Research Article

Open Access

Kumaraswamy harris generalized kumaraswamy distribution and its Application in Survival Analysis

Abstract

In this paper, we introduced a new family of distributions called Kumaraswamy Harris Gnralized family of distributions. We explore some statistical properties of the family. Maximum likelihood method is used for estimation of unknown parameters. Special model, called Kumaraswamy Harris Generalized Kumaraswamy distribution is developed and studied. A Monte Carlo simulation study is conducted to validate the performance of the model. This distribution is applied to a real data set on survival analysis and verified that the new distribution is a better model than the Beta Kumaraswamy Weibull distribution.

Keywords: Kumaraswamy distribution, Harris Extended distribution, Beta Kumaraswamy Weibull distribution, survival analysis

Volume 11 Issue 1 - 2022

Jisha Varghese,[1] K.K. Jose[2]

[1]Department of Statistics, St. Thomas College, Mahatma Gandhi University, India

[2]Honorable Director, School of Mathematics and Statistics, Mahatma Gandhi University, India

Correspondence: K.K. Jose, Hon. Director, School of Mathematics and Statistics, Mahatma Gandhi University, Kottayam, Kerala 686 560, India, Email kkjstc@gmail.com

Received: January 07, 2022 | **Published:** February 22, 2022

Introduction

Addition of new parameters to expand classical distributions in order to get more flexible families of distributions has been investigated by several authors in the existing literature. In many applied areas such as lifetime analysis, finance and insurance, there is a clear need for extended forms of the classical distributions, that is, new distributions which are more flexible to model real data in these areas since the data can present a high degree of skewness and kurtosis. Recent developments focus on new techniques for building meaningful distributions.

In 1980, Kumaraswamy[1] introduced a two-parameter distribution on [0, 1], called Kumaraswamy distribution (Kw), for double bounded random processes for hydrological applications.

The probability density function (pdf) is

$$f(x;a,b) = abx^{(a-1)}(1-x^a)^{(b-1)}, \quad x \in [0,1]$$

where a and b are non- negative shape parameters.

The survival function (sf) is

$$\bar{F}(x;a,b) = (1-x^a)^b$$

Cordeiro and de Castro[2] defined the Kumaraswamy G distribution specified by the sf and the pdf

$$\bar{F}(x;a,b) = (1-G(x)^a)^b \tag{1}$$

and

$$f(x;a,b) = abg(x)G(x)^{(a-1)}(1-G(x)^a)^{(b-1)} \tag{2}$$

where G(x) is a general cdf with corresponding pdf g(x) and $a,b>0$ are shape parameters. The Kumaraswamy G families of distributions are more flexible than the baseline distribution in the sense that the families allow for greater flexibility of tail properties. Kumaraswamy G distributions can be tractable and effective models for skewed and censored data.

In the last several decades various forms of Kumaraswamy G family of distributions have appeared in the literature. For more details see de Pascoa et al.,[3] El-Sherpieny et al.,[4] de Santana et al.,[5] Saulo et al.,[6] Shahbaz et al.,[7] Correa et al.,[8] Bourguignon et al.,[9] Nadarajah et al.,[10] Paranaiba et al.,[11] Elbatal,[12-14] Muthulakshmi and Selvi,[15] Shams,[16] Lemonte et al.,[17] Cordeiro et al.,[18] Gomes et al.,[19] Huang and Oluyede[20] and Eldin et al.[21]

Aly and Benkherouf[22] introduced a new family of distributions, called Harris Extended (HE) family by adding two new parameters to a baseline distribution. Since the method is based on the probability generating function introduced by Harris,[23] the resulting family of distributions is known as Harris Extended (HE) family of distributions. The new method is based on the probability generating function (p.g.f.) of Harris[23] distribution. If f(x), $\bar{F}(x)$ and $h_F(x)$ denote the probability density function (pdf), survival function (sf) and hazard rate function (hrf) of a parent distribution and the parameters η and k are additional shape parameters, then the sf of HE family of distributions is given by

$$\bar{G}(x) = \left[\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right]^{\frac{1}{k}} \quad ; \quad x>0, \eta>0, k>0, \bar{\eta}=1-\eta \tag{3}$$

The pdf of HE distribution is

$$g(x) = \frac{\eta^{\frac{1}{k}} f(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]^{\frac{k+1}{k}}} \quad ; \quad x>0, \eta>0, k>0, \bar{\eta}=1-\eta \tag{4}$$

The hrf of the HE distribution is given by

$$h(x) = \frac{h_F(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]} \quad ; \quad x>0 \tag{5}$$

wher e $h_F(x;\theta)$ denotes the hazard rate function of the baseline distribution. When k=1, the above equations (3), (4) and (5) becomes the sf, pdf and hrf of Marshall-Olkin family of distributions respectively, which was introduced by Marshall-Olkin[24] and studied by Krishna et al.,[25] and many other researchers. Hence the HE family of distributions generalizes the well-known Marshall-Olkin class of distributions.

