Estimating Entropy of the Kumaraswamy Distribution with Contaminated Samples

Layla M. Nass[1], Waleed Abdullah Araheemah[2] and Husam Abdulrazzak Rasheed[3]
[1]College of Engineering, Mustansiriyah University, 10052 Baghdad, Iraq
[2]Technical College of Management - Baghdad Business, Middle Technical University, 10047 Baghdad, Iraq
[3]College of Administration and Economics, Mustansiriyah University, 10052 Baghdad, Iraq
Layla_matter@uomustansiriyah.edu.iq, dr.waleed@mtu.edu.iq, husamstat@uomustansiriyah.edu.iq

Keywords: Kumaraswamy Distribution, Maximum Likelihood Estimation Method, Moment Method, Mixed Method, Entropy Function.

Abstract: Kumaraswamy distribution (KD) is essential for many applications and experiments in renewable energy and geophysics. The distribution has many applied functions that depend on the accuracy of the distribution parameter estimates. Several simulation experiments were carried out based on the change in each of (polluted ratios (ϖ_i), sample sizes (n_i), and proposal parameter values (α_i , β_i) and estimation methods with the (Maximum Likelihood Estimation Method (MLE), Moment Method (MOM) and Mixed Method (MM))and studying the effect of the change in each of them on the distribution parameter estimates and the entropy function of the distribution function. The results of the simulation experiments were compared through the mean square error, and the results showed the effect of the function estimator on the change in each of the (ϖ_i, n_i, α_i , β_i) and the estimation method. Simulation experiments can be carried out on other distributions (Weibull, beta, Normal), and methods of estimating ((Bayesian, Shrinkage)) can be carried out on the (KD).

1 INTRODUCTION

Entropy function for the (KD) has vast importance in renewable energy research applications and the presence of polluted observations affects the capabilities of the distribution parameters and entropy function many researches have been done. This paper presents methods for estimating the entropy function of the Kumaraswamy distribution. The simulation results of comparisons between different experiments were introduced, which were characterized by several factors between them (sample size, pollution percentage, distribution parameters) they received significant attention from many researchers, given the importance of this topic because of its applications.

The research aims to know the effect of the different estimation methods on pollution percentages, and the study seeks to find the best estimates for the distribution parameters, which definitely affect the estimation processes of the entropy function and thus determine the best estimation method.

The effect of statistical distribution pollutants is one of the essential factors. When observed, it is necessary to know the extent of their impact on the distribution parameters. The ability of estimation methods to overcome these effects may vary, so it is essential to know which ways are better and which are stable towards the presence of different proportions of pollution.

2 RELATED WORKS

In this field, many studies have been presented. In [1], methods for estimating (MLE, MM) and different simulation experiments were used to search for the effect of changes in simulation experiments on the (KD) location and shape parameters The results showed that estimation methods can provide capabilities close to the real values of the distribution parameters [2]. In [3], the practical results showed the ability of each of (Bayesian and Moment) estimation methods to provide estimates of the distribution parameters close to the real values and thus possess the reliability function of an estimated distribution close to the actual reality. [4] represents a new family of continuous distributions called (Marshall-Olkin Kumaraswamy), the Lorenz curve was adopted to ensure that the distribution represents the research data, and the results showed the ability of the estimation method to provide optimal estimates. In [5], the effect of changing the sample size on the classification error was studied by assuming unequal sizes by applying the two methods (Bayesian and moment) on the distribution parameters of (KD). The results were compared by adopting the loss function criterion and showed that the estimation method was affected by changing the sample size.

Another study represented in [6], which includes using the Deep Learning-Based Method to Detect Components in Scanned Structural Drawings images, showed the presented algorithms' ability to deal with drawn images. This research included the Wavelet transformation in recognizing several images of drawn faces that were dealt with according to different noise models.

3 KUMARASWAMY DISTRIBUTION (K.D.)

This distribution is considered one of the most important statistical distributions and has a family of continuous distributions with [0,1] random variable intervals. The distribution was first assumed by the Indian researcher (Poondi Kumaraswamy). The distribution has many applications, especially in determining the reliability and distribution of failure times for various engineering devices [7].

