

Estimation of parameters of the Kumaraswamy distribution based on general progressive type II censoring

Mostafa Mohie Eldin¹, Nora Khalil², Montaser Amein³

¹Professor in department of Mathematics, faculty of Science, El Azhar University, Cairo, Egypt

²Lecturer department of Mathematics, faculty of Science, Helwan University, Cairo, Egypt

³Lecturer department of Mathematics, faculty of Science, El Azhar University, Cairo, Egypt

Email address:

Nora.hassan194@gmail.com (N. Khalil)

To cite this article:

Mostafa Mohie Eldin, Nora Khalil, Montaser Amein. Estimation of Parameters of the Kumaraswamy Distribution Based on General Progressive Type II Censoring. *American Journal of Theoretical and Applied Statistics*. Vol. 3, No. 6, 2014, pp. 217-222.

doi: 10.11648/j.ajtas.20140306.17

Abstract: In this paper, we produced a study in Estimation for parameters of the Kumaraswamy distribution based on general progressive type II censoring. These estimates are derived using the maximum likelihood and Bayesian approaches. In the Bayesian approach, the two parameters are assumed to be random variables and estimators for the parameters are obtained using the well known squared error loss (SEL) function. The findings are illustrated with actual and computer generated data.

Keywords: Kumaraswamy's Distribution, Bayes Estimation, Bayes Prediction, General Progressive Type II Censoring

1. Introduction

The Kumaraswamy distribution is very similar to the Beta distribution, but has the important advantage of an invertible closed form cumulative distribution function. Kumaraswamy (1976, 1978) has showed that the well known probability distribution functions such as the normal, log-normal, beta and empirical distributions such as Johnson's and polynomial-transformed-normal, etc., do not fit well hydrological data, such as daily rainfall, daily stream flow, etc. and developed a new probability density function known as the sine power probability density function. Furthermore, Kumaraswamy (1980) developed a more general probability density function for double bounded random processes, which is known as Kumaraswamy's distribution. This distribution is applicable to many natural phenomena whose outcomes have lower and upper bounds, such as the heights of individuals, scores obtained on a test, atmospheric temperatures, hydrological data, etc. Also, this distribution could be appropriate in situations where scientists use probability distributions which have infinite lower and/or upper bounds to fit data, when in reality the bounds are finite. Gupta and Kirmani (1988) discussed the connection between non-homogeneous Poisson process (NHPP) and record values. So, the results in this paper may be used for some applied situation such as preventive maintenance.

The paper by Kumaraswamy (1980) proposed a new probability distribution for double bounded random processes with hydrological applications. The continuous part of Kumaraswamy's distribution has the probability density function (pdf) and the cumulative distribution function (cdf) specified by

$$f(x) = pqx^{p-1}(1-x^p)^{q-1} \quad (1)$$

and

$$F(x) = 1 - (1-x^p)^q \quad (2)$$

respectively, for $0 \leq x \leq 1$, $p > 0$ and $q > 0$. Kumaraswamy (p,q), where p and q are two positive shape parameters. It has many of the same properties as the beta distribution but has some advantages in terms of tractability. This distribution appears to have received considerable interest in hydrology and related areas, see Sundar and Subbiah (1989), Fletcher and Ponnambalam (1996), Seifi et al. (2000), Ponnambalam et al. (2001), and Ganji et al. (2006).

Kumaraswamy distributions are special cases of the three parameter distribution with density

$$\frac{p}{B(\gamma, q)} x^{\gamma p - 1} (1 - x^p)^{q-1}, 0 \leq x \leq 1, q \text{ and } p > 0.$$

Kumaraswamy distributions has special cases, Kumaraswamy (p,1) distribution is the power function distribution, Kumaraswamy (1,p) distributions is the distribution of one minus that power function random variable. Kumaraswamy (1,1) distribution is the uniform distribution.

A further special case of the Kumaraswamy distribution has also appeared elsewhere. The Kumaraswamy (2,q) distribution is that of the" generating variate $R = \sqrt{x_1^2 + x_2^2}$ when $\{x_1, x_2\}$ Follow a bivariate Pearson Type II distribution [K.T. Fang(1990)], Section 3.4.1).