Batsidis and Lemonte[26] considered the HE family of distributions with respect to some lifetime models. Recently, Jose and Remya,[27] Pinho et al.,[28] Jose and Paul[29] studied various forms of HE models. This paper is a new generalization of Kumaraswamy Family of distributions introduced by Roshini and Thobias.[30]

This paper is organized as follows. Section 1 gives a brief introduction about Kumaraswamy G Family and Harris Extended Family of distributions. In section 2, introduced a new family, called Kumaraswamy Harris Generalized family (KwHG) family and explored its statistical properties such as probability density function (pdf), hazard rate function (hrf), expressions for cumulative distribution function (cdf), quantiles, survival function and Mill's Ratio, mixture representation, Renyi entropy and its shape properties. Method of maximum likelihood estimation is used for estimation of unknown parameters of the new family of distributions. In section 3, introduced and discussed about some special models of the new family. In section 4, the simulation study is conducted based on a special model, called Kumaraswamy Harris Generalized Kumaraswamy (KwHGKw) distribution. In section 5, the KwHGKw distribution is applied to the real data set on survival analysis to show the effectiveness of the new model. The conclusions are given in section 6.

Kumaraswamy harris generalized family of distributions

For a baseline random variable having pdf g(x) and sf $\bar{G}(x)$ then the two-parameter Kw-G distribution has sf (1) and the pdf (2).

Now we propose a new extension of the Harris Extended family for a given baseline distribution with cdf $G(x;\theta)$ and pdf $g(x;\theta)$ depending on a parameter vector θ.

Using (3) in (1), the survival function is

$$\bar{G}(x)=\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x))^k}{1-\bar{\eta}(\bar{F}(x))^k}\right)^{1/k}\right]^a\right\}^b, \quad x,a,b,\eta,k>0. \tag{6}$$

The pdf of the new Kumaraswamy Harris Generalized (KwHG) family of distributions is

$$g(x;a,b,\eta,k,\theta)=ab\frac{\eta^{1/k}f(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]^{\left(\frac{k+1}{k}\right)}}\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^{(a-1)}$$
$$\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^a\right\}^{(b-1)}, \quad x,a,b,\eta,k>0. \tag{7}$$

The hazard rate function and mill's ratio

The hazard rate function can be obtained using $h(x)=\frac{g(x)}{\bar{G}(x)}$ where $\bar{G}(x)=1-G(x)$ is the survival function of the new family.

Then

$$h(x;a,b,\eta,k)=\frac{ab\frac{\eta^{1/k}f(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]^{\left(\frac{k+1}{k}\right)}}\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^{(a-1)}}{\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^a\right\}} \tag{8}$$

The Mill's ratio is defined as

$$m(x)=\frac{\bar{G}(x)}{g(x)}$$

where $\bar{G}(x)$ is the survival function and $g(x)$ is the probability density function of the new family.

The Mill's ratio is related to the hazard rate function by the relation $m(x)=\frac{1}{h(x)}$.

Then

$$m(x;a,b,\eta,k)=\frac{\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^a\right\}}{ab\frac{\eta^{1/k}f(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]^{\left(\frac{k+1}{k}\right)}}\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^{(a-1)}}$$

The inverse Mill's ratio is the ratio of the probability density function to the cumulative distribution function of a distribution. A common application of the inverse Mill's ratio arises in regression analysis to take account of a possible selection bias.

The shape of density and hazard rate function

We obtain the shapes of density and hazard rate functions analytically. The modes of g(x) are the roots of the equation $\frac{\partial(g(x))}{\partial x}=0$ or equivalently $\frac{\partial\log(g(x))}{\partial x}=0$ Consider the logarithm of (7),

$$\log(g(x))=\log(ab)+\frac{1}{k}\log(\eta)+\log(f(x;\theta))-\frac{k+1}{k}\log[1-\bar{\eta}(\bar{F}(x;\theta))^k]$$
$$+(a-1)\log\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]+(b-1)\log\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]^a\right\}$$

In order to study the shape properties, we consider $\frac{\partial\log(g(x))}{\partial x}=0$ which gives

$$\frac{f'(x;\theta)}{f(x;\theta)}+\frac{(k+1)\bar{\eta}(\bar{F}(x;\theta))^{(k-1)}\bar{F}'(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]}$$
$$-\frac{(a-1)}{k}\frac{\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}-1}}{\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]}\frac{\eta k\left(\bar{F}(x;\theta)\right)^{k-1}F'(x;\theta)}{\left[1-\bar{\eta}\bar{F}(x;\theta)\right]^2}$$
$$+\frac{a(b-1)}{k}\frac{1}{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]^a}$$
$$\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]^{a-1}\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}-1}\frac{\eta k\left(\bar{F}(x;\theta)\right)^{k-1}F'(x;\theta)}{\left[1-\bar{\eta}\bar{F}(x;\theta)\right]^2}=0$$

