SciencePG
Science Publishing Group

Parameter Estimation of Kumaraswamy Distribution Based on Progressive Type II Censoring Scheme Using Expectation-Maximization Algorithm

Wafula Mike Erick[*], Kemei Anderson Kimutai, Edward Gachangi Njenga

Department of Statistics and Actuarial Science, Kenyatta University (KU), Nairobi, Kenya

Email address:
okhwafula@gmail.com (Wafula M. E.)
[*]Corresponding author

Abstract: This project considers the parameter estimation problem of test units from Kumaraswamy distribution based on progressive Type-II censoring scheme. The progressive Type-II censoring scheme allows removal of units at intermediate stages of the test other than the terminal point. The Maximum Likelihood Estimates (MLEs) of the parameters are derived using Expectation-Maximization (EM) algorithm. Also the expected Fisher information matrix based on the missing value principle is computed. By using the obtained expected Fisher information matrix of the MLEs, asymptotic 95% confidence intervals for the parameters are constructed. Through simulations, the behaviour of these estimates are studied and compared under different censoring schemes and parameter values. It's concluded that for an increasing sample; the estimated parameter values become closer to the true values, the variances and widths of the confidence intervals reduce. Also, more efficient estimates are obtained with censoring schemes concerned with removals of units from their right.

Keywords: Kumaraswamy Distribution, Progressive Type II Censoring, Maximum Likelihood Estimation, EM Algorithm

1. Introduction

Censored sampling arises in a life testing experiment whenever the experimenter does not observe (either intentionally or un-intentionally) the failure times of all units placed on a life test.

"According to Horst, a data sample is said to be censored when, either by accident or design the value of the variables under investigation is unobserved for some of the items in the sample."[1] Inference based on censored sampling has been studied during the past over 50 years by numerous authors for a wide range of lifetime distributions.

In this study, we assume that the lifetimes have Kumaraswamy distribution. This distribution was introduced by Kumaraswamy as a probability density function for double bounded random processes. [2] 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.

The two parameter Kumaraswamy distribution has a PDF and CDF given respectively by;

$$f(x;\theta,\lambda) = \lambda\theta x^{\lambda-1}(1-x^\lambda)^{\theta-1}, \tag{1}$$

$$F(x;\theta,\lambda) = 1-(1-x^\lambda)^\theta, \ 0 < x < 1 ; \ \lambda,\theta > 0 \tag{2}$$

Kumaraswamy and Ponnambalam et al. [2, 3] have pointed out that depending on the choice of the parameters, this distribution can be used to approximate many distributions, such as uniform, triangular, or almost any single model distribution and can also reproduce results of beta distribution. The basic properties of the distribution have been given by Jones. [4]

Inferential issues for the Kumaraswamy distribution based on censored data have been addressed by Gholizadeh et al. [5] who considered the Bayesian estimation of Kumaraswamy distribution under progressively Type II censored samples. Tabassum et al. [6] explored the Bayesian analysis of

Kumaraswamy distribution under failure censoring sampling scheme. Feroze et al [7] estimated the parameters of Kumaraswamy distribution under progressive type II censoring with random removals using maximum likelihood method.

Most recently, Mostafa et al [8] derived parameter estimators of Kumaraswamy distribution based on general progressive type II censoring scheme using maximum likelihood and Bayesian approaches. Also, some of the recent work on progressive censoring include but not limited to [9-14]. As far as we know, no one has described the EM algorithm for determining the MLEs of the parameters of the Kumaraswamy distribution based on progressive type-II censoring scheme.

The purpose of this study is to estimate the shape and scale parameters of the Kumaraswamy distribution under progressive type-II censoring using the EM algorithm and to compare the results under different censoring schemes.

In this work, we propose to use EM algorithm for computing MLEs. This is because the EM algorithm is relatively robust against the initial values compared to the traditional Newton-Raphson (NR) method. [15, 16] For some of the recently relevant references on EM algorithm and censoring include [17 and 20].

2. Parameter Estimation

2.1. Progressive Type II Censoring

Suppose n identical units are put on a test and the lifetime distributions of the n units are denoted by $X_1, ..., X_n$.

The integer m < n is fixed at the beginning of the experiment and they are the units which are observed completely until failure.