3.1 Distribution Features

The probability density function for the distribution is given by:

$$f(x,\alpha,\beta) = \alpha\,\beta\,x^{\alpha-1}(1-x^{\alpha})^{\beta-1}, \qquad (1)$$

$with\ \ 0 \le x \le 1\ ,\ \ \alpha,\beta\ > 0$

And the cumulative density for the distribution can be:

$$F(x,\alpha,\beta) = 1-(1-x^{\alpha})^{\beta}. \qquad (2)$$

Figures 1 and 2 represent the probability density function and cumulative distribution function for Kumaraswamy distribution with various (α, β) values.

The entropy function for the distribution can be [4]:

$$H = \left(1-\frac{1}{\beta}\right)+\left(1-\frac{1}{\alpha}\right)Hb-(\alpha\,\beta), \qquad (3)$$

with $Hb = \sum_{i=1}^{b} \frac{1}{i}$.

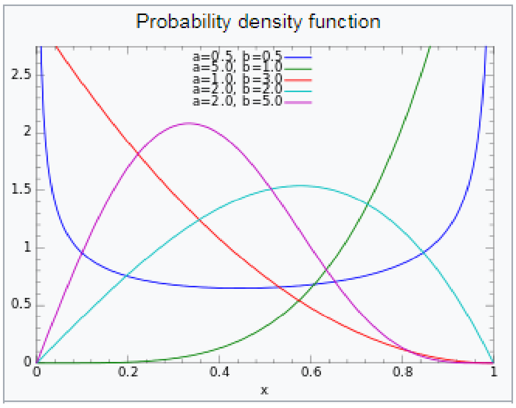

Figure 1: Probability density function for KD.

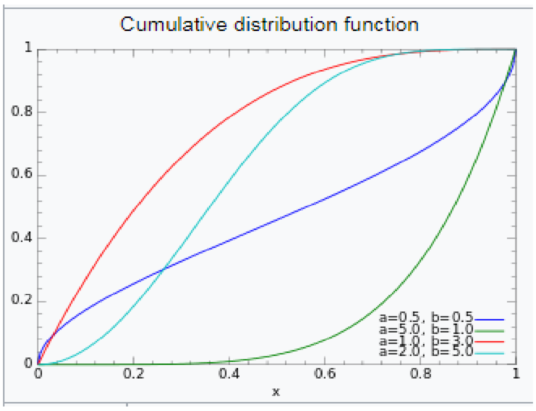

Figure 2: Cumulative density function for KD.

3.2 The Relation with other Distributions

The Kumaraswamy distribution has a relationship with many other distributions through transformations that can be made on the random variable of this distribution:

1) When both shape parameters are set to 1, the Kumaraswamy distribution reduces to the uniform distribution between 0 and 1.

$$x = 1-\left(1-x^{\frac{1}{\beta}}\right)^{\frac{1}{\alpha}}. \qquad (4)$$

2) By taking a logarithm transformation of a Kumaraswamy random variable, we obtain a random variable that follows an exponential distribution:

$$x = -Ln(y). \qquad (5)$$

3.3 Contaminated Distribution

In practical applications, assuming a random variable with a pure statistical distribution is difficult, especially with an increased sample size. This is because the random sample may not all follow the assumed distribution function. Consequently, the pollution rates included in the distribution cannot be completely isolated. Thus, it becomes a high or low impact depending on many factors, including (Pollution percentage, sample size, and extremes in some pollutants) The contaminated distribution can be:

$$f_t(\alpha_t, \beta_t, z) = P f_1(\alpha_1, \beta_1, x) + (1-P) f_2(\alpha_2, \beta_2, y), \quad (6)$$

With $f_t(\alpha_t, \beta_t, z)$ Represents the probability density function of contaminated distribution with (α_t, β_t) parameters. $f_1(\alpha_1, \beta_1, x)$ Represents the probability density function of the polluted sample with (α_1, β_1) parameters. $f_2(\alpha_2, \beta_2, y)$ Represents the probability density function of the unpolluted sample with (α_2, β_2) parameters. P represents polluted ratio

