Statistical Estimation Based on Generalized Order Statistics from Kumaraswamy Distribution

Conference Paper · June 2013

Samir Safi
Islamic University of Gaza

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013)
International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Statistical Estimation Based on Generalized Order Statistics from Kumaraswamy Distribution

Samir K. Safi
Department of Statistics, The Islamic University of Gaza, Palestine
(E-mail: **samirsafi@gmail.com**)

Rehab H. Al Sheikh Ahmed
1Statistician Consultant, Gaza, , Palestine
(E-mail: **rehab4ps@yahoo.com**)

Abstract. The Kumaraswamy distribution is similar to the Beta distribution but has the key advantage of a closed-form cumulative distribution function. In this paper we present the estimation of Kumaraswamy distribution parameters based on Generalized Order Statistics (GOS) using Maximum Likelihood Estimators (MLE). We proved that the parameters estimation for Kumaraswamy distribution can not be obtained in explicit forms, and therefore it has been implemented using the simulated data for illustrative purposes. We compare the performances of parameters estimation through an extensive numerical simulation for different sample sizes. These simulations examine the sensitivity of estimation to different sample sizes. In particular, how do estimations perform for small, moderate and large sample sizes? The main findings are: First, the worst performance estimation for small sample size selection for different values of the parameters estimation. Secondly, as the sample size increases the MSE of the estimation decreases. Finally, the estimation accuracy reaches its superiority for large sample sizes.

Keywords: Kumaraswamy Distribution, Generalized Order Statistics, Simulation, Maximum Likelihood Estimators.

1 Introduction

Poondni Kumaraswamy was a leading Indian engineer and hydrologist. Kumaraswamy[9] introduced the distribution for variables that are lower and upper bounded. With its two non-negative shape parameters p and q, it was originally conceived to model hydrological phenomena, (See for example Mitnik[10]). The Kumaraswamy distribution is a continuous probability distribution with double-bounded support, defined on the interval [0,1] differing in the values of their two non-negative shape parameters p and q. It is similar to the Beta distribution but has the key advantage of a closed form cumulative distribution function, Carrasco *et al.*[4].

Generalized Order Statistics (GOS) concept was introduced by Kamps[8] as a unified approach to several models of ordered random variables such as upper order statistics, upper record values, sequential order statistics, ordering via truncated distributions, censoring schemes, among others. Ateya and Ahmad[3], Jaheen[7], Habibullah and Ahsanullah[6], Raqab and Ahsanullah [11]among others, utilized the GOS in their works.

Abu El-Fotouh and Nassar[1] have investigated the estimation problem for the unknown parameters of Weibull extension model based on GOS by

Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013) International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

Maximum Likelihood Estimators (MLE). Alkasasbeh and Raqab[2] considered the MLE of the different parameters of a generalized logistic distribution and compared the performances of these procedures through an extensive numerical simulation.

This paper is organized as follows. Section 2 presents some basic definitions; Section 3 demonstrates the estimation of Kumaraswamy distribution parameters based on GOS Using MLE, the main results of this paper are stated and proved. Simulation study is shown in section 4; and Section 5 summarizes the important results.

2 Preliminaries

In this section, we introduce some basic definitions.

Definition 2.1 The random variables $X(1,n,\tilde{m},k),\ldots,X(n,n,\tilde{m},k)$ are called GOS based on the Cumulative Distribution Function *(cdf)*, $F(x)$, if their joint probability density function *(pdf)* is given by Kamps[8].

$$f(x_1,\ldots,x_n) = k\left(\prod_{j=1}^{n-1}\gamma_j\right)\left[\prod_{i=1}^{n-1}\left(1-F(x_i)\right)^{m_i} f(x_i)\right]\left(1-F(x_n)\right)^{k-1} f(x_n), \quad (2.1)$$

on the cone $F^{-1}(0) < X_1 \le X_2 \le \ldots \le X_n < F^{-1}(1)$ of $\mathbb{R}^n$, with parameters $n \in \mathbb{N}$, $n \ge 2$, $k > 0$, $\tilde{m} = (m_1,\ldots,m_{n-1}) \in \mathbb{R}^{n-1}$, $M_r = \sum_{j=r}^{n-1} m_j$, such that $\gamma_r = k + n - r + M_r > 0$ for all $r \in \{1,\ldots,n-1\}$, let $c_{r-1} = \prod_{j=1}^{r}\gamma_j$, $r = 1,2,\ldots,n-1$ and $\gamma_n = k$.