That is

Citation: Varghese J, Jose KK. Kumaraswamy harris generalized kumaraswamy distribution and its Application in Survival Analysis. *Biom Biostat Int J.* 2022;11(1):28‒34. DOI: 10.15406/bbij.2022.11.00351

$$\frac{f'(x_i;\theta)}{f(x_i;\theta)}+\frac{(k+1)\bar{\eta}(\bar{F}(x_i;\theta))^{(k-1)}\bar{F}'(x_i;\theta)}{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^k\right]}-\eta\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}-1}\frac{\left(\bar{F}(x_i;\theta)\right)^{k-1}F'(x_i;\theta)}{\left[1-\bar{\eta}\bar{F}(x_i;\theta)\right]^2}$$

$$\left\{\frac{(a-1)}{\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}}\right]}-\frac{a(b-1)\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}}\right]^{a-1}}{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}}\right]^a}\right\}=0$$

There may be more than one root to (9). If $x=x_0$ is the root of (9), then it corresponds to a local maximum or local minimum depending on whether $\frac{\partial^2\log(g(x))}{\partial x^2}<0\,or>0.$

Consider the logarithm of (8) which gives,

$$\log(h(x))=\log(ab)+\frac{1}{k}\log(\eta)+\log(f(x;\theta))-\frac{k+1}{k}\log[1-\bar{\eta}(\bar{F}(x;\theta))^k]+(a-1)\log\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]-\log\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{\frac{1}{k}}\right]^a\right\}$$

$$\frac{\partial\log(h(x))}{\partial x}=\frac{f'(x_i;\theta)}{f(x_i;\theta)}+\frac{(k+1)\bar{\eta}(\bar{F}(x_i;\theta))^{(k-1)}\bar{F}'(x_i;\theta)}{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^k\right]}-$$

$$\eta\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}-1}\frac{\left(\bar{F}(x_i;\theta)\right)^{k-1}F'(x_i;\theta)}{\left[1-\bar{\eta}\bar{F}(x_i;\theta)\right]^2}$$

$$\left\{\frac{(a-1)}{\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}}\right]}+\frac{a\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}}\right]^{a-1}}{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^k}{1-\bar{\eta}(\bar{F}(x_i;\theta))^k}\right)^{\frac{1}{k}}\right]^a}\right\}$$

Then (10) corresponds to decreasing, increasing or constant hazard rates depending on whether $\frac{\partial\log(h(x))}{\partial x}<0,>0\,or=0$ respectively.

Quantile function

The u^{th} quantile of KwHE Family can be obtained by inverting $G(x)=u$ and is given by

$$x_u=1-\left\{\frac{\{1-[1-(1-u)^{\frac{1}{b}}]^{\frac{1}{a}}\}^k}{\{\eta+\bar{\eta}[1-[1-(1-u)^{\frac{1}{b}}]^{\frac{1}{a}}]^k\}}\right\}^{\frac{1}{k}}$$

where $0<u<1$ and $G(x)=1-\bar{G}(x)$

Measure of skewness and kurtosis

Galton[31] defined the formula for the measure of skewness by using quantile and is given by

$$S_k=\frac{Q(6/8)-2Q(4/8)+Q(2/8)}{Q(6/8)-Q(2/8)} \tag{9}$$

and Moors[32] defined the formula for the measure of kurtosis by using quantile and is given by

$$K=\frac{Q(7/8)-Q(5/8)+Q(3/8)-Q(1/8)}{Q(6/8)-Q(2/8)} \tag{10}$$

It is difficult to get the expression for the moments in KwHE family of distributions is in the closed form. So skewness and kurtosis of the KwHE family of distributions can obtain using Galton's skewness and Moor's kurtosis given in (11) and (12).

Skewness is a measure of the asymmetry of the probability distribution, in which the curve appears skewed either to the left or to the right.If $S_k=0$, the distribution is symmetric. If $S_k>0$, the distribution is skewed to the right and if $S_k<0$, the distribution is skewed to the left. Kurtosis is the degree of peakedness. As K increases, the tail of the distribution becomes heavier.

Order statistics

Let $X_1,X_2,...,X_n$ be a random sample of size n from a population with KwHE (a, b, η, k, θ) family of distributions. Let $X_{s:n}$ denote the s^{th} order statistics. Then the probability density function of the random variable $X_{s:n}$ is

$$g_{s:n}(x)=\frac{n!}{(s-1)!(n-s)!}g(x)G^{s-1}(x)(\bar{G}(x))^{n-s}$$

$$=\frac{ab\eta^{1/k}n!}{(s-1)!(n-s)!}\frac{f(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]^{\left(\frac{k+1}{k}\right)}}\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^{(a-1)}$$

$$\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^a\right\}^{[b(n-s+1)-1]}\left\{1-\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x))^k}{1-\bar{\eta}(\bar{F}(x))^k}\right)^{1/k}\right]^a\right\}^b\right\}^{(s-1)}$$