3.4 Derivation of Entropy for KD

The entropy function of the Kumaraswamy distribution can be obtained through the following series of steps (4-5)

$$h_x = -\int_0^1 \alpha\,\beta\, x^{\alpha-1}(1-x^{\alpha})^{\beta-1}\, Ln\left(\alpha\,\beta\, x^{\alpha-1}(1-x^{\alpha})^{\beta-1}\right) dx. \quad (7)$$

By using the following logarithm properties $Ln(\varphi\,\tau) = Ln(\varphi) + Ln(\tau)$, $Ln(\varphi)^{\tau} = \tau\, Ln(\varphi)$, we get:

$$h_x = -\int_0^1 \alpha\,\beta\, x^{\alpha-1}(1-x^{\alpha})^{\beta-1}\, Ln(\alpha\,\beta) + (\alpha-1)Ln(x) + (\beta-1)Ln(1-x^{\alpha}))dx, \quad (8)$$

$$h_x = -Ln(\alpha\,\beta) - \alpha\,\beta(\alpha-1)\int_0^1 x^{\alpha-1}(1-x^{\alpha})^{\beta-1}Ln(x)dx - \alpha\,\beta(\beta-1)\int_0^1 x^{\alpha-1}(1-x^{\alpha})^{\beta-1}Ln(1-x^{\alpha})dx. \quad (9)$$

To get integration results by the following transformation $y = 1 - x^{\alpha} \Leftrightarrow dy = -\alpha x^{\alpha-1}dx$, and we get:

$$h_y = -Ln(\alpha\,\beta) - \alpha\,\beta(\alpha-1)\frac{1}{\alpha^2}\, B(1-\beta)\left(\varphi(1) - \varphi(1+\beta)\right) - \alpha\,\beta(\beta - 1)\int_0^1 \frac{1}{\alpha}\, y^{\beta-1}Ln(y)dy.$$

Since we have $B(1-\beta) = \frac{\Gamma(1)\Gamma(\beta)}{\Gamma(\beta+1)}$, and $\Gamma(\lambda) = (\lambda-1)!$, then $B(1-\beta) = \frac{1}{\beta}$, By using Partisan integration, let $u = Ln(y) \Leftrightarrow \frac{1}{y}dy, dv = y^{\beta-1}\, dy \Leftrightarrow v = \frac{1}{\beta}y^{\beta}$. The integration results:

$$h_y = Ln\left(\frac{1}{\alpha\,\beta}\, e^{\frac{\beta-1}{\beta} + \frac{\alpha-1}{\alpha}\left(\gamma + \varphi(\beta) + \frac{1}{\beta}\right)}\right). \quad (11)$$

With γ represent (Euler's constant), and $\varphi(\beta+1) = \varphi(\beta) + \frac{1}{\beta}$.

3.5 Simulation Experiments for KD

Simulation experiments can be performed with sufficient iterations to obtain the entropy estimator through the formula (11) and assuming the random generation function (R) as a binomial generation function and represent the basis of simulation for various distributions that represent random values within the period [0-1]. The simulation process for a random sample with Kumaraswamy distribution can be used by the formula (2) and letting $(F(x, \alpha, \beta) = R)$ then:

$$R = 1 - (1-x^{\alpha})^{\beta}, \quad (12)$$
$$(1-x^{\alpha})^{\beta} = 1 - R,$$
$$(1-x^{\alpha}) = (1-R)^{\frac{1}{\alpha}},$$
$$x = (1-(1-R)^{\frac{1}{\beta}})^{\frac{1}{\alpha}}.$$

Which (x) is a random variable that follows the Kumaraswamy distribution with (α, β) parameters.

4 ESTIMATION METHODS

Parameters estimators of the distribution parameters can be obtained using several estimation methods, including [7].

