**KUMARASWAMY DISTRIBUTIONS: A NEW FAMILY OF GENERALIZED DISTRIBUTIONS**

**Pankaj Das**

**PhD scholar, Roll No: 10683**

**Email ID: pankaj.iasri@gmail.com**

**ICAR-I.A.S.R.I., Library Avenue, New Delhi-110012**

**Abstract**

Kumaraswamy introduced a distribution for double bounded random processes with hydrological applications. For any continuous baseline G distribution, G.M. Cordeiro and M. de Castro describe a new family of generalized distributions (denoted with the prefix "*Kw*") to extend the normal, Weibull, gamma distributions, among several well-known distributions. Some special distributions in the new family such as the *Kw*-normal, *Kw*-Weibull, and *Kw*-gamma distribution are discussed. We discuss the ordinary moments of any *Kw* generalized distribution as linear functions of probability weighted moments of the parent distribution. We also obtain the ordinary moments of order statistics as functions of probability weighted moments of the baseline distribution. We use the method of maximum likelihood to fit the distributions in the new class and illustrate the potentiality of the new model with two application to real data.

**Keywords:** gamma distribution; Kumaraswamy distribution; moments; normal distribution; order statistics; Weibull distribution

1. **Introduction:**

Beta distributions are very versatile and a variety of uncertainties can be usefully modeled by them. In practical situation, many of the finite range distributions encountered can be easily transformed into the standard beta distribution. In econometrics, many times the data are modeled by finite range distributions. Generalized beta distributions have been widely studied in statistics and numerous authors have developed various classes of these distributions. Eugene *et al.* (2002) proposed a general class of distributions for a random variable defined from the logit of the beta random variable by employing two parameters whose role is to introduce skewness and to vary tail weight. Following the work of Eugene *et al*. (2002), who defined the beta normal distribution, Nadarajaha and Kotz (2004) introduced the beta Gumbel distribution, Nadarajaha and Gupta (2004) proposed the beta Frechet distribution and Nadarajaha and Kotz (2004) worked with the beta exponential distribution. However, all these works lead to some mathematical difficulties because the beta distribution is not fairly tractable and, in particular, its cumulative distribution function (cdf) involves the incomplete beta function ratio.

Poondi Kumaraswamy proposed a new probability distribution for variables that are lower and upper bounded. In probability and statistics, the Kumaraswamy's double bounded distribution is a family of continuous probability distributions defined on the interval [0, 1] differing in the values of their two non-negative shape parameters, *a* and *b*. In reliability and life testing experiments, many times the data are modeled by finite range distributions.

Eugene *et al* (2004) and Jones (2004) constructed a new class of Kumaraswamy generalized distribution (*Kw-G* distribution) on the interval (0, 1). The probability density function (pdf) and the cdf with two shape parameters *a* >0 and *b* > 0 defined by

 (1)

Where 

respectively, where 𝑓(𝑥) = 𝑑𝐹(𝑥)/𝑑𝑥 and𝑎,𝑏 > 0 are additional shape parameters to the distribution F. Except for some special choices of the function 𝐹(𝑥).The associated hazard rate function (hrf) is 

1. **Conversion of a distribution into *Kw-G* distribution:**

Let a parent continuous distribution having cdf *G(x)* and pdf g(x). Then by applying the quantile function on the interval (0, 1) we can construct *Kw-G* distribution (Cordeiro and de Castro, 2009). The cdf F(x) of the *Kw-G* is defined as

**** (2)

Where a > 0 and b > 0 are two additional parameters whose role is to introduce skewness and to vary tail weights.

Similarly the density function of this family of distributions has a very simple form

 (3)

1. **Some Special *Kw* generalized distributions:** 
   1. ***Kw*- *normal*:**

The *KN* density is obtained from (3) by taking G (.) and g (.) to be the cdf and pdf of the normal  distribution, so that

 (4)

where  is a location parameter, *σ* > 0 is a scale parameter, a, b > 0 are shape parameters, and  and Ф (.) are the pdf and cdf of the standard normal distribution, respectively. A random variable with density *f (x)* above is denoted by *X* ~ *Kw-N*For *µ*= 0 and *σ* = 1 we obtain the standard *Kw-N* distribution. Further, the *Kw-N* distribution with a = 2 and b = 1 coincides with the skew normal distribution with shape parameter equal to one.