2. Parameter Estimation

Suppose n independent units are placed on life test with corresponding failure times (y1, y2..., ym) being independently distributed with cumulative distribution

$$L(\theta) = c^* [F(y_{r+1}, \theta)]^r \prod_{i=r+1}^m f(y_i, \theta) [1 - F(y_i, \theta)]^{R_i}, y_{r+1} < y_{r+2} < \dots < y_m \quad (3)$$

where,

$$c^* = \frac{n!}{r!(n-r-1)!} (n - R_{r+1} - r - 1) \dots (n - R_{r+1} - \dots - R_{m-1} - m + 1)$$

Using eq. (1) the likelihood function can be expressed as:

$$L(t; p, q) = c^* \left(1 - (1 - y_{r+1}^p)^q\right)^r \prod_{i=r+1}^m pq y_i^{p-1} (1 - y_i^p)^{q-1} \left((1 - y_i^p)^q\right)^{R_i} = c^* (pq)^{m-r} \left(1 - (1 - y_{r+1}^p)^q\right)^r \prod_{i=r+1}^m y_i^{p-1} (1 - y_i^p)^{q(R_i+1)-1} \quad (4)$$

The log-likelihood function $\ell(y; p, q)$, is given by:

$$\ell(y; p, q) = \text{constant} + (m-r) \text{Log}(pq) + r \text{Log}\left(1 - (1 - y_{r+1}^p)^q\right) + (p-1) \sum_{i=r+1}^m \text{Log}(y_i) + \sum_{i=r+1}^m (q(R_i+1)-1) \text{Log}(1 - y_i^p) \quad (5)$$

The ML estimates of p and q can be obtained by maximizing the log-likelihood function, Eq. 5. By taking

derivatives with respect to p and q the MLE are obtained by satisfying the following equations

$$\frac{\partial \ell}{\partial p} = \frac{m-r}{p} + \frac{qr \text{Log}(y_{1+r}) y_{1+r}^p (1 - y_{1+r}^p)^{-1+q}}{1 - (1 - y_{1+r}^p)^q} + \sum_{i=1+r}^m \text{Log}(y_i) - \sum_{i=1+r}^m \text{Log}(y_i) (-1 + q(1 + R_i)) y_i^p = 0$$

$$\frac{\partial \ell}{\partial q} = \frac{m-r}{q} - \frac{r \text{Log}(1 - y_{1+r}^p) (1 - y_{1+r}^p)^q}{1 - (1 - y_{1+r}^p)^q} + \sum_{i=1+r}^m (1 + R_i) (1 - y_i^p) = 0$$

The above system is nonlinear, but can be easily solved using numerical techniques.

2.2. Bayes Estimation of the Unknown Parameter(s)

In this subsection we consider the Bayes estimation of the unknown parameter(s). We will assume that the parameters p and q of the Kumaraswamy's distribution are random variables with a joint bivariate prior density function that was first suggested by Al-Hussaini and Jaheen (1995) as,

function F(y), probability density function f(y), and m prefixed numbers of failures that observed. The scheme (R1, ..., Rm) is the progressively type II censored, $y_{i;m;n}$ be completely observed times, $i=1, 2, \dots, m$. The failure times of the first r units were not observed, at time (r+1)th failure R_{r+1} number of surviving units are withdrawn from the test randomly, and so on. Finally at the time mth failure, the remaining $R_m = n - m - R_{r+1} - R_{r+2} - \dots - R_{m-1}$ are withdrawn from the test.

2.1. ML Estimation

The likelihood function to be maximized when general progressive type II censored sample based on n independent units with F(y) is L(θ).

$$\pi(p, q) = g_1(q|p) g_2(p) \quad (6)$$

where

$$g_1(q|p) = \frac{p^{\alpha+1}}{\Gamma(\alpha+1)} q^{\alpha} e^{-pq/\gamma}, \alpha > -1, \gamma > 0 \quad (7)$$

is the gamma conjugate prior. This prior was first introduced by Papadopoulos (1978) and was also used later on by Al-Hussaini and Jaheen (1992). The prior of p is

$$g_2(p) = \frac{p^{\delta-1}}{\Gamma(\delta) \beta^{\delta}} e^{-p/\beta}, \delta > 0, \beta > 0 \quad (8)$$

which is the gamma (δ, β) density.