Special Case

Given Definition 2.1, let $k = 1$ and $m_1 = m_2 = \ldots = m_{n-1} = zero$, then the joint *pdf* of all Ordinary Order Statistics (OOS) is $f(x_1,\ldots,x_n) = \left[\prod_{j=1}^{n-1}\gamma_j\right]\left[\prod_{i=1}^{n} f(x_i)\right]$

$\left[\prod_{j=1}^{n-1}\gamma_j\right] = \prod_{j=1}^{n-1}\left(k + n - j + M_j\right)$, $M_j = \sum_{r=j}^{n-1} m_r = zero$. Then

$\left[\prod_{j=1}^{n-1}\gamma_j\right] = \prod_{j=1}^{n-1}(1 + n - j) = n(n-1)(n-2)\ldots 3\times 2\times 1 = n!$.

Therefore, the joint *pdf* of all OOS $f(x_1,\ldots,x_n) = n!\prod_{i=1}^{n} f(x_i)$, *which is the well known pdf of all OOS*.

Definition 2.2 The *pdf* of the Kumaraswamy distribution is given by

$$f_Z(z) = \frac{1}{(b-c)} p\, q \left(\frac{z-c}{b-c}\right)^{p-1} \left[1 - \left(\frac{z-c}{b-c}\right)^{p}\right]^{q-1} \quad , c < z < b, \qquad (2.2)$$

with shape parameters $p > 0$ and $q > 0$, and boundary parameters *c* and *b*. The standard form of the Kumaraswamy density function $(c = 0,\, b = 1)$, $kum\,(p,q)$ is given by

$$f_X\,(x\,;p,q) = p\, q\, x^{p-1}\left(1 - x^{p}\right)^{q-1} \qquad (2.3)$$

The closed form of the *cdf* of the Kumaraswamy distribution is given by

$$F(x\,;p,q) = 1 - \left(1 - x^{p}\right)^{q} \qquad (2.4)$$

Definition 2.3 The joint *pdf* of $X\,(1,n,\tilde{m},k), X\,(2,n,\tilde{m},k),\ldots,X\,(n,n,\tilde{m},k)$ for Kumaraswamy distribution is

$$f\,(x_1,\ldots,x_n) = k\left[\prod_{j=1}^{n-1}\gamma_j\right]\prod_{i=1}^{n-1}\left[\left(1-x_i^{\,p}\right)^{q\,m_i}\, p\, q\, x_i^{\,p-1}\left(1-x_i^{\,p}\right)^{q-1}\right]\left(1-x_n^{\,p}\right)^{q(k-1)} \times$$
$$p\, q\, x_n^{\,p-1}\left(1-x_n^{\,p}\right)^{q-1} \qquad (2.5)$$

3. Estimation of Kumaraswamy Distribution Parameters Based on GOS Using MLE

The method of MLE is, by far, the most popular technique for deriving estimators and a reasonable choice for an estimator. The MLE is the parameter point for which the observed sample is most likely. In general, the MLE is and a good point estimator, possessing some of the optimality properties such as invariance property of MLE, Casella and Berger[5]. In this section, the estimation of Kumaraswamy distribution parameters based on GOS using MLE will be derived. Furthermore, the estimation of Kumaraswamy distribution parameters based on OOS will be derived as special case when $k = 1$ and $m = 0$.

Theorem 3.1 Let $X\,(1,n,\tilde{m},k), X\,(2,n,\tilde{m},k),\ldots,X\,(n,n,\tilde{m},k)$ be *n* GOS for Kumaraswamy with parameters *p* and *q*, i.e. *X* has *kum(p,q)*. The estimation of Kumaraswamy distribution parameters based on GOS for *p* and *q* are given by

$$\hat{p} = -n\left[\sum_{i=1}^{n-1}\left(\frac{\ln x_i\left(1-\hat{q}\left(m_i+1\right)x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right) + \frac{\ln x_n\left(1-qkx_n^{\hat{p}}\right)}{1-x_n^{\hat{p}}}\right]^{-1} \qquad (3.1)$$

and

$$\hat{q} = -n\left[\sum_{i=1}^{n-1}(m_i+1)\ln\left(1-x_i^{\hat{p}}\right)+k\,\ln\left(1-x_n^{\hat{p}}\right)\right]^{-1}, \qquad (3.2)$$

respectively.