* 1. ***Kw-weibull*:**

The cdf of the Weibull distribution with parameters β > 0 and *c* > 0 is for *x* > 0. Correspondingly, the density of the *Kw-Weibull* distribution, say *Kw-W* *(a, b, c, β),* reduces to

 (5)

Here *x, a, b, c, β* > 0

If c = 1 we obtain the *Kw*-exponential distribution. The *Kw-W* (1, b, 1, β) distribution corresponds to the exponential distribution with parameter β\* = bβ.

* 1. **Kw-gamma:**

Let Y be a gamma random variable with cdf  for *y, α, β* > 0, where (-) is the gamma function and is the incomplete gamma function. The density of a random variable *X* following a *Kw-Ga* distribution, say *X* ~ *Kw-Ga (a, b, β, α)*, can be expressed as

 , *x, α, β, a, b* >0 (6)

For α=1, we obtain the *Kw-*exponential distribution.

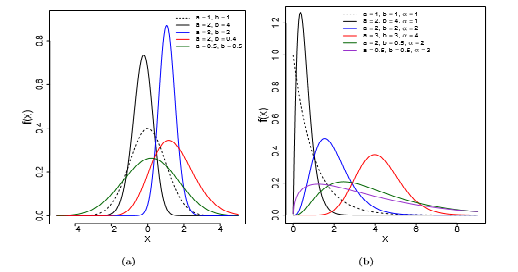
****

Figure 1. Some possible shapes of density function of *Kw-G* distribution. (a) *Kw-normal* (*a, b, 0, 1*) and (b) *Kw- gamma* (*a, b, 1, α*) density functions (dashed lines represent the parent distributions)

1. **A general expansion for the density function:**

Cordeiro and de Castro (2009) elaborate a general expansion of the distribution.

For b > 0 real non-integer, the form of the distribution

 (7)

where the binomial coefficient is defined for any real. From the above expansion and formula (3), we can write the *Kw-G* density as

 (8)

Where the coefficients are  and 

1. **General formulae for the moments:**

The s-th moment of the Kw-G distribution can be expressed as an infinite weighted sum of PWMs of order (s, r) of the parent distribution G from equation (8) for a integer and for a real non-integer. We assume Y and X following the baseline G and Kw-G distribution, respectively. The *s*-th moment of X, say µ's, can be expressed in terms of the (s, r)-th PWMs   of Y for r = 0, 1 ..., as defined by Greenwood et al. (1979). For a integer,

 (9)

Whereas for *a* real non integer the formula

 (10)

We can calculate the moments of the *Kw-G* distribution in terms of infinite weighted sums of PWMs of the G distribution. Established power series expansions to calculate the moments of any *Kw-G* distribution can be more efficient than computing these moments directly by numerical integration of the expression.

1. **Probability weighted moments:**

A general theory for PWMs covers the summarization and description of theoretical probability distributions, the summarization and description of observed data samples, nonparametric estimation of the underlying distribution of an observed sample, estimation of parameters and quantiles of probability distributions and hypothesis testing for probability distributions. (Barakat and Abdelkader, 2004)

The *(s,r*)-th PWM of *X* following the *Kw-G* distribution, say , is formally defined by   (11)

This formula also can be written in the following form

 (12)

Where  is the (*s,m+l)*-th PMW of G distribution and the coefficients



1. **Order statistics:**

The density  of the *i*-th order statistic, for *i = 1,..., n*, from *i.i.d*. random variables *X1,... ,Xn* following any *Kw-G* distribution, is simply given by



= (13)

Where *B(*.,.) denote the beta function and then

  (14)

After expanding all the terms we get the following two forms

When *a*= non integer

 (15)

And when *a*= integer

 (16)

Formulae (15) and (16) immediately yield the density of order statistics of the *Kw-G* distribution as a function of the density of the baseline distribution multiplied by infinite weighted sums of powers of G(x). Hence, the ordinary moments of order statistics of the *Kw-G* distribution can be written as infinite weighted sums of PWMs of the G distribution. These generalized moments for some baseline distributions can be accurate computationally by numerical integration as mentioned before.