Substituting (6) and (7) in (5) we obtain the bivariate prior density of p and q given as

$$\pi(p, q) \propto p^{\delta+\alpha} q^{\alpha} \exp(-p(1/\beta + q/\gamma)) \quad (9)$$

where $\alpha > -1$, β, δ and γ are positive real numbers.

From (3) and (9), the joint posterior distribution is given by

$$\pi(p, q | data) = \frac{p^{m+\alpha+\delta-r} q^{m+\alpha-r} e^{r \ln(1-(1-y_{r+1}^p)^q)} e^{p \sum_{i=r+1}^m \ln(y_i) + \sum_{i=r+1}^m (q(1+R_i)-1) \ln(1-y_i^p) - p(q/\gamma + 1/\beta)}}{\int_0^\infty \int_0^\infty p^{m+\alpha+\delta-r} q^{m+\alpha-r} e^{r \ln(1-(1-y_{r+1}^p)^q)} e^{p \sum_{i=r+1}^m \ln(y_i) + \sum_{i=r+1}^m (q(1+R_i)-1) \ln(1-y_i^p) - p(q/\gamma + 1/\beta)} dp dq}$$

$$\pi(p, q | data) = \frac{p^{m+\alpha+\delta-r} q^{m+\alpha-r} e^{r \ln(1-(1-y_{r+1}^p)^q)} e^{p \sum_{i=r+1}^m \ln(y_i) + \sum_{i=r+1}^m (q(1+R_i)-1) \ln(1-y_i^p) - p(q/\gamma + 1/\beta)}}{\Gamma(m+\alpha-r+1) \psi(0, 1, 0, 0)} \quad (10)$$

Where

$$\psi(a, c, d, f) = \int_0^\infty \sum_{j=0}^r (-1)^j \binom{r}{j} \frac{z^{m+\alpha+\delta-r+a} e^{-z \left(d - \sum_{i=r+1}^m \ln(y_i) - \sum_{i=r+1}^m \ln(1-y_i^z) \right)}}{\left(z/\gamma - \sum_{i=r+1}^m (1+R_i) \ln(1-y_i^z) - j \ln(1-y_{r+1}^z) + f \right)^{m+\alpha-r+c}} dz$$

If the loss function is the well known squared error loss function, the Bayes estimators for the parameters p and q are the given by their marginal posterior expectations as

$$\hat{p}_B = E(p | y) = \frac{\psi(1, 1, 0, 0)}{\psi(0, 1, 0, 0)}$$

And

$$\hat{q}_B = E(q | y) = (m + \alpha - r + 1) \frac{\psi(0, 2, 0, 0)}{\psi(0, 1, 0, 0)}$$

3. Numerical Examples

A simulation study with real data is conducted in order to compare the performances between MLE and Bayes estimator.

3.1. Illustrative Examples

Example 1(real data)

In order to illustrate the findings of Sects. 2 and 3 two examples are given. The former is a real data set obtained from the Shasta reservoir in California while the latter example uses simulated data set. In both examples the mathematical package was used to obtain the estimates of the parameters a and b and Bayes predictive estimates.

The first example deals with the monthly water capacity data from the Shasta reservoir in California, USA and were

taken for the month of February from 1991 to 2010, http://cdec.water.ca.gov/reservoir_map.html. The maximum capacity of the reservoir is 4552000 AF and the data were transformed to the interval $[0, 1]$.