Record values

Record values and associated statistics are of much importance in many real life situations involving data relating to sports, weather, economics, life testing etc. Balakrishnan and Ahsanullah[33] and Arnold, Balakrishnan and Nagaraja[34] contributed significantly to the theory of record values. Let X_i, $i\geq 1$ be a sequence of i.i.d. random variables having an absolutely continuous c.d.f. F(x) and p.d.f. f(x). An observation X_j will be called an upper record value if its value exceeds that of all previous observations. Thus X_j is an upper record if $X_j\geq X_i$ for every $i<j$. The p.d.f. of nth record value say R_n is given by

$$f_{R_n}(x)=\frac{f(x)[-log(\bar{F}(x))]^n}{n!}$$

If $g_{R_n}(x)$ denote the density function of n^{th} record value from KwHE (a, b, η, k,θ), we have

$$g_{R_n}(x)=\frac{1}{n!}\frac{ab\eta^{1/k}f(x;\theta)}{\left[1-\bar{\eta}(\bar{F}(x;\theta))^k\right]^{\left(\frac{k+1}{k}\right)}}\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^{(a-1)}$$

$$\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x;\theta))^k}{1-\bar{\eta}(\bar{F}(x;\theta))^k}\right)^{1/k}\right]^a\right\}^{(b-1)}$$

Citation: Varghese J, Jose KK. Kumaraswamy harris generalized kumaraswamy distribution and its Application in Survival Analysis. *Biom Biostat Int J.* 2022;11(1):28‒34. DOI: 10.15406/bbij.2022.11.00351

$$\left[-log\left(\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x))^{k}}{1-\bar{\eta}(\bar{F}(x))^{k}}\right)^{1/k}\right]^{a}\right\}^{b}\right)\right]^{n}$$

where x, a, b, η, $k>0$.

Maximum likelihood estimation

In this section, the maximum likelihood estimators of the unknown parameters a,b,η *and* k of the KwHE distribution are derived. Let $X_1,X_2,X_3,\ldots X_n$ be a random sample of size n from KwHE distribution, then the likelihood function is

$$L=(ab)^{n}\frac{\eta^{n/k\sum_{i=1}^{n}f(x_i;\theta)}}{\sum_{i=1}^{n}\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]^{\left(\frac{k+1}{k}\right)}}\sum_{i=1}^{n}\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{(a-1)}$$

$$\sum_{i=1}^{n}\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a}\right\}^{(b-1)}$$

The log likelihood function is given by

$$\log L=n\log(ab)+\frac{n}{k}\log(\eta)+\sum_{i=1}^{n}\log f(x_i,\theta)-\frac{k+1}{k}\sum_{i=1}^{n}\log\left[1-\bar{\eta}(\bar{F}(x_i,\theta))^{k}\right]$$
$$+(a-1)\sum_{i=1}^{n}\log\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]+(b-1)\sum_{i=1}^{n}\log\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a}\right\}$$

The partial derivatives of the log likelihood with respect to a, b, η, k and θ are obtained as

$$\frac{\partial \log L}{\partial a}=\frac{n}{a}+\sum_{i=1}^{n}\log\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]\left\{\frac{1+(b-2)\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a}}{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a}}\right\}$$

$$\frac{\partial \log L}{\partial b}=\frac{n}{b}+\sum_{i=1}^{n}\log\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a}\right\}$$

$$\frac{\partial \log L}{\partial \eta}=\frac{n}{k\eta}-\frac{k+1}{k}\sum_{i=1}^{n}\frac{(\bar{F}(x_i;\theta))^{k}}{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]}$$

$$-\frac{1}{k}\sum_{i=1}^{n}\left[\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right]^{\frac{1}{k}-1}\frac{\left[\left(1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right)(\bar{F}(x_i;\theta))^{k}-\eta(\bar{F}(x_i;\theta))^{2k}\right]}{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]^{2}}$$

$$\frac{\partial \log L}{\partial k}=-\frac{n}{k^{2}}\log\eta+\bar{\eta}(k+1)\sum_{i=1}^{n}\frac{(\bar{F}(x_i;\theta))^{k-1}}{\left(1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right)}+\frac{1}{k^{2}}\sum_{i=1}^{n}\log\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]$$

$$-\frac{(a-1)}{k}\sum_{i=1}^{n}\frac{\left[\eta(\bar{F}(x_i;\theta))^{k}\right]^{\frac{1}{k}}\log\bar{F}(x_i;\theta)}{\left\{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]^{\frac{1}{k}}-\left[\eta(\bar{F}(x_i;\theta))^{k}\right]^{\frac{1}{k}}\right\}\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]}$$