The censoring occurs progressively in m stages. These m stages offer failure times of the m completely observed units. At the time of the first failure (the first stage), R_1 of the $n - 1$ surviving units are randomly withdrawn from the experiment. At the time of the second failure (the 2nd stage), R_2 of the $n - 2 - R_1$ surviving units are withdrawn and so on. Finally, at the time of the mth failure (the mth stage), all the remaining $R_m = n - m - R_1 - R_2 - \cdots - R_{m-1}$ surviving units are withdrawn. According to Childs and Balakrishnan, we refer to this as progressive Type-II right censoring with scheme $(R_1, R_2, ... R_m)$. [21]

2.2. Maximum Likelihood Estimation

Let $X_{1:n}, X_{2:n}, ..., X_{m:n}$ denote a progressive Type II censored sample from Kumaraswamy distribution. Then according to [21] the likelihood function based on progressively Type II censored sample is given by;

$$L(\theta, \lambda; x_{1:n}, ..., x_{m:n}) = k \prod_{i=1}^{m} f(x_i; \theta, \lambda)[1 - F(x_i; \theta, \lambda)]^{R_i} \quad (3)$$

From equations (1) and (2), the likelihood function based on progressive Type II censored sample is as follows;

$$L(\theta, \lambda; x_{1:n}, ..., x_{m:n}) \propto \lambda^m \theta^m \prod_{i=1}^{m} x^{\lambda-1} \left(1 - x^\lambda\right)^{\theta-1} \left[\left(1 - x^\lambda\right)^\theta\right]^{R_i} \quad (4)$$

The log-likelihood function of equation (4) can be written as follows:

$$l(\theta, \lambda; x_{1:n}, ..., x_{m:n}) \propto m \ln \lambda + m \ln \theta + (\lambda - 1) \sum_{i=1}^{m} \ln(x_i)$$
$$+ (\theta - 1) \sum_{i=1}^{m} \ln(1 - x_i^\lambda) + \theta \sum_{i=1}^{m} R_i \ln\left(1 - x_i^\lambda\right) \quad (5)$$

2.3. EM Algorithm

We propose the EM algorithm, introduced by Dempster et al. [22] to find the MLEs.

Let

$$Z = (Z_1, Z_2, ..., Z_m), with..Z_j = (Z_{j1}, Z_{j2}, ..., Z_{jR_J}),$$

$j = 1, 2, ..., m,$ be the censored data.

We consider the censored data as missing data. The combination (X, Z) = W forms the complete data set. The log-likelihood function based on W can be written respectively as:

$$H(w; \theta, \lambda) \propto n \ln \lambda + n \ln \theta + (\lambda - 1) \sum_{j=1}^{m} \ln(x_j) +$$
$$(\theta - 1) \sum_{j=1}^{m} \ln(1 - x_j^\lambda) + (\lambda - 1) \sum_{j=1}^{m} \sum_{k=1}^{R_J} \ln Z_{jk}$$
$$+ (\theta - 1) \sum_{j=1}^{m} \sum_{k=1}^{R_j} \ln(1 - Z^\lambda_{jk}) \quad (6)$$

In the E-step, one requires to compute the pseudo-likelihood function. This can be obtained from $H(w; \theta, \lambda)$ by replacing any function of Z_{jk} $\left\{say, h(z_{jk}) \; by \; E\left(h(z_{jk})/z_{jk} > x_j\right).\right\}$

Therefore equation (6) becomes;

$$H^*(w; \theta, \lambda) \propto n \ln \lambda + n \ln \theta$$
$$+ (\lambda - 1) \sum_{j=1}^{m} \ln(x_j) + (\theta - 1) \sum_{j=1}^{m} \ln(1 - x_j^\lambda) + (\lambda - 1) \sum_{j=1}^{m} \sum_{k=1}^{R_J} E\left[\ln(Z_{jk})/Z_{jk} > x_j\right]$$
$$+ (\theta - 1) \sum_{j=1}^{m} \sum_{k=1}^{R_J} E\left[\ln(1 - Z^\lambda_{jk})/Z_{jk} > x_j\right] \quad (7)$$

Therefore, $given, X_j = x_j$, the conditional distribution of Z_{jk} follows a truncated Kumaraswamy distribution with left truncation at x_j. That is

$$f_{Z/X}(Z_j/X) = f_w(z_j)/\left[1 - F_w(x_j)\right], z_j > x_j \quad (8)$$

Therefore the conditional expectations in equations (6) and (7) can be obtained as follows:

$$E_1(v;\theta,\lambda) = E\left(\ln\left(1 - z^{\lambda}_{jk}\right)\Big/z_{jk} > v\right) = \ln\left(1 - v^{\lambda}\right) - \frac{1}{\theta} \quad (9)$$

$$\because E_2(v;\theta,\lambda) = E\left(\ln z_{jk}\Big/z_{jk} > v\right)$$

$$= \frac{\theta(1-v)}{\lambda\left(1-v^{\lambda}\right)^{\theta}}\beta(1,\theta)\left\{\Psi(1) - \Psi(1+\theta)\right\} \quad (10)$$