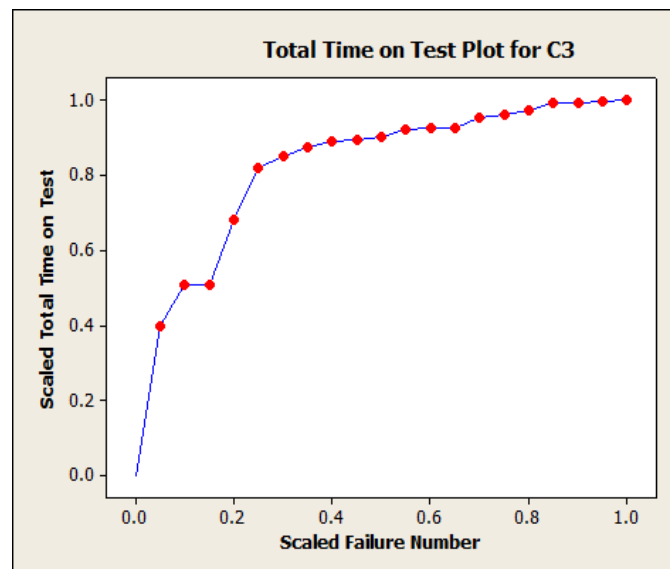


Fig. 1. The empirical hazard function.

Table 1, gives the date, actual and transformed data.

Table 1. Monthly capacity for August and proportion of total capacity for Shasta reservoir.

Year	capacity	Proportion of total capacity	Year	capacity	Proportion of total capacity
1991	1542838	0.338936	2001	3495969	0.768007
1992	1966077	0.431915	2002	3839544	0.843485
1993	3459209	0.759932	2003	3584283	0.787408
1994	3298496	0.724626	2004	3868600	0.849868
1995	0.759932	0.757583	2005	3168056	0.695970
1996	3694201	0.811556	2006	3834224	0.842316
1997	3574861	0.785339	2007	3772193	0.828689
1998	3567220	0.783660	2008	2641041	0.580194
1999	3712733	0.815627	2009	1960458	0.430681
2000	3857423	0.847413	2010	3380147	0.742563

The 20 values were used to verify that the transformed data follow Kumaraswamy's distribution, by examine the empirical hazard function of the observed data by applying the scaled Total Time on Test (TTT) plot, see Aarset (1987). This provides a very good idea about the shape of the hazard function of a distribution. For a family with the survival function $s(y) = 1 - F(y)$, The scaled TTT with

$$H^{-1}(u) = \int_0^{F(u)} s(y) dy \text{ define for } 0 < u < 1 \text{ is } g(u) = \frac{H^{-1}(u)}{H^{-1}(1)}.$$

The corresponding empirical version of the scaled TTT

$$\text{transform is given by } g_n(r|n) = \frac{H_n^{-1}(r|n)}{H_n^{-1}(1)} = \frac{\left(\sum_{i=1}^r y_{(i)} + (n-r)y_{(r)} \right)}{\sum_{i=1}^n y_{(i)}},$$

Failure time vector $Y = (y_1, \dots, y_n)$ and censoring scheme $R = (R_1, \dots, R_r), r=3$.

i	1	2	3	4	5	6	7	8	9
Y_i	-	-	-	0.430681	0.431915	0.580194	0.695970	0.724626	0.785339
R_i	-	-	-	1	2	0	4	2	2

MLE, posterior mean, median and mode of the parameter

Parameter	MLE	Posterior mean	Posterior mode	Posterior median
p	3.38005	3.2005	3.16566	3.1436
q	1.34034	1.15226	1.09477	1.0563

Example 2 (generation data)

The following data is general progressively type II censored sample from Kumaraswamy distribution ($p = 3, q = 2$) was simulated by using $n=60, r=9, m=29$ and censored scheme $[R_i = \{4, 1, 1, 1, 1, 3, 1, 0, 3, 0, 2, 0, 3, 2, 3, 0, 0, 3, 2, 1\},$

$i=r+1, \dots, m]$: 0.4520, 0.4654, 0.4702, 0.4925, 0.5052, 0.5182, 0.5284, 0.5528, 0.5654, 0.5890, 0.6274, 0.6402, 0.6696, 0.6812, 0.7056, 0.7092, 0.7323, 0.7566, 0.7871, 0.8223.

