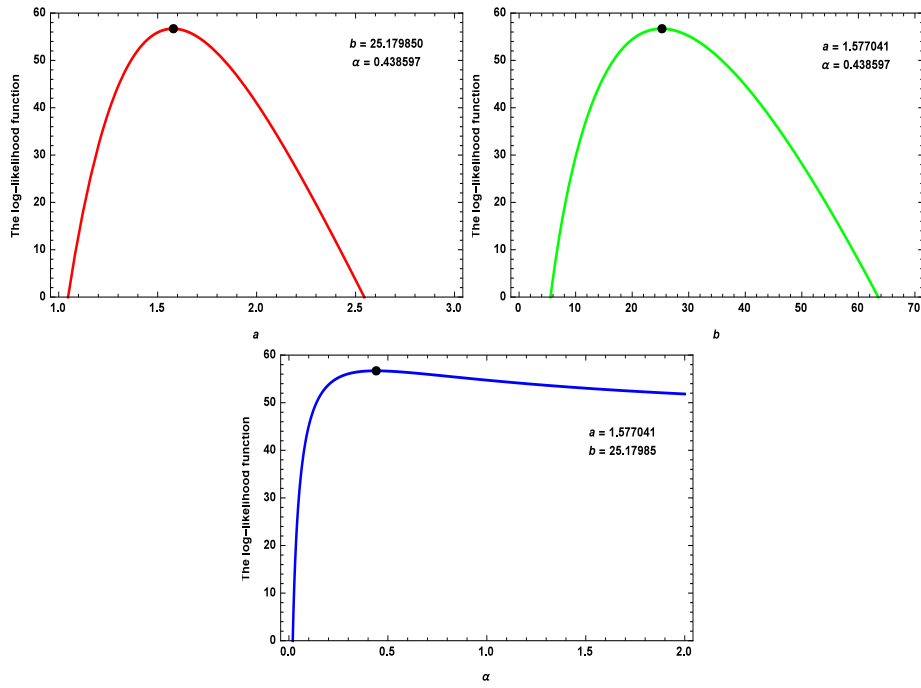


Figure 6: The log-likelihood as a function of different parameters of the extended half-Kumaraswamy distribution for Data set 1.

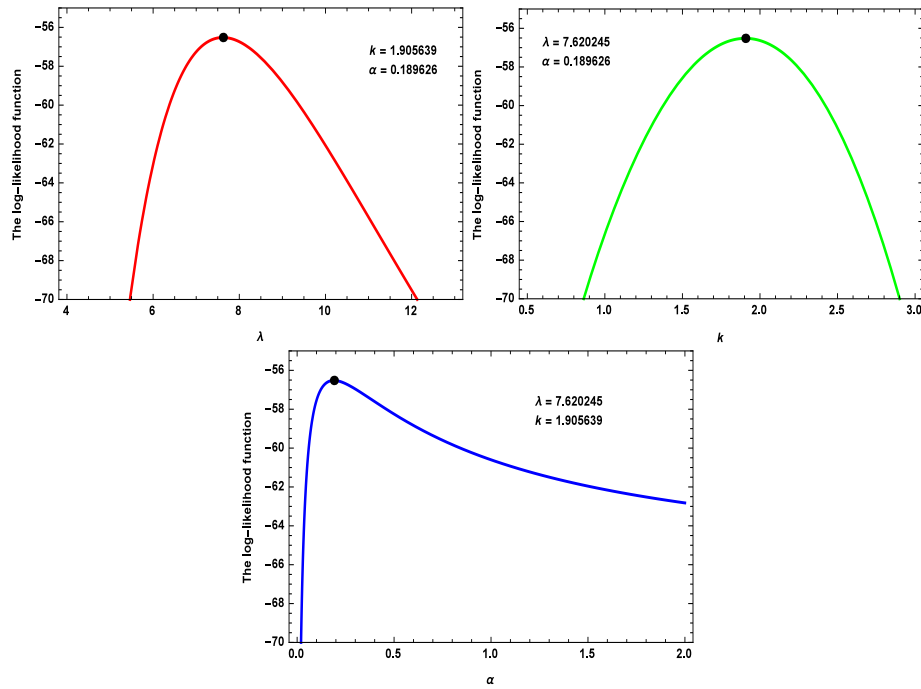


Figure 7: The log-likelihood as a function of different parameters of the extended half-Weibull distribution for Data set 2.

Acknowledgements.

Many thanks to the referee for careful reading and valuable comments which improved the paper representation.

References

- [1] Alizadeh M., Cordeiro G. M., Abrao D. C., Nascimento A. D. C., Lima M. do C. S., Ortega Edwin M. M., Odd-Burr generalized family of distributions with some applications, *Journal of statistical computation and simulation* 87 (2016) 367-389.
- [2] Alzaatreh A., Famoye F., Lee C., A new method for generating families of continuous distributions, *Metron* 71 (2013) 63-79.
- [3] Bakouch H. S., Chesneau C., Khan M. N., The extended odd family of probability distributions with practice to a submodel, *Filomat* 33:12 (2019) 3855–3867.
- [4] Castellares F, Lemonte AJ, On the Marshall-Olkin extended distributions, *Communications in Statistics-Theory and Methods* 45 (2016) 4537-4555.
- [5] Chesneau C., Bakouch H. S., Hussain T., A new class of probability distributions via cosine and sine functions with applications, *Communications in Statistics - Simulation and Computation* 48 (2019) 2287-2300.
- [6] Chesneau C., Bakouch H. S., Sharma V. K., A new family of distributions based on a poly-exponential transformation, *Journal of Testing and Evaluation* 48 (2020) 289-307.
- [7] Dasgupta R., On the distribution of Burr with applications, *Sankhya* 73 (2011) 1-19.
- [8] Eugene N., Lee C., Famoye F., Beta-normal distribution and its applications, *Communications in Statistics - Theory and Methods* 31 (2002) 497-512.
- [9] Ghitany M. E., Al-Mutairi D. K., Balakrishnan N., Al-Enezia L. J., Power Lindley distribution and associated inference, *Computational Statistics and data analysis* 64 (2013) 20-33.
- [10] Gupta R. C., Gupta P. L., Gupta, R. D., Modeling failure time data by Lehmann alternatives, *Communications in Statistics - Theory and Methods* 27 (1998) 887-904.
- [11] Marshall A., Olkin, I., A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika* 84 (1997) 641-652.
- [12] Murthy D. N. P., Xie M., Jiang R., *Weibull Models*. John Wiley & Sons, Inc., Publication. (2004) ISBN: 978-0-471-36092-6
- [13] Nasir A., Bakouch H. S., Jamal, F., Kumaraswamy odd Burr G family of distributions with applications to reliability data, *Studia Scientiarum Mathematicarum Hungarica* 55:1 (2018) 94-114.
- [14] Ristic M. M., Balakrishnan N., The gamma-exponentiated exponential distribution, *Journal of Statistical Computation and Simulation* 82 (2012) 1191-1206.
- [15] Zografos K., Balakrishnan N., On families of beta and generalized gamma-generated distributions and associated inference, *Statistical Methodology* 6 (2009) 344-362.