Thus, in the M-step of the $(k+1)^{th}$ iteration of the EM algorithm, the value of $\theta^{(k+1)}$ is first obtained by solving the following equation:

$$\frac{dH^*(w;\theta,\lambda)}{d\theta} = \frac{n}{\theta} + \sum_{j=1}^{m}\ln(1 - x_j^{\lambda^{(k)}}) + \sum_{j=1}^{m}R_j E_2\left(x_j;\theta^{(k)},\lambda^{(k)}\right) = 0$$

$$\theta^{(k+1)} = \hat{\theta} = \frac{-n}{\sum_{j=1}^{m}\ln(1 - x_j^{\lambda^{(k)}}) + \sum_{j=1}^{m}R_j E_2\left(x_j;\theta^{(k)},\lambda^{(k)}\right)} \quad (11)$$

Once $\theta^{(k+1)}$ is obtained, $\lambda^{(k+1)}$ is obtained by solving the equation

$$\frac{dH^*(w;\theta,\lambda)}{d\lambda}$$

$$= \frac{n}{\lambda} + \sum_{j=1}^{m}\ln(x_j) - (\theta-1)\sum_{j=1}^{m}\frac{x_j^{\lambda}\ln x_j}{1 - x_j^{\lambda}} + \sum_{j=1}^{m}R_j E_1\left(x_j;\theta^{(k+1)},\lambda^k\right) = 0$$

$$\lambda^{(k+1)} = \frac{-n}{\sum_{j=1}^{m}\ln(x_j) - (\hat{\theta}-1)\sum_{j=1}^{m}\frac{x_j^{\lambda}\ln x_j}{1 - x_j^{\lambda}} + \sum_{j=1}^{m}R_j E_1\left(x_j;\theta^{(k+1)},\lambda^k\right)} \quad (12)$$

2.4. Asymptotic Variance-Covariance Matrix of the MLEs

The variance–covariance matrix is used to provide a measure of precision for parameter estimators by utilizing the log-likelihood function. We first compute the variance–covariance matrix of parameters θ and λ by considering a complete data set from the Kumaraswamy distribution.

For such a case, the log likelihood function based on X is obtained as follows;

$$l_c(x;\theta,\lambda) = n\ln\lambda + n\ln\theta + (\lambda-1)\sum_{i=1}^{n}\ln(x_i) + (\theta-1)\sum_{i=1}^{n}\ln\left(1 - x_i^{\lambda}\right) \quad (13)$$

Using equation (13), the Fisher information matrix for the complete data set is given as;

$$I_c(\theta,\lambda) = -E\begin{pmatrix} \dfrac{d^2\ln l_c}{d\theta^2} & \dfrac{d^2\ln l_c}{d\theta d\lambda} \\ \dfrac{d^2\ln l_c}{d\theta d\lambda} & \dfrac{d^2\ln l_c}{d\lambda^2} \end{pmatrix} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

And the variance-covariance matrix of parameters θ and λ is given by

$$\begin{pmatrix} \text{var}(\hat{\theta}) & \text{cov}(\hat{\theta},\hat{\lambda}) \\ \text{cov}(\hat{\theta},\hat{\lambda}) & \text{var}(\hat{\lambda}) \end{pmatrix} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1} \quad (14)$$

Where

$$I_{11} = \frac{n}{\theta^2}$$

$$I_{12} = \frac{n\theta\,\beta(2,\theta-1)\left\{\Psi(2) - \Psi(1+\theta)\right\}}{\lambda}$$

$$I_{22} = \frac{n}{\lambda^2} + \frac{n\theta(\theta-1)}{\lambda^2}\beta(2,\theta-2)\left\{\psi'(2) - \psi'(\theta) + \left[\psi(2) - \psi(\theta)\right]^2\right\}$$

where $\psi(x)$ and $\psi'(x)$ are the digamma and trigamma functions respectively.

In this work, we are interested in deriving the asymptotic variance–covariance matrix for the MLEs based on the EM algorithm. For this we will use the procedure that was established by Louis and Tanner. [23, 24] The idea of this procedure is given by

$$I_{obs}(\eta) = I_c(\eta) - I_{miss}(\eta) \quad (15)$$

where $I_c(\eta), I_{obs}(\eta)$ and $I_{miss}(\eta)$ denote the complete, observed, and missing (expected) information, respectively, and η = (θ, λ). The Fisher information matrix for a single observation which is censored at the time of the jth failure is given by

$$I^{(j)}_{miss}(\eta) = -E\left(\frac{d^2\ln f_{z/x}\left(z_{jk}\Big/z_{jk} > x_j;\eta\right)}{d\eta^2}\right)$$