1. **L moments:**

The L-moments are analogous to the ordinary moments but can be estimated by linear combinations of order statistics. The L-moments have several theoretical advantages over the ordinary moments. They exist whenever the mean of the distribution exists, even though some higher moments may not exist. They are able to characterize a wider range of distributions and, when estimated from a sample, are more robust to the effects of outliers in the data. Unlike usual moment estimates, the parameter estimates obtained from L-moments are sometimes more accurate in small samples than even the maximum likelihood estimates (MLEs). The L-moments are linear functions of expected order statistics defined as



(17)   
the first four L-moments are  ,, 

and . The L-moments can also be calculated in terms of PWMs given in (12) as

 (18)

In particular, 

1. **Other measure:**

**9.1. Mean deviations**:

Mean deviation denotes the amount of scatter in a population. This is evidently measured to some extent by the totality of deviations from the mean and median. Let *X* ∼ *Kw-G (a, b*). The mean deviations about the mean (*δ1(X))* and about the median *(δ2(X))* can be expressed as

 and  Where, *M* = median, is come from pdf and 

**9.2. Extreme values:**

If  denotes the sample mean from *i.i.d*. random variables following (2), then by the usual central limit theorem approaches the standard normal distribution as *n* →∞ under suitable conditions. Sometimes one would be interested in the asymptotics of the extreme values Mn = max(*X1... Xn*) and mn = min(*X1,...,Xn*).

First, suppose that G belongs to the max domain of attraction of the Gumbel extreme value distribution. Then by Leadbetter *et al*. there must exist a strictly positive function, say h(t), such that  for every x ∈ (−∞,∞).So, it follows by Leadbetter *et al.*(1987) that F also belongs to the max domain of attraction of the Gumbel extreme value distribution with 

1. **Estimation of parameter:**

Let γ be the *p*-dimensional parameter vector of the baseline distribution in equations (2) and (3). We consider independent random variables *X1,..., Xn*, each *X*i following a *Kw-G* distribution with parameter vector θ = (a,b, γ). The log-likelihood  function for the model parameters obtained from (3) is



The elements of the score vector are given by





And 

These partial derivatives depend on the specified baseline distribution. Numerical maximization of the log-likelihood above is accomplished by using the RS method (Rigby and Stasinopoulos, 2005) available in the gamlss package in R. Since numerically the maximum likelihood estimation of the parameters of the *Kw-G* distributions is much simpler than the estimation of the parameters of the generalized beta distributions, we recommend to use *Kw-G* distributions in place of the second family of distributions.

We can compare non-nested Kw-G distributions by penalizing over-fitting using the Akaike information criterion given by AIC =, where p\* is the number of model parameters. The distribution with the smallest value of AIC (among all distributions considered) is usually taken as the best model for describing the given data set. This comparison is based on the consideration of a model that shows a lack of fit with one that does not.

1. **Relation to the Beta distribution**: The density function of beta distribution is defined as



Assume that is a Kumaraswamy distributed random variable with parameters *a* and *b*. Then is the *a*-th root of a suitably defined Beta distributed random variable. More formally, Let denote a Beta distributed random variable with parameters  and. One has the following relation between  and.



With equality in distribution.



Jones (2008) explored the background and genesis of the *Kw* distribution and, more importantly, made clear some similarities and differences between the beta and *Kw* distributions. For example, the *Kw* densities are also unimodal, increasing, decreasing or constant depending in the same way as the beta distribution on the values of its parameters. He highlighted several advantages of the *Kw* distribution over the beta distribution: the normalizing constant is very simple; simple explicit formulae for the distribution and quantile functions which do not involve any special functions; a simple formula for random variate generation; explicit formulae for L-moments and simpler formulae for moments of order statistics. Further, according to Jones (2008), the beta distribution has the following advantages over the *Kw* distribution: simpler formulae for moments and moment generating function; a one-parameter sub-family of symmetric distributions; simpler moment estimation and more ways of generating the distribution via physical processes.

1. **Application:**

We illustrate the superiority of some new Kw-G distributions proposed here as compared with some of their sub-models. We give two applications (uncensored and censored data) using well- known data sets to demonstrate the applicability of the proposed regression model.