$$+\frac{a(b-1)}{k}\sum_{i=1}^{n}\frac{\frac{\left[\eta(\bar{F}(x_i;\theta))^{k}\right]\log\bar{F}(x_i;\theta)}{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]^{2}}\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a-1}\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{\frac{1}{k}-1}}{\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{1/k}\right]^{a}\right\}}$$

$$\frac{\partial \log L}{\partial \theta}=\sum_{i=1}^{n}\frac{f'(x_i;\theta)}{f(x_i;\theta)}+\bar{\eta}(k+1)\sum_{i=1}^{n}\frac{(\bar{F}(x_i;\theta))^{(k-1)}\bar{F}'(x_i;\theta)}{\left[1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}\right]}$$

$$-\eta\sum_{i=1}^{n}\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{\frac{1}{k}-1}\frac{\left(\bar{F}(x_i;\theta)\right)^{k-1}F'(x_i;\theta)}{\left[1-\bar{\eta}\bar{F}(x_i;\theta)\right]^{2}}$$

$$\left\{\frac{(a-1)}{\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{\frac{1}{k}}\right]}-\frac{a(b-1)\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{\frac{1}{k}}\right]^{a-1}}{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{\frac{1}{k}}\right]^{a}}\right\}$$

In order to estimate the parameters, we have to solve the normal equations

$$\frac{\partial \log L}{\partial a}=0;\quad \frac{\partial \log L}{\partial \eta}=0;\quad \frac{\partial \log L}{\partial k}=0;\quad \frac{\partial \log L}{\partial \theta}=0. \qquad (11)$$

Since (13) cannot be solved analytically, numerical iteration technique is used to get a solution for the parameters a, η, k, and θ. One may use the nlm package in R software to get the maximum likelihood estimator (MLE) for the parameters.

The MLE of b can be obtained by solving $\frac{\partial \log L}{\partial b}=0$ which gives

$$\hat{b}=\frac{-n}{\sum_{i=1}^{n}\log\left\{1-\left[1-\left(\frac{\eta(\bar{F}(x_i;\theta))^{k}}{1-\bar{\eta}(\bar{F}(x_i;\theta))^{k}}\right)^{\frac{1}{k}}\right]^{a}\right\}}$$

Kumaraswamy harris generalized kumaraswamy distribution

Let X follows Kumaraswamy distribution with parameters α and β with pdf, cdf and survival function $f(x)=\alpha\beta x^{(\alpha-1)}(1-x^{\alpha})^{(\beta-1)}$, $F(x)=1-(1-x^{\alpha})^{\beta}$ and $\bar{F}(x)=(1-x^{\alpha})^{\beta}$ respectively. If we apply the sf and pdf of Kumaraswamy distribution in Kumaraswamy Harris Generalized Family given in (6) and (7), we get the sf and pdf of the new distribution called Kumaraswamy Harris Generalized Kumaraswamy distribution with parameters a, b, η, k, α, β and is denoted by KwHGKw.

The survival function is

$$\bar{G}(x;a,b,\eta,\alpha,\beta,k)=\left\{1-\left[1-\left(\frac{\eta((1-x^{\alpha})^{\beta})^{k}}{1-\bar{\eta}\left((1-x^{\alpha})^{\beta}\right)^{k}}\right)^{1/k}\right]^{a}\right\}^{b}$$

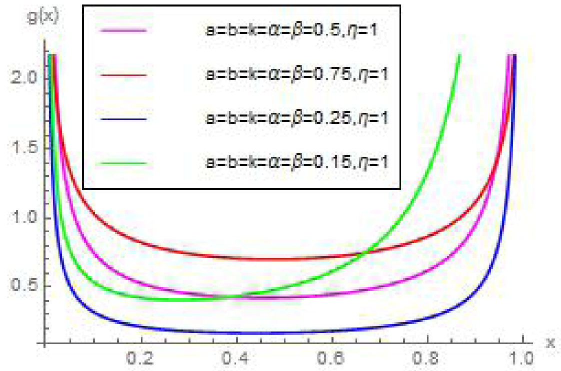
Citation: Varghese J, Jose KK. Kumaraswamy harris generalized kumaraswamy distribution and its Application in Survival Analysis. *Biom Biostat Int J.* 2022;11(1):28‒34. DOI: 10.15406/bbij.2022.11.00351