where $f_{z/x}\left(z_{jk}\Big/z_{jk} > x_j;\eta\right)$ is given in Equation (8). The expected values of the second partial of the log-likelihood function of Z given X are calculated as follows;

$$f\left(z_j/z_j = x_j;\theta,\lambda\right) = \frac{\lambda\theta z_j^{\lambda-1}\left(1 - z_j^{\lambda}\right)^{\theta-1}}{\left(1 - x_j^{\lambda}\right)^{\theta}}, z_j > x_j$$

$$\ln f_{z/x}\left(z_{jk}\Big/z_{jk} > x_j;\eta\right)$$
$$= \ln(\lambda\theta) + (\lambda-1)\ln z_j$$
$$- \theta\ln\left(1 - x_j^{\lambda}\right) + (\theta-1)\ln\left(1 - z_j^{\lambda}\right) \quad (16)$$

$$-E\left(\frac{d^2\ln f_{z/x}\left(z_{jk}\Big/z_{jk} > x_j;\eta\right)}{d\theta^2}\right) = \frac{1}{\theta^2} = I_{11}^{(j)} \quad (17)$$

$$-E\left(\frac{d^2 \ln f_{z/x}\left(z_{jk}/z_{jk} > x_j;\eta\right)}{d\theta d\lambda}\right) \quad (18)$$

$$= -\frac{x_j^\lambda \ln x_j}{1-x_j^\lambda} + \theta \frac{(1-x_j)}{\lambda\left(1-x_j^\lambda\right)^\theta} \beta(2,\theta-1)\left\{\Psi(2)-\Psi(\theta+1)\right\} = I_{12}^{(j)}$$

$$-E\left(\frac{d^2 \ln f_{z/x}\left(z_{jk}/z_{jk} > x_j;\eta\right)}{d\lambda^2}\right)$$

$$= \frac{1}{\lambda^2} - \frac{\theta x_j^\lambda \left(\ln x_j\right)^2}{\left(1-x_j^\lambda\right)^2} + \quad (19)$$

$$\frac{\theta(\theta-1)(1-x_j)\beta(2,\theta-2)}{\lambda^2\left(1-x_j^\lambda\right)^\theta}\left\{\psi'(2)-\psi'(\theta)+\left[\psi(2)-\psi(\theta)\right]^2\right\}$$

$$= I_{22}^{(j)}$$

Note that $I^{(j)}_{miss}$ is a function of x_j and η, since the expectation is taken with respect Z_j; therefore, the expected information matrix is simply

$$I_{miss}(\eta) = \sum_{j=1}^m R_j I^{(j)}_{miss}(\eta) \quad (20)$$

$$I_{miss}(\eta) = \begin{pmatrix} \sum_{j=1}^m R_j I_{11}^{(j)} & \sum_{j=1}^m R_j I_{12}^{(j)} \\ \sum_{j=1}^m R_j I_{12}^{(j)} & \sum_{j=1}^m R_j I_{22}^{(j)} \end{pmatrix}$$

Hence $I_{obs}(\eta) = I_c(\eta) - I_{miss}(\eta)$

$$I_{obs}(\eta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{pmatrix} - \begin{pmatrix} \sum_{j=1}^m R_j I_{11}^{(j)} & \sum_{j=1}^m R_j I_{12}^{(j)} \\ \sum_{j=1}^m R_j I_{12}^{(j)} & \sum_{j=1}^m R_j I_{22}^{(j)} \end{pmatrix}$$

Therefore, the variance–covariance matrix of parameter η can be obtained by

$$I^{-1}_{obs}(\eta) = \left[I_c(\eta) - I_{miss}(\eta)\right]^{-1} \quad (21)$$

Using equation (21) an approximate $100(1-\alpha)$ % confi-dence intervals for θ and λ is obtained respectively, as;

$$\left[\hat{\theta} \pm Z_{\alpha/2}\sqrt{\mathrm{var}\left(\hat{\theta}\right)}\right] \ and \ \left[\hat{\lambda} \pm Z_{\alpha/2}\sqrt{\mathrm{var}\left(\hat{\lambda}\right)}\right] \quad (22)$$

Where $Z_{\alpha/2} = \left(1-\alpha/2\right)^{th}$ is the percentile of the standard normal distribution.

3. Numerical Results and Discussion

In this section a simulation study is conducted to investigate how the above estimators perform in estimating the parame-ters of Kumaraswamy distribution based on progressive type II censored data. The samples were generated based on the algorithms of Balakrishnan and Sandhu and Aggarwala and Balakrishnan (1998). [25, 26] The censoring schemes con-sidered are given in table 1 below;

Table 1. Censoring schemes.