**12.1. Censored data:**

In this section we present an example with data from adult numbers of Flour beetle (*T. confusum*) cultured at 29°C presented by Cordeiro and de Castro (2009).

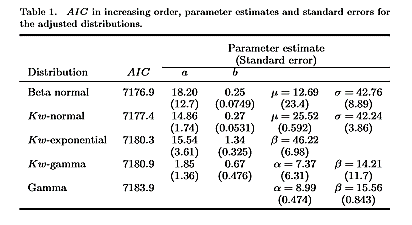


Table 1 gives AIC values in increasing order for some fitted distributions and the MLEs of the parameters together with its standard errors. According to AIC, the beta normal and *Kw*-normal distributions yield slightly different fittings, outperforming the remaining selected distributions. Notice that for the beta normal distribution the variability in the estimates of *a*, *µ* and *σ* is appreciably greater.

The fitted distributions superimposed to the histogram of the data in Figure 3 reinforce the result in Table 1 for the gamma distribution. The beta normal and the *Kw-normal* distributions are almost indistinguishable. This claim is further strengthened by the comparison between observed and expected frequencies in Table 2. The mean absolute deviation between expected and observed frequencies reaches the minimum value for the Kw-normal distribution.

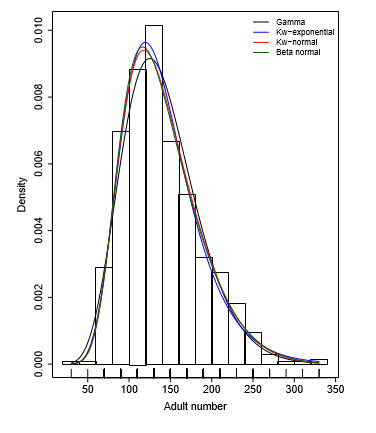
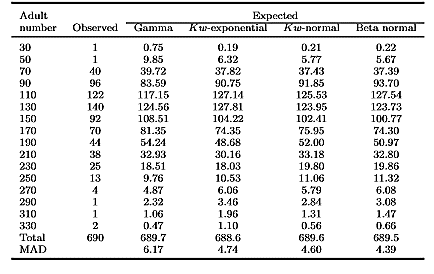


Figure 3. Histogram of adult number and fitted probability density functions.

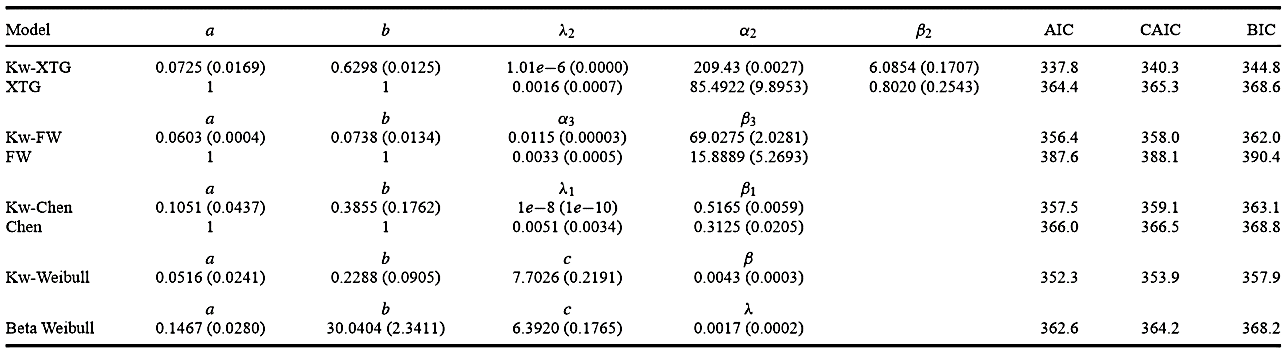
Table 2. Observed and expected frequencies of adult numbers for *T. confusum* cultured at 29°C and mean absolute deviation (MAD) between the frequencies



Based on the values of the LR statistic (Section 8), the *Kw-gamma* and the *Kw-exponential* distributions are not significantly different yielding *LR = 1.542 (1 d.f., p-value = 0.214).* Comparing the *Kw-gamma* and the *gamma* distributions, we find a significant difference (*LR* = *6.681, 2 d.f., p-value = 0.035).*