MLE, posterior mean, median and mode of the parameter

Parameter	MLE	Posterior mean	Posterior mode	Posterior median
p	3.08216	3.0326	2.88523	2.7625
q	1.8516	1.6276	1.60599	1.4276

3.2. Monte Carlo Simulations

In the following, Monte Carlo simulation study is conducted in order to compare the performance of the Bayes estimator with MLE for different sample sizes and censoring schemes. We generate progressively type II censored samples

from Kumaraswamy distribution with parameter $p=3, q=2$, different combinations of the effective sample size $m' = (m-r)$ and different progressive censoring schemes $(R_{r+1}, R_{r+2}, \dots, R_m)$. The different progressive censoring schemes that used were I) random censoring schemes, II)

$\left(\frac{m'-1}{2} * 0, n-m, \frac{m'-1}{2} * 0\right)$ and III) $(m'-1 * 0, n-m)$. For

simplicity in notation, we denote these censoring schemes, for example, by $(3*0,1,5)$ which represents the censoring scheme $(R_4=0, R_5=0, R_6=0, R_7=1, R_8=5)$ for fixed $r=3$. The programs were written by Mathematica 9. In computing the estimates we generated 1000 samples from the

Kumaraswamy distribution, and we replicated the process 1000 times. The averages and mean squared errors (MSE) in parentheses of estimators of p and q are presented in Tables 2 and 3, respectively. For prior information we have used prior1 with $\alpha=1, \gamma=2, \delta=1, \beta=1$.

Table 2. Average estimates of p and the associated MSEs when $r=3$.

n	m	Effective sample size m'	$(R_{r+1}, R_{r+2}, \dots, R_m)$ Censoring schemes	MLE	Bayes Prior1
20	10	7	(6*0,10)	3.1131 (0.2662)	3.18885 (0.3062)
			(3*0,10,3*0)	3.0516 (0.3159)	3.09802 (0.2809)
			(0, 2*1,2,1,3)	3.1593 (0.3305)	3.2018 (0.3053)
	12	9	(8*0,8)	3.02529 (0.1200)	2.9806 (0.2230)
			(4*0,8, 4*0)	3.09034 (0.3073)	3.17932 (0.2960)
			(2,0,2,0,2,0,2,2*0)	3.10669 (0.3453)	3.19834 (0.3526)
30	10	7	(6*0,20)	3.01308 (0.17909)	3.17069 (0.18989)
			(3*0,20, 3*0)	3.04555 (0.19109)	3.18085 (0.18782)
			(5,0,5,0,5,0,5)	3.03692 (0.18950)	3.16 (0.1933)
	12	9	(8*0,18)	3.04414 (0.1761)	3.22823 (0.2244)
			(4*0,18, 4*0)	3.02868 (0.211869)	3.18064 (0.2013)
			(16*0,10)	3.01585 (0.13514)	3.05194 (0.1363)
	20	17	(8*0,10, 8*0)	3.00253 (0.1609)	3.01027 (0.1556)
			(2,0,3*1,6*0,2*1,3*0,3)	3.00044 (0.1531)	3.20177 (0.2317)

Table 3. Average estimates of q and the associated MSEs when $r=3$.

n	m	Effective sample size m'	$(R_{r+1}, R_{r+2}, \dots, R_m)$ Censoring schemes	MLE	Prior1
20	10	7	(6*0,10)	2.13186 (0.17392)	2.0108 (0.2042)
			(3*0,10,3*0)	2.15666 (0.22698)	2.0717 (0.2024)
			(0, 2*1,2,1,3)	2.1697 (0.2462)	2.0343 (0.2133)
	12	9	(8*0,8)	2.06577 (0.15831)	2.06063 (0.1782)
			(4*0,8, 4*0)	2.11246 (0.2005)	2.08378 (0.2572)
			(2,0,2,0,2,0,2,2*0)	2.11698 (0.1766)	2.08138 (0.25484)
30	10	7	(6*0,20)	2.02441 (0.2238)	2.12304 (0.2782)
			(3*0,20, 3*0)	2.0848 (0.2157)	2.09572 (0.2142)
			(5,0,5,0,5,0,5)	2.07875 (0.2237)	2.1284 (0.31579)
	12	9	(8*0,18)	2.00134 (0.178693)	2.12667 (0.302296)
			(4*0,18, 4*0)	2.08847 (0.200834)	2.15622 (0.283017)
			(4,0,4,0,4,0,4,0,2)	2.04804 (0.1069)	1.92063 (0.1241)
	20	17	(16*0,10)	2.05152 (0.1399)	1.9915 (0.1267)
			(8*0,10, 8*0)	2.06452 (0.1571)	2.2525 (0.3071)
			(2,0,3*1,6*0,2*1,3*0,3)		