Scheme	Censoring rate
1 :	$r_1 = ... = r_{11} = 0, r_{12} = 6$
2 :	$r_1 = ... = r_5 = r_7 = ... = r_{12} = 0, r_6 = 6$
3 :	$r_2 = ... = r_{12} = 0, r_1 = 6$
4 :	$r_1 = ... = r_{14} = 0, r_{15} = 3$
5 :	$r_1 = ... = r_5 = r_9 = ... = r_{15} = 0, r_6 = r_7 = r_8 = 1$
6 :	$r_2 = ... = r_{15} = 0, r_1 = 3$
7 :	$r_1 = ... = r_{17} = 0, r_{18} = 7$
8 :	$r_1 = r_8 = r_{11} = ... = r_{18} = 0, r_9 = 3, r_{10} = 4$
9 :	$r_2 = ... = r_{18} = 0, r_1 = 7$
10 :	$r_1 = ... = r_{22} = 0, r_{18} = 3$
11 :	$r_1 = ... = r_8 = r_{11} = r_{22} = 0, r_9 = 2, r_{10} = 1$
12 :	$r_2 = ... = r_{22} = 0, r_1 = 3$
13 :	$r_1 = ... = r_{29} = 0, r_{30} = 10$
14 :	$r_1 = ... = r_{14} = r_{17} = ... = r_{30} = 0, r_{15} = r_{16} = 5$
15 :	$r_2 = ... = r_{30} = 0, r_1 = 10$
16 :	$r_1 = ... = r_{35} = 0, r_{36} = 4$
17 :	$r_1 = ... = r_{17} = r_{20}... = r_{36} = 0, r_{18} = r_{19} = 2$
18 :	$r_2 = ... = r_{36} = 0, r_1 = 4$

Clearly from table 1, schemes 1, 4, 7, 10, 13 and 16 are right censored schemes; 2, 5, 8, 11, 14 and 17 are centre censored while 3, 6, 9, 12, 15 and 18, are left censored schemes. The right, centre and left censored schemes are respectively denoted as n:m-R, n:m-C and n:m-L.

All the computational results were computed using R software

Table 2. MLEs, variances and confidence intervals of MLEs of Kumaraswamy distribution when λ=0.6 and θ=1.0.

Scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
1	18:12-R	0.7645	1.19158	0.09871	0.06021	0.0938	1.26118	0.46808	1.42109
2	18:12-C	0.76556	1.20544	0.10043	0.07087	0.08922	1.2679	0.47492	1.44196
3	18:12-L	0.7799	1.23841	0.10648	0.07514	0.0841	1.30169	0.4912	1.49161

Scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
4	18:15-R	0.73095	1.16709	0.09782	0.05675	0.06311	1.22478	0.49642	1.34377
5	18:15-C	0.73096	1.17661	0.09803	0.05771	0.06983	1.2353	0.5064	1.35768
6	18:15-L	0.73957	1.18491	0.0984	0.05855	0.06983	1.23548	0.50154	1.36982
7	25:18-R	0.70526	1.15949	0.08395	0.05137	0.08526	1.15125	0.51389	1.31109
8	25:18-C	0.71061	1.16257	0.09091	0.05163	0.0661	1.18113	0.51562	1.31551
9	25:18-L	0.722941	1.16529	0.09452	0.0519	0.06311	1.20278	0.51715	1.31943
10	25:22-R	0.67157	1.08667	0.08123	0.03801	0.06148	1.10766	0.5117	1.16765
11	25:22-C	0.6744	1.10452	0.09203	0.03923	0.05838	1.11043	0.51955	1.18949
12	25:22-L	0.60212	1.14617	0.08358	0.04214	0.06138	1.13377	0.54489	1.24747
13	40:30-R	0.63614	1.05911	0.06865	0.03183	0.07146	1.02082	0.51958	1.09867
14	40:30-C	0.63612	1.06611	0.06993	0.03223	0.06637	1.02603	0.52407	1.10815
15	40:30-L	0.6439	1.07324	0.06475	0.03258	0.05516	1.05262	0.52865	1.11783
16	40:36-R	0.60099	1.02478	0.0483	0.02668	0.12739	0.89458	0.52168	1.02787
17	40:36-C	0.62299	1.03134	0.0521	0.0271	0.13385	0.9381	0.5291	1.03958
18	40:36-L	0.62611	1.04616	0.05868	0.02762	0.10667	0.97156	0.53908	1.05925

From table 2, it is observed that irrespective of the censoring rate and the position at which the censored units are removed from the sample, for increasing sample size;