4.1 Maximum Likelihood Estimation Method (MLE)

Assume that a sample of size (n) has a distribution of Kumaraswamy distribution such that the function of the Maximum Likelihood will be [8]:

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i\,, \alpha, \beta). \quad (13)$$

By substituting the formula (1), the function is as follows

$$L(x_1, x_2, \dots, x_n) = (\alpha\beta)^n \prod_{i=1}^{n} x_i^{\alpha-1} \prod_{i=1}^{n} (1 - x_i^{\alpha})^{\beta-1}. \quad (14)$$

By taking the logarithm, it could be getting

$$Log\big(L(x_1, x_2, \dots, x_n)\big) = nLog(\alpha) + nLog(\beta) + (\alpha - 1)\sum_{i=1}^{n} Log(x_i) + (\beta - 1)\sum_{i=1}^{n} Log(1 - x_i^{\alpha}). \quad (15)$$

And taking the partial derivative to $(Log\big(L(x_1, x_2, \dots, x_n)\big))$ for (α) the first time and (β) a second time, and Each derivative is equal to zero, we get:

$$\hat{\beta}_{mle} = \frac{-n}{\sum_{i=1}^{n} Log(1 - x_i^{\hat{\alpha}_{mle}})}, \quad (16)$$

$$\hat{\alpha}_{mle} = -\frac{n}{\sum_{i=1}^{n} Log(x_i) - (\hat{\beta}_{mle} - 1)\sum_{i=1}^{n} \frac{x_i^{\hat{\alpha}_{mle}} Log(\hat{\alpha}_{mle})}{1 - x_i^{\hat{\alpha}_{mle}}}}, \quad (17)$$

with $(\hat{\alpha}_{mle}$ *and* $\hat{\beta}_{mle})$ represent Maximum Likelihood Estimators for $(\alpha$ *and* $\beta)$ respectively. One of the numerical methods is adopted to obtain the estimators from the previous formulas (Newton Raphson).

4.2 Moment Method (MOM)

This method depends on equal the sample moments with the distribution moments to obtain the estimators of the distribution parameters according to the following steps. The (k_{th}) moment can be [9]:

$$E(x^k) = \alpha\beta \frac{\Gamma\left(\frac{k}{\alpha} + 1\right)\Gamma(\beta)}{\Gamma\left(\frac{k}{\alpha} + 1 + \beta\right)}. \quad (18)$$

Obtain the (1_{st}) moment by letting $(k = 1)$ in formula (18) we get:

$$E(x^1) = \alpha\beta \frac{\Gamma\left(\frac{1}{\alpha} + 1\right)\Gamma(\beta)}{\Gamma\left(\frac{1}{\alpha} + 1 + \beta\right)}. \quad (19)$$

The (2_{nd}) moment b letting $(k = 2)$ in formula (18) we get

$$E(x^2) = \alpha\beta \frac{\Gamma\left(\frac{2}{\alpha} + 1\right)\Gamma(\beta)}{\Gamma\left(\frac{2}{\alpha} + 1 + \beta\right)}, \quad (20)$$

while the sample moments $\tau_1 = \frac{\sum_{i=1}^{n} x_i}{n}$, $\tau_2 = \frac{\sum_{i=1}^{n} {x_i}^2}{n}$, by using $(E(x^1) = \tau_1)$ and $(E(x^2) = \tau_2)$ we get:

$$\tau_1 = \alpha\beta \frac{\Gamma\left(\frac{1}{\alpha} + 1\right)\Gamma(\beta)}{\Gamma\left(\frac{1}{\alpha} + 1 + \beta\right)}, \quad (21)$$

$$\alpha\,\tau_1 = \beta \frac{\Gamma\left(\frac{1}{\alpha} + 1\right)\Gamma(\beta)}{\Gamma\left(\frac{1}{\alpha} + 1 + \beta\right)}. \quad (22)$$

Since we have $\Gamma(\beta + 1) = \beta\Gamma(\beta)$, Then:

$$\hat{\alpha}_{mom} = \tau_1 \frac{\Gamma\left(\frac{1}{\alpha} + 1 + \beta\right)}{\Gamma\left(\frac{1}{\alpha} + 1\right)\Gamma(\beta + 1)}, \quad (23)$$

$$\hat{\beta}_{mom} = \tau_2 \frac{\Gamma\left(\frac{2}{\alpha} + 1 + \beta\right)}{\Gamma\left(\frac{2}{\