4. Conclusion

In this paper we have considered Estimation problems for parameters of the Kumaraswamy based on progressive Type-II censored data by using Maximum Likelihood Estimation and the Bayesian inference. The prior belief of the model is represented by the independent gamma conjugate priors on the shape parameters. The squared error loss function is used as it is appropriate when large errors of the estimation are considered to be more serious compared to other loss functions. We used Mathematica program to generate the samples and bisection method to compute the estimated parameters. The details have been explained using a real life example and computer generated data. Simulation study was conducted in order to compare the performance of the Bayes estimators with MLE for different sample sizes and censoring schemes, it was seen that the behavior of the MLE and Bayes estimates depends on the kind of censoring schemes, and the results were summarized as follow:

- i. Comparing the behavior of MLE and Bayes estimators: where when used random censoring schemes the MLE and Bayes estimators were almost same values in terms of MSE; however, it was seen that the Bayes estimators were better than their corresponding MLEs when used the following censoring schemes $\left(\frac{m'-1}{2} * 0, n-m, \frac{m'-1}{2} * 0\right)$ and the MLEs were better than their corresponding the Bayes estimators when used the following censoring schemes $(m'-1 * 0, n-m)$. An important problem will be to extend these results for other censoring schemes such as Type-I, hybrid censoring schemes. The work is in progress.
- ii. behavior of the MLE: the behavior of MLE in censoring scheme number I was better than number II, III.
- iii. behavior of Bayes estimators: the different kind of censoring schemes behavior of does not effect on Bayes estimators behavior.

References

- [1] Aarset, M.V (1987) How to identify a bathtub hazard function. IEEE Transactions on Reliability 36: 106–108.
- [2] Al-Hussaini EK, Jaheen ZF (1992) Bayesian estimation of the parameters, reliability and failure rate functions of the Burr Type XII failure model. J Stat Comput Simul 41:37–40
- [3] Chansoo Kim, Keunhee Han (2009) Estimation of the scale parameter of the Rayleigh distribution under general progressive censoring. Journal of the Korean Statistical Society 38: 239_246
- [4] Fletcher SC, Ponnambalam K (1996) Estimation of reservoir yield and storage distribution using moments analysis. J Hydrol 182:259–275
- [5] Gupta RC, Kirmani SNUA (1988) Closure and monotonicity properties of nonhomogeneous Poisson processes and record values. Probab Eng Inf Sci 2:475–484
- [6] K.T. Fang, S. Kotz, K.W. Ng, Symmetric Multivariate and Related Distributions, Chapman and Hall, London, 1990.
- [7] Kumaraswamy P (1976) Sinpower probability density function. J Hydrol 31:181–184
- [8] Kumaraswamy P (1978) Extended sinpower probability density function. J Hydrol 37:81–89
- [9] Kumaraswamy P (1980) A generalized probability density function for double-bounded random processes. J Hydrol 46:79–88
- [10] Ponnambalam K, Seifi A, Vlach J (2001) Probabilistic design of systems with general distributions of parameters. Int J Circuit Theory Appl 29:527–536
- [11] Seifi A, Ponnambalam K, Vlach J (2000) Maximization of manufacturing yield of systems with arbitrary distributions of component values. Ann Oper Res 99:373–383
- [12] Sundar V, Subbiah K (1989) Application of double bounded probability density function for analysis of ocean waves. Ocean Eng 16:193–200