(i) the estimated value of the parameter becomes closer to the true value,

(ii) the variances of the MLEs decrease

Table 3. *Effect on the Confidence intervals of the estimates.*

Scheme	n:m	width of λ	width of θ
1	18:12-R	1.16738	0.95301
2	18:12-C	1.17868	0.96704
3	18:12-L	1.21759	1.00041
4	18:15-R	1.16167	0.84735
5	18:15-C	1.16305	0.85614
6	18:15-L	1.16547	0.86382
7	25:18-R	1.06599	0.7972
8	25:18-C	1.11503	0.79989
9	25:18-L	1.13967	0.80228
10	25:22-R	1.04618	0.65595
11	25:22-C	1.05205	0.66994
12	25:22-L	1.06329	0.70258
13	40:30-R	0.94936	0.57909
14	40:30-C	0.95966	0.58408
15	40:30-L	0.99746	0.58918
16	40:36-R	0.76719	0.50619
17	40:36-C	0.80428	0.51048
18	40:36-L	0.86489	0.52017

Table 3 clearly shows that the widths of 95% confidence intervals tend to be lesser as the sample size increases.

Table 4. *Effect of the number of censored units on estimates.*

scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
1	18:12-R	0.7645	1.19158	0.09871	0.06021	0.0938	1.26118	0.46808	1.42109
2	18:12-C	0.76556	1.20544	0.10043	0.07087	0.08922	1.2679	0.47492	1.44196
3	18:12-L	0.7799	1.23841	0.10648	0.07514	0.0841	1.30169	0.4912	1.49161
4	18:15-R	0.73095	1.16709	0.09782	0.05675	0.06311	1.22478	0.49642	1.34377
5	18:15-C	0.73096	1.17661	0.09803	0.05771	0.06983	1.2353	0.5064	1.35768
6	18:15-L	0.73957	1.18491	0.0984	0.05855	0.06983	1.23548	0.50154	1.36982

Table 4 has been extracted from table 2, so as to clearly illustrate the effect of censored units on the parameter estimates. The results in table 4 show that when the sample size is kept constant, then better dstimates are obtained when the censored units are reduced. Schemes 4-6 have better estimates compared to schemes 1-3 because the number of censored units in schemes 4-6 are each 3 units while in schemes 1-3, we have 6 units censored from each.

Table 5. *Effect of position of removal of units in the scheme on estimates.*

scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
1	18:12-R	0.7645	1.19158	0.09871	0.06021	0.0938	1.26118	0.46808	1.42109
2	18:12-C	0.76556	1.20544	0.10043	0.07087	0.08922	1.2679	0.47492	1.44196
3	18:12-L	0.7799	1.23841	0.10648	0.07514	0.0841	1.30169	0.4912	1.49161

The removal of units in scheme 1, 2 and 3 was done at the 12th, 6th, and 1st failures respectively and from the results it was observed that scheme 1 which is right censored, gave a better estimate followed by scheme 2 (centre censored scheme) and lastly scheme 3 (left censored scheme). The same trend was observed across all the censoring schemes i.e all the right censored schemes resulted in better estimates followed by centre censored and left censored in that order.

Table 6. *MLEs, variances and confidence intervals of MLEs of Kumaraswamy distribution when λ=2.2 and θ=3.5.*

scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
1	18:12-R	2.60598	3.83061	0.19357	0.76331	1.82916	3.46218	2.09551	5.46764
2	18:12-C	2.62011	3.84161	0.19533	0.76442	1.84026	3.48123	2.10042	5.4826
3	18:12-L	2.62426	3.84951	0.20058	0.76795	1.8317	3.49804	2.10432	5.49451
4	18:15-R	2.56329	3.82357	0.19383	0.63367	1.83556	3.37081	2.23806	5.30889
5	18:15-C	2.57632	3.826872	0.18148	0.61554	1.84254	3.41778	2.24078	5.31627
6	18:15-L	2.5967	3.83107	0.17294	0.61644	1.82001	3.4502	2.24209	5.31985
7	25:18-R	2.50917	3.85857	0.15119	0.53091	1.83915	3.2591	2.28034	5.1366
8	25:18-C	2.51364	3.78588	0.15352	0.53985	1.83733	3.26975	2.29568	5.17589
9	25:18-L	2.5252	3.82177	0.15546	0.5517	1.84372	3.28649	2.31584	5.2275
10	25:22-R	2.45423	3.69152	0.13098	0.40743	1.84118	3.14707	2.39035	4.8925
11	25:22-C	2.45571	3.72075	0.14127	0.41493	1.81306	3.17816	2.40812	4.93319
12	25:22-L	2.47865	3.75515	0.14529	0.42384	1.82476	3.21234	2.42903	4.98107
13	40:30-R	2.38023	3.62095	0.12241	0.32467	1.79288	3.04738	2.45403	4.68767
14	40:30-C	2.38451	3.63553	0.12243	0.32771	1.7971	3.05173	2.46339	4.70746
15	40:30-L	2.55376	3.64016	0.12327	0.32868	1.86378	3.12353	2.46637	4.71376
16	40:36-R	2.30453	3.61267	0.10232	0.26913	1.78206	2.90681	2.54577	4.57938
17	40:36-C	2.33327	3.61775	0.11919	0.27001	1.75586	2.99047	2.54918	4.58612
18	40:36-L	2.37704	3.61798	0.12067	0.27005	1.79505	3.03882	2.54934	4.58642