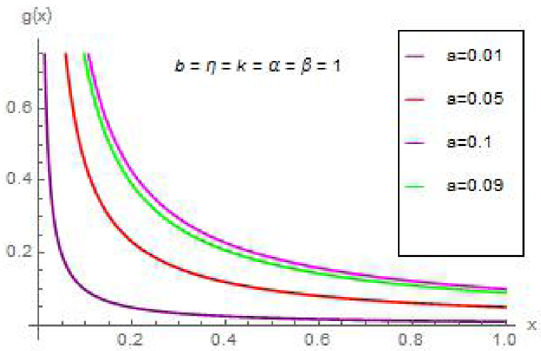
The pdf is

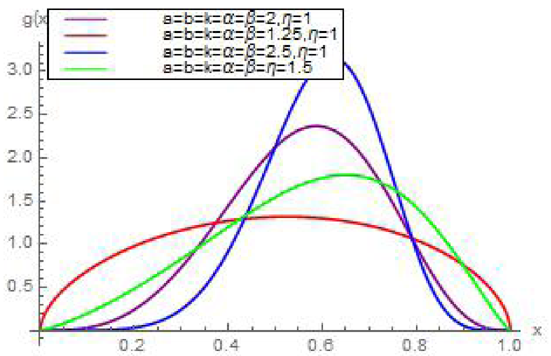
$$g(x;a,b,\eta,\alpha,\beta,k)=ab\frac{\eta^{1/k}\alpha\beta x^{\alpha-1}(1-x^{\alpha})^{\beta-1}}{\left[1-\bar{\eta}\left((1-x^{\alpha})^{\beta}\right)^{k}\right]^{\left(\frac{k+1}{k}\right)}}\left[1-\left(\frac{\eta((1-x^{\alpha})^{\beta})^{k}}{1-\bar{\eta}\left((1-x^{\alpha})^{\beta}\right)^{k}}\right)^{1/k}\right]^{(\alpha-1)}$$

$$\left\{1-\left[1-\left(\frac{\eta((1-x^{\alpha})^{\beta})^{k}}{1-\bar{\eta}\left((1-x^{\alpha})^{\beta}\right)^{k}}\right)^{1/k}\right]^{a}\right\}^{(b-1)}, \quad x,a,b,\eta,\alpha,\beta,k>0.$$

The density plot for different values of the parameters are given in the following figure.







The u^{th} quantile function is obtained as

$$x_u=\left\{1-\left[\frac{A}{\{\eta+\bar{\eta}A^{k}\}^{\frac{1}{k}}}\right]^{\frac{1}{\beta}}\right\}^{\frac{1}{\alpha}}$$

where $A=1-[1-(1-u)^{\frac{1}{b}}]^{\frac{1}{a}}$ and U follows U(0,1).

The log likelihood function is given by

$$logL=n\log(ab)+\frac{n}{k}\log(\eta)+\sum_{i=1}^{n}\log\left[\alpha\beta x^{(\alpha-1)}(1-x^{\alpha})^{(\beta-1)}\right]-\frac{k+1}{k}\sum_{i=1}^{n}\log\left[1-\bar{\eta}((1-x^{\alpha})^{\beta})^{k}\right]$$

$$+(a-1)\sum_{i=1}^{n}\log\left[1-\left(\frac{\eta((1-x^{\alpha})^{\beta})^{k}}{1-\bar{\eta}((1-x^{\alpha})^{\beta})^{k}}\right)^{1/k}\right]+(b-1)\sum_{i=1}^{n}\log\left\{1-\left[1-\left(\frac{\eta((1-x^{\alpha})^{\beta})^{k}}{1-\bar{\eta}((1-x^{\alpha})^{\beta})^{k}}\right)^{1/k}\right]^{a}\right\}$$

Simulation study

In this section, we assess the performance of the KwHGKw distribution by conducting various simulations for different sample sizes and parameter values. We generate samples by using Monte Carlo simulation with sample size n = 25, 75, 90, 150 and the parameter values a = 0.9, b = 0.3, η =0.9, k=0.5, $\alpha=0.6$ and $\beta=0.5$ were considered. We find the biases and mean squared errors (mse). The results of the Monte Carlo study are given in Table 1.

The values in the Table 1 show that mse of the estimate of the parameters decreases as the sample size increases. This fact indicates that the normal distribution is an asymptotic distribution of the estimates.

Application in survival analysis

In order to show the applicability of the new model, the Kumaraswamy Harris Generalized Kumaraswamy distribution (KwHGKw) is applied to data set on survival analysis. The performance of the proposed model is compared with the Beta Kumaraswamy Weibull (BKwW) model using log-likelihood, AIC, BIC and AICC. The data represents the ordered remission times (in months) of a random sample of 142 bladder cancer patients, reported in Lee and Wang.[35] The result reveals that the applicability of Kumaraswamy Harris Generalized Kumaraswamy (KwHGKw) distribution in survival analysis. The results are given in Table 1. The data are as follows.