**12.2. Uncensored data (voltage):**

Here, we compare the results of Nadarajaha *et al* (2011). They ﬁts some distributions to a data set which gives the times of failure and running times for a sample of devices from a ﬁeld-tracking study of a larger system. At a certain point in time, 30 units were installed in normal service conditions. Two causes of failure were observed for each unit that failed: the failure caused by an accumulation of randomly occurring damage from power-line voltage spikes during electric storms and failure caused by normal product wear. The required numerical evaluations were implemented using the SAS procedure NLMIXED. Table 3 lists the MLEs (and the corresponding standard errors in parentheses) of the parameters and the values of the following statistics for some ﬁtted models: AIC (Akaike information criterion), BIC (Bayesian information criterion) and CAIC (Consistent Akaike information criterion). These results indicate that the Kw-Weibull model has the lowest AIC, CAIC and BIC values among all ﬁtted models, and so it could be chosen as the best model.



1. **Conclusion:**

Following the idea of the class of beta generalized distributions and the distribution by Kumaraswamy, we define a new family of *Kw* generalized (*Kw-G*) distributions to extend several widely-known distributions such as the normal, Weibull, gamma and Gumbel distributions. For each distribution G, we can define the corresponding *Kw-G* distribution using simple formulae.

We show how some mathematical properties of the Kw-G distributions are readily obtained from those of the parent distributions. The moments of the Kw-G distribution can be expressed explicitly in terms of infinite weighted sums of probability weighted moments (PWMs) of the G distribution. The same happens for the moments of order statistics and PWMs of the *Kw-G* distributions.

We discuss maximum likelihood estimation and inference on the parameters. The maximum likelihood estimation in *Kw-G* distributions is much simpler than the estimation in beta generalized distributions. Further, we can easily compute the maximum values of the unrestricted and restricted log-likelihoods to construct likelihood ratio statistics for testing nested models in the new family of distributions. An application of the new family to real data is given to show the feasibility of our proposal.

**References:**

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics.* **12**:171-178.

Barakat, H. M. and Abdelkader, Y. H. (2004). Computing the moments of order statistics from nonidentical random variables. *Statistical Methods and Applications.* **13**:15-26.

Barlow, R. E. and Proschan, F. (1975). *Statistical theory of reliability and life testing: probability models.* Holt, Rinehart and Winston, New York, London.

Cordeiroa, Gauss M. and Castrob, Mario de (2009). A new family of generalized distributions. *Journal of Statistical Computation & Simulation*. **00:** 1-17.

Eugene, N., Lee, C., and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics. Theory and Methods.* **31**:497-512.

Fletcher, S. C. and Ponnambalam, K. (1996). Estimation of reservoir yield and storage distribution using moments analysis. *Journal of Hydrology.* **182**: 259-275.

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 expressable in inverse form. *Water Resources Research*. **15**:1049-1054.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*. **Series B**. 52:105- 124.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion).*Test.* **13**:1-43.

Jones, M. C. (2008). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*. **6**:70-81.

Kumaraswamy, P. (1980). Generalized probability density-function for double-bounded random- processes. *Journal of Hydrology*. **462**:79-88.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1987). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York, London.

Nadarajaha, S. and Gupta, A. K. (2004). The beta Frechet distribution. *Far East Journal of Theoretical Statistics*. **14**:15-24.

Nadarajaha, S. and Kotz, S. (2006). The beta exponential distribution. *Reliability Engineering & System Safety.* **91**: 689-697.

Nadarajaha, S., Cordeirob, Gauss M. and Ortegac, Edwin M. M. (2011). General results for the Kumaraswamy-G distribution. *Journal of Statistical Computation and Simulation*. **iFirst**: 1-29.

Rigby, R. A. and Stasinopoulos, D. M.(2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics.* **54**:507-554.

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Sundar, V. and Subbiah, K. (1989). Application of double bounded probability density-function for analysis of ocean waves. *Ocean Engineering*. **16**:193-200.

Seifi, A., Ponnambalam, K. and Vlach, J. (2000). Maximization of manufacturing yield of systems with arbitrary distributions of component values. *Annals of Operations Research*. **99**:373- 383.

Stasinopoulos, D. M. and. Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R*. Journal of Statistical Software*. **23:**1-46.