Proof. The likelihood function for GOS for *kum(p,q)* is defined by

$$L(p,q;x) = k\left[\prod_{j=1}^{n-1}\gamma_j\right]\prod_{i=1}^{n-1}\left[\left(1-x_i^{\,p}\right)^{qm_i}\; p\,q\,x_i^{\,p-1}\left(1-x_i^{\,p}\right)^{q-1}\right]\left(1-x_n^{\,p}\right)^{q(k-1)}\times$$
$$p\,q\,x_n^{\,p-1}\left(1-x_n^{\,p}\right)^{q-1}$$

Collecting terms, $L(p,q;x)$ can be written as

$$L(p,q;x) = k\left[\prod_{j=1}^{n-1}\gamma_j\right]\left[\prod_{i=1}^{n-1}(1-x_i^{\,p})^{q(m_i+1)-1}pqx_i^{\,p-1}\right](1-x_n^{\,p})^{qk-1}pqx_n^{\,p-1} \qquad (3.3)$$

While this function in (3.3) is not all that hard to differentiate, it is much easier to differentiate the log likelihood. Now take logarithm on both sides of (3.3) to get

$$ln\left[L\left(p,q;x\right)\right] =$$
$$ln\,k+\sum_{j=1}^{n-1}ln\,\gamma_j+\sum_{i=1}^{n-1}\left[\left(q\left(m_i+1\right)-1\right)ln\left(1-x_i^{\,p}\right)+ln\,p+ln\,q+\left(p-1\right)ln\,x_i\right]$$
$$+\left(qk-1\right)ln\left(1-x_n^{\,p}\right)+ln\,p+ln\,q+\left(p-1\right)ln\,x_n$$

Take the first partial derivatives, with respect to p and q, and collecting terms, we find that

$$\frac{\partial ln\left[L(p,q;x)\right]}{\partial p} = \sum_{i=1}^{n-1}\left[\frac{\ln x_i\left(1-q\left(m_i+1\right)x_i^{\,p}\right)}{1-x_i^{\,p}}\right]+\frac{\ln x_n\left(1-qkx_n^{\,p}\right)}{1-x_n^{\,p}}+\frac{n}{p}, \text{ and}$$

$$\frac{\partial ln\left[L(p,q;x)\right]}{\partial q} = \sum_{i=1}^{n-1}\left[\left(m_i+1\right)\ln\left(1-x_i^{\,p}\right)\right]+k\,\ln(1-x_n^{\,p})+\frac{n}{q},$$

respectively.

Setting these first partial derivatives equal to zero and solving for p and q, yield the solution

$$\hat{p} = -n\left[\sum_{i=1}^{n-1}\left(\frac{\ln x_i\left(1-\hat{q}\left(m_i+1\right)x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right)+\frac{\ln x_n\left(1-qkx_n^{\hat{p}}\right)}{1-x_n^{\hat{p}}}\right]^{-1}, \text{ and}$$

$$\hat{q} = -n\left[\sum_{i=1}^{n-1}(m_i+1)\ln\left(1-x_i^{\hat{p}}\right)+k\,\ln\left(1-x_n^{\hat{p}}\right)\right]^{-1},$$

respectively.

Evaluating the second derivative at $p = \hat{p}$ and $q = \hat{q}$ yield $\left.\frac{\partial^2 ln\left[L(p,q;x)\right]}{\partial p^2}\right|_{p=\hat{p}} = -\frac{n}{\hat{p}^2} < 0$, and $\left.\frac{\partial^2 ln\left[L(p,q;x)\right]}{\partial q^2}\right|_{q=\hat{q}} = -\frac{n}{\hat{q}^2} < 0$, then each of $\hat{p}$ and $\hat{q}$ is the local maximum, and since they are the only values obtained when the first partial derivatives are equal to zero, then $\hat{p}$ and $\hat{q}$ are the global maximum for the likelihood function $ln\left[L\left(p,q;x\right)\right]$. This completes the proof of the Theorem. ⊔

Corollary 3.1 The estimation of Kumaraswamy distribution parameters based on OOS for *p* and *q* are given by

$$\hat{p} = -n\left(\sum_{i=1}^{n}\left[\frac{\ln x_i\left(1-\hat{q}x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right]\right)^{-1}, \qquad (3.4)$$

and

$$\hat{q} = -n\left(\sum_{i=1}^{n}\ln\left(1-x_i^{\hat{p}}\right)\right)^{-1}, \qquad (3.5)$$

respectively.