Table 6 also shows that for increasing sample size the estimated value of the parameter becomes closer to the true value and the variances of the MLEs decrease.

However, these variances are much higher than those obtained in table 2.

Table 7. *Effect on the Confidence intervals of the estimates.*

Scheme	n:m	width of λ	width of θ
1	18:12-R	1.163302	3.37213
2	18:12-C	1.64097	3.38218
3	18:12-L	1.66704	3.39019
4	18:15-R	1.53525	3.07083
5	18:15-C	1.57524	3.07549
6	18:15-L	1.6302	3.07776
7	25:18-R	1.41985	2.85626
8	25:18-C	1.43242	2.88021
9	25:18-L	1.44277	2.91166
10	25:22-R	1.30589	2.50215
11	25:22-C	1.3651	2.55204
12	25:22-L	1.38758	2.52507
13	40:30-R	1.2545	2.23364
14	40:30-C	1.25463	2.24407
15	40:30-L	1.25975	2.24739
16	40:36-R	1.12475	2.03361
17	40:36-C	1.23461	2.03694
18	40:36-L	1.24377	2.03708

The widths of the confidence intervals are also higher under these set of parameter values and tend to be lesser for an increasing sample size.

Table 8. *Effect of the number of censored units on estimates.*

scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
7	25:18-R	2.50917	3.85857	0.15119	0.53091	1.83915	3.2591	2.28034	5.1366
8	25:18-C	2.51364	3.78588	0.15352	0.53985	1.83733	3.26975	2.29568	5.17589
9	25:18-L	2.5252	3.82177	0.15546	0.5517	1.84372	3.28649	2.31584	5.2275
10	25:22-R	2.45423	3.69152	0.13098	0.40743	1.84118	3.14707	2.39035	4.8925
11	25:22-C	2.45571	3.72075	0.14127	0.41493	1.81306	3.17816	2.40812	4.93319
12	25:22-L	2.47865	3.75515	0.14529	0.42384	1.82476	3.21234	2.42903	4.98107

Table 6 as well reveals that reducing the censored units leads to better estimates for a constant sample size. In schemes 7-9, the number of units censored are each 7, while in schemes 10-12, the censored units are each 3 and we see from the results that schemes 10-12 gave better estimates compared to schemes 7-9.

Table 9. *Effect of position of removal of units in the scheme on estimates.*

scheme	n:m	$\hat{\lambda}$	$\hat{\theta}$	v(λ)	v(θ)	LL(λ)	UL(λ)	LL(θ)	UL(θ)
7	25:18-R	2.50917	3.85857	0.15119	0.53091	1.83915	3.2591	2.28034	5.1366
8	25:18-C	2.51364	3.78588	0.15352	0.53985	1.83733	3.26975	2.29568	5.17589
9	25:18-L	2.5252	3.82177	0.15546	0.5517	1.84372	3.28649	2.31584	5.2275

The removal of units in scheme 7, 8 and 9 was done at the 18th, 9th and 10th, and 1st failures respectively and from the results it was observed that scheme 7, gave a better estimate followed by scheme 8 and finally scheme 9. This trend was observed to cut across all the censoring schemes i.e all the right censored schemes resulted in better estimates followed by centre censored and left censored in that order.

4. Conclusion

This study has addressed the problem of estimation of parameters of the Kumaraswamy distribution based on progressive Type-II censored data. It is shown that the MLEs of the scale and shape parameters can be obtained by using EM algorithm.

A comparison of the MLEs and their variances as well as their confidence intervals is made by simulation for different censoring schemes. It is observed that:

i.   for an increasing sample size, the estimated value of the parameter becomes closer to the true value, the variances of the MLEs decrease and the widths of the confidence intervals become less.

ii. better estimates are obtained when the removal of units is from the right, followed by those at the centre and poorest for those removed from the left.

iii. reducing the number of units to be removed in the censoring scheme, leads to better estimates for a fixed sample size.

iv. an increase in the true parameter values leads to estimates with large variances and increased widths of the confidence intervals.