Remission times (in months) of a random sample of 142 bladder cancer patients.

ccccccccccccc 0.08 0.20 0.40 0.50 0.51 0.81 0.87 0.90 1.05 1.19 1.26 1.35 1.40

1.46 1.76 2.02 2.02 2.07 2.09 2.23 2.26 2.46 2.54 2.62 2.64 2.69

2.69 2.75 2.83 2.87 3.02 3.25 3.31 3.36 3.36 3.36 3.48 3.52 3.57

3.64 3.70 3.82 3.88 4.18 4.23 4.26 4.33 4.33 4.34 4.40 4.50 4.51

4.65 4.70 4.87 4.98 5.06 5.09 5.17 5.32 5.32 5.34 5.41 5.41 5.49

5.62 5.71 5.85 6.25 6.54 6.76 6.93 6.94 6.97 7.09 7.26 7.28 7.32

7.39 7.59 7.62 7.28 7.32 7.39 7.59 7.62 7.63 7.66 7.87 7.93 8.26

8.37 8.53 8.60 8.65 8.66 9.02 9.22 9.47 9.74 10.06 10.34 10.66 10.75

10.86 11.25 11.64 11.79 11.98 12.02 12.03 12.07 12.69 13.11 13.29 13.80 14.24

14.76 14.77 14.83 15.96 16.62 17.12 17.14 17.36 18.10 19.13 19.36 20.28 21.73

22.69 23.63 24.80 25.74 25.82 26.31 32.15 34.26 36.66 43.01 46.12 79.05

Citation: Varghese J, Jose KK. Kumaraswamy harris generalized kumaraswamy distribution and its Application in Survival Analysis. Biom Biostat Int J. 2022;11(1):28‒34. DOI: 10.15406/bbij.2022.11.00351

From Table 1, we can observe that the values of -logL, AIC, BIC and AICC for the Kumaraswamy Harris Generalized Kumaraswamy (KwHGKw) distribution is smaller than that of Beta Kumaraswamy Weibull distribution (BKwW). So the KwHEKw distribution is a better model for the survival data than BKwW distribution.

Table 1 Simulation Study

Sample (m)	Parameters	$\hat{a}$	$\hat{b}$	$\hat{\eta}$	$\hat{k}$	$\hat{\alpha}$	$\hat{\beta}$
	bias	0.04335	-0.2999	0.03828	0.43828	0.33828	0.43828
	mse	0.001879	0.08999	0.001465	0.1920918	0.1144352	0.1920918
	bias	0.069799	-0.2999	0.045912	0.445912	0.345912	0.445912
	mse	0.00487	0.08999	0.002107	0.198838	0.11965	0.198838
	bias	0.03173722	-0.299998	0.033630	0.4336306	0.3336306	0.4336306
	mse	0.00100725	0.08999	0.001131	0.1880355	0.111309	0.188035
	bias	0.004273113	-0.2999	0.01680361	0.4168036	0.3168036	0.4168036
	mse	0.0000182	0.0899	0.0002823	0.1737252	0.1003645	0.1737252

Table 2 MLE's, -logL, AIC, BIC and AICC for the survival data

ibution	Distr	$\hat{a}$	$\hat{b}$		$\hat{\eta}$	$\hat{k}$		$\widehat{a}$		$\hat{\beta}$	$\hat{\beta}$	-logL	AIC	BIC	AICC
EKw	KwH	0.921	0.783	99	0.999	0.565	02	0.502	499	0.499	0	40.454	92.908	93.82	92.622
W	BKw	0.862	0.766	99	0.999	0.501	02	0.402	499	0.474	0	52.297	116.594	117.506	117.216

Conclusion

In this paper, we introduced a new family of distributions called Kumaraswamy Harris Generalized family of distributions and is denoted by 'KwHG'. We explore the statistical properties such as probability density function (pdf), hazard rate function (hrf), expressions for cumulative distribution function (cdf), quantiles, measure of skewness and Kurtosis, order statistics, record values, survival function and Mill's Ratio and its shape properties. Method of maximum likelihood estimation is used for estimation of unknown parameters of the new distribution. A special model called Kumaraswamy Harris Generalized Kumaraswamy model is developed for this new family. The probability density function (pdf), cumulative distribution function (cdf) and hazard rate function (hrf) are obtained and shape property is considered for the new model. Kumaraswamy Harris Generalized Kumaraswamy distribution (KwHGKw) is applied to the real data set on survival analysis to show the effectiveness of the new model. Based on the analysis, it is concluded that the Kumaraswamy Harris Genralized Kumaraswamy distribution (KwHGKw) is a better model for the survival data analysis than the Beta Kumaraswamy Weibull distribution (BKwW). Simulation study is also conducted for various values.

Acknowledgment

The authors are grateful to the reviewers for the valuabule suggestions to improve the paper.