Proof. Let $m = 0$ and $k = 1$ in (3.1), then $\hat{p} = -n\left[\sum_{i=1}^{n-1}\left[\frac{\ln x_i\left(1-\hat{q}x_i^{\hat{p}}\right)}{1-x_i^{\hat{p}}}\right] + \frac{\ln x_n\left(1-\hat{q}x_n^{\hat{p}}\right)}{1-x_n^{\hat{p}}}\right]^{-1}$ and collecting terms, we get (3.4).

Let $m = 0$ and $k = 1$ in (3.2), then $\hat{q} = -n\left[\sum_{i=1}^{n-1}\ln\left(1-x_i^{p}\right) + \ln\left(1-x_n^{\hat{p}}\right)\right]^{-1}$ and collecting terms, we get (3.5). This completes the proof of the Corollary. ⊔

Equations (3.1), (3.2), (3.4) and (3.5) are complicated and consequently computer facilities and numerical solutions are needed to compute $\hat{p}$ and $\hat{q}$.

4 Simulation Study

In this section, since there are no closed forms for the estimation of Kumaraswamy distribution parameters, we consider the simulation technique for the estimation of Kumaraswamy distribution parameters p and q for different sample sizes. These simulations examine the sensitivity of estimation to different sample sizes. In particular, how do estimations perform for small, moderate and large sample sizes?

Definition 1. The efficiency of the parameter estimation for sample size n_1 relative to that of n_2 in terms of the Mean Squared Error (MSE) of the parameter p, $RE(\hat{p})$, is given by

$$RE(\hat{p}) = \frac{n_2 \sum_{i=1}^{r} \left(\hat{p}_{i.n_1} - p_{n_1}\right)^2}{n_1 \sum_{i=1}^{r} \left(\hat{p}_{i.n_2} - p_{n_2}\right)^2}, \qquad (4.1)$$

where r represents the number of simulations. Note that, p_{n_1} and p_{n_2} are the true parameters values for the two samples sizes n_1 and n_2, respectively. A ratio greater than one indicates that the parameter estimation for the sample size n_1 is less efficient than sample size n_2 estimate, and if $RE(\hat{p})$ is close to one, then the parameter estimation for the sample size n_1 is nearly as efficient as sample size n_2 estimate. We will look for different pairs of parameters for Kumaraswamy distribution that we can use to characterize the efficiency ratio, such as $p = q = 1$, $p = 1, q = 2$, $p = 2, q = 1$ and $p = q = 2$. We will try to find an answer to the following question: How robust are Kumaraswamy parameter estimations for different sample sizes?

4.1 The Simulation Setup

Three finite sample sizes (50, 200, and 500) and four values for the parameters p and q. We also generated a simulation of length 500 observations for each of the selected parameters; (p,q): (1,1), (1,2), (2,1) and (2,2).

4.2 The Simulation Results for $RE(\hat{p})$

Table (4.1) shows the complete simulation results for all selected parameters for Kumaraswamy distribution; (p,q): (1,1), (1,2), (2,1) and (2,2) for three finite sample small, moderate and large sizes (50, 200, and 500). The estimated values of the parameters and their corresponding MSEs are given. In addition, the ratios of parameter estimation p, $RE(\hat{p})$, for sample size n_1 relative to that of n_2 in terms of the mean squared error are shown.

Looking at the Table (4.1), we see that for the parameter $p = 1$, the relative efficiency of the MSE for estimating the parameter $p = 1$ with sample size 50 with respect to 200 equals 19.3. This means the parameter estimation error for small sample size ($n = 50$) is about 19 times for moderate sample size ($n = 200$). While the relative efficiency of the MSE for estimating the parameter $p = 1$ with sample size 50 with respect to 500 equals 113.0. This means the parameter estimation error for small sample size ($n = 50$) is about 113 times for large sample size ($n = 500$). This result is the worst performance of small sample size selection as compared to large sample size for estimating the parameter p. In addition, the relative efficiency of the MSE for estimating the parameter $p=1$ with sample size 200 with respect to 500 equals 5.9. This means the parameter estimation error for moderate sample size ($n=200$) is about 6 times for large sample size ($n=500$). This result indicates that as the sample size increases the MSE of the estimated parameters decreases. This indicates that the MLE tend to its true parameters values. In other words, the estimation accuracy reaches its superiority as the sample size gets larger and larger. Results for the other sample sizes and different parameter choices for Kumaraswamy distribution demonstrate a similar pattern as shown in Table (4.1).