References

[1]   Horst R. (2009). *The Weibull Distribution Handbook*. CRC Press, Taylor and Francis Group LLC, New York.

[2]   Kumaraswamy P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46, 79-88.

[3]   Ponnambalam K., Seifi A. and Vlach J. (2001). Probabilistic design of systems with general distributions of parameters. *Integrated Journals on Circuit Theory Applications,* 29, 527-536.

[4]   Jones M. C (2009). Kumaraswamy's Distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6, 70-81.

[5]   Gholizadeh R., Khalilpor M. and Hsadian M. (2011). Bayesian Estimation in the Kumaraswamy Distribution under Progressively Type II Censoring Data. 3(9), 47-65.

[6]   Tabassum N. S, Navid Feroze and Muhammed Aslam (2013). Bayesian analysis of the Kumaraswamy distribution under failure censoring sampling scheme. *International Journal of Advanced Science and Technology*, 51, 39-58.

[7]   Feroze Navid and El-Batal Ibrahim (2013). Parameter Estimation based on Kumaraswamy Progressive Type II Censored data with Random Removals. *Journal of Modern Applied Statistical Methods*, 12(19).

[8]   Mostafa Mohie Eldin, Nora Khalil and Montaser Amein (2014). Estimation of Parameters of the Kumaraswamy Distribution Based on General Progressive Type II Censoring. *American Journal of Theoretical and Applied Statistics,* 3(6), 217-222.

[9]   Mahmoud, M. A. W., El-Sagheer, R. M., Soliman, A. A. and Abd Ellah, A. H., Inferences of the lifetime performance index with Lomax distribution based on progressive Type-II censored data. *Economic Quality Control*, 29(1), 39–51, (2014).

[10] Soliman, A. A., Abd Ellah, A. H., Abou-Elheggag, N. A. and El-Sagheer, R. M., Inferences for Burr-X model using Type-II progressively censored data with binomial removals. *Arabian journal of Mathematics*, 4, 127-139, (2015).

[11] EL-Sagheer, R. M. and Ahsanullah, M., Bayesian estimation based on progressively Type-II censored samples from compound Rayleigh distribution. *Journal of Statistical Theory and Applications*, 14(2), 107-122, (2015).

[12] EL-Sagheer, R. M., Estimation of the parameters of life for distributions having power hazard function based on progressively Type-II censored data. *Advances and Applications in Statistics*, 45(1), 1-27, (2015).

[13] EL-Sagheer, R. M., Estimation using progressively Type-II censored data from Rayleigh distribution with binomial removals: Bayesian and non-Bayesian approach. *JP Journal of Fundamental and Applied Statistics*, 8(1), 17-39, (2015).

[14] EL-Sagheer, R. M. Bayesian prediction based on general progressive censored data from generalized Pareto distribution. *Journal of Statistics Applications & Probability*, 5(1), 43-51, (2016).

[15] Watanabe M, Yamaguchi K. The EM algorithm and related statistical models. New York: Marcel Dekker; 2004.

[16] Ng HKT, Chan PS, Balakrishnan N. (2002) Estimation of parameters from progressively censored data using EM algorithm. *Computational Statistics and Data Analysis;* 39(4):371–386.

[17] Kus C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*; 51, 4497-4509.

[18] Tahmasbi R. and Rezaei S. (2008). A two-parameter lifetime distribution with decreasing failure rate. *Computational Statistics and Data Analysis*, 52, 3889-3901.

[19] Amal Helu, Hani Samawi and Mohammad Z. Raqab (2013). Estimation on lomax progressive censoring using the EM algorithm. *Journal of Statistical Computation and Simulation,* 837-861.

[20] Juan Li and Lina Ma (2015). Inference for the Generalized Rayleigh Distribution Based on Progressively Type II hybrid Censored data. *Journal of Information and Computational Science*, 1101-1112.

[21] Childs A. and Balakrishnan N. (2000). Conditional Inference Procedures for the Laplace Distribution when the observed samples are progressively censored. *Metrika*, 52(3), 253-265.

[22] Dempster A. P, Laird N. M, Rubin D. D (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

[23] Louis T. A (1982). Finding the observed matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B.*, 44, 226-233.

[24] Tanner M. A (1993). *Tools for Statistical Inference: observed data and data augmentation methods.2nd edition.* New York; Springer.

[25] Balakrishnan N. and Sandhu A. (1995). A simple simulation algorithm for generating progressive Type II censored sample. *American journal of Statistics,* 49, 229-230.

[26] Aggarwala R. and Balakrishnan N. (1998). Some Properties of Progressive Censored Order Statistics from arbitrary and uniform distributions with applications to Inference and Simulations. *Journal of Statistics and Planning Inference*, 70, 35-49.