Conflicts of interest

None

References

1. Kumaraswamy P. A generalized probability density function for double–bounded random processes. *Journal of Hydrology.* 1980;46: 79–88.
2. Cordeiro GM, de Castro M. A new family of generalized distributions. *J Stat Comput Simul*. 2011;81(7):883–893.
3. de Pascoa Mar, Ortega EMM, Cordeiro GM. The Kumaraswamy generalized gamma distribution with application in survival analysis. *Stat Methodol*. 2011;8(5):411–433.
4. El–Sherpieny ESA, Ahmed MA. On the Kumaraswamy–Gumbel distribution. Paper presented at 46th Ann Conf Statist Comput Sci Oper Res, ISSR–Cairo University, Egypt. 2011.
5. de Santana TVF, Ortega EMM, Cordeiro GM, et al. The Kumaraswamy log–logistic distribution. *J Stat Theor Appl*. 2012;11:265–291.
6. Saulo H, Leao J, Bourguignion M. The Kumaraswamy Birnbaum–Sanders distribution. *J Stat Theory Pract*. 2012;6:754–759.
7. Shahbaz MQ, Shahbaz S, Butt NS. The Kumaraswamy inverse–Weibull distribution. *Pakistan J Statist Oper Res*. 2012;8:479–489.
8. Correa MA, Nougueira DA, Ferreira EB. Kumaraswamy normal and Azzalini's skew normal modeling asymmetry. *Sigmae*. 2012;1:65–83.
9. Bourguignion M, Silva MB, Zea LM, et al. The Kumaraswamy Pareto distribution. *J Stat Theor Appl.* 2013;12(2):129–144.
10. Nadarajah S, Cordeiro GM, Ortega EMM. General results for the Kumaraswamy G distribution. *J Stat Comput Simul.* 2011;87:951–979.
11. Paranaiba PF, Ortega EMM, Cordeiro GM. et al. The Kumaraswamy burr XII distribution: theory and practice. *J Stat Comput Simul*. 2013;83:2117–2143.
12. Elbatal I. Kumaraswamy linear exponential distribution. *Pioneer J Theor Appl Statist.* 2013a;5:59–73.
13. Elbatal I. Kumaraswamy generalized linear failure rate distribution. *Indian J Comput Appl Math*. 2013b;1:61–78.
14. Elbatal I. Kumaraswamy exponentiated Pareto distribution. Econ Qual Control. 2013c;28:1–8.
15. Muthulakshmi S, Selvi BGG. Double sampling plan for truncated life test based on Kumaraswamy–log–logistic distribution. *IOSR J Math.* 2013;7:29–37.
16. Shams TM. The Kumaraswamy–generalized Lomax distribution. *Middle–East J Sci Res*. 2013;17: 641–646.

17. Lemonte AJ, Barreto–Souza W, Cordeiro GM. The exponentiated Kumaraswamy distribution and its log–transform. *Brazilian Journal of Probability and Statistics.* 2013;27(1):31–53.
18. Cordeiro GM, Ortega EMM, Silva GO. The Kumaraswamy modified Weibull distribution: Theory and applications. *J Stat Comput Simul.* 2014;84(7):1387–1411.
19. Gomes AE, Da Silva CQ, Cordeiro GM, et al. A new lifetime model: The Kumaraswamy generalized Rayleigh distribution. *J Stat Comput Simul.* 2014;84:290–309.
20. Huang S, Oluyede BO. Exponentiated kumaraswamy–dagum distribution with applications to income and lifetime data. journal of statistical distributions and applications. 2014;1:8.
21. Eldin MM, Khalil N, Amein M. Estimation of parameters of the Kumaraswamy distribution based on general progressive type II censoring. *American Journal of Theoretical and Applied Statistics.* 2014;3(6):217–222.
22. Aly EAA, Benkherouf L. A new family of distributions based on probability generating functions. *Sankhya B.* 2011;73:70–80.
23. Harris TE. Branching processes. *Ann Math Stat*. 1948;19:474–494.
24. Marshall AW, Olkin I. A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*. 1997;84:641–652.
25. Krishna E, Jose KK, Alice T, et al. The marshall–olkin fre'chet distribution. *Communications in statistics theory and methods*. 2013a;42(22):4091–4107.
26. Batsidis A, Lemonte AJ. On the harris extended family of distributions. *Statistics.* 2014.
27. Jose KK, Remya S. Harris extended burr XII and exponentiated exponential distribution. *International Journal of Computer and Mathematical Sciences (in press).* 2014.
28. Pinho LGB. Cordeiro GM, Nobre JS. The Harris Extended exponential distribution. *Communications in Statistics– Theory and Methods.* 44:3486–3502.
29. Jose KK, Paul A. reliability test plans for percentiles based on the harris generalized linear exponential distribution. *Stochastics and Quality Control*. 2018.
30. George R, Thobias S. Kumaraswamy marshall–olkin exponential distribution. *Commun Statist Theor Meth*. 2018.
31. https://galton.org/books/human-faculty/text/galton-1883-human-faculty-v4.pdf
32. Moors JJA. A quantile alternative for kurtosis. *The Statistician.* 1988;37(1):25–32.
33. Balakrishnan N, Ahsanullah M. Relations for single and product moments of record values from Lomax distributions. *Sankhya Series B.* 1994;56:140–146.
34. Arnold BC, Balakrishnan N, Nagaraja HN. Records, New York: John Wiley and Sons, INC. 1998;344p.
35. Lee W, statistical methods for survival data analysis, third edition, 2003.