Table 4.1. Estimation of the Parameters for Kumaraswamy Distribution p and q for Different Sample Sizes

Parameters	Sample Size	$\hat{p}$	$MSE(\hat{p})$	$RE(\hat{p})$	$\hat{q}$	$MSE(\hat{q})$	$RE(\hat{q})$
p = q =1	50	1.00115	0.000379	19.3*	0.988747	0.000166691	6.2*
	200	0.994804	0.000020	113.0**	0.988783	0.000027	37.4**
	500	0.994319	0.000003	5.9***	0.985543	0.000004	6.1***
p = 1 , q = 2	50	0.852075	0.002560	17.9*	1.447103	0.020512	27.2*
	200	0.945802	0.000143	387.6**	1.767986	0.000755	173.9**
	500	0.971881	0.000007	21.7***	1.877155	0.000118	6.4***
p = 2 , q =1	50	1.638147	0.005442	7.9*	0.931092	0.000379	5.1*
	200	1.660116	0.000692	29.7**	0.893285	0.000074	9.6**
	500	1.718804	0.000183	3.8***	0.91861	0.000039	1.9***
p = q =2	50	1.320175	0.019067	50.4*	1.291178	0.050228212	77.1*
	200	1.852513	0.000379	299.5**	1.829603	0.000652	462.1**
	500	1.940557	0.000064	5.9***	1.894831	0.000109	6.0***

* RE estimate of n_1=50 relative to n_2=200.
**RE estimate of n_1=50 relative to n_2=500.
*** RE estimate of n_1=200 relative to n_2=500.

5. Conclusions

This paper deals with the estimation of Kumaraswamy distribution parameters using maximum likelihood estimators. Statistical estimation of Kumaraswamy distribution parameters have been derived based on generalized order statistics. Special cases are also deduced for ordinary order statistics. The resulting equations are complicated and numerical solutions for parameters p and q is recommended. The simulation technique is discussed.

References

1. S. Abu El-Fotouh and M. Nassar. Estimation for the Parameters of the WeibullExtension Model Based on Generalized Order Statistics. *Int. J. Contemp. Math. Sciences*, 6, 36, 1749 – 1760, 2011.
2. M. Alkasasbeh and M. Raqab. Estimation of the Generalized Logistic Distribution Parameters: Comparative Study. *Statistical Methodology*, 6, 3, 262-279, 2009.
3. S. Ateya and A. Ahmad. Inferences based on generalized order statistics under truncated type I generalized logistic distribution, *Statistics,* 45, 389-402, 2011.
4. J. Carrasco, S. Ferrari and G. Cordeiro. A New Generalized Kumaraswamy Distribution, 2010.
5. G. Casella and R. Berger. *Statistical Inference,* second edition, Duxbury, California, 2002.
6. M. Habibullah and M. Ahsanullah. Estimation of parameters of a Pareto distribution by generalized order statistics, *Commun. Statist.-Theor. Meth., 29, 1597-1609,* 2000.
7. Z. Jaheen. Estimation based on generalized order statistics from the Burr model, *Commun. Statist.- Theor. Meth.,* 34, 785-794, 2005.
8. U. Kamps. A Concept of Generalized Order Statistics, *ELSEVIER Journal of Statistical planning and inference,* 48, 1995.
9. P. Kumaraswamy. A generalized probability density function for double bounded random processes. *Journal of Hydrology*. 46, 79-88, 1980.
10. P. Mitnik. New Facts About Kumaraswamy Distribution. Department of Sociology, University of Wisconsin-Madison, USA, 2008.
11. M. Raqab and M. Ahsanullah. Estimation of the location and scale parameters of the generalized exponential distributions based on order statistics. *J. Statist. Comput. Simul.* 69, 109- 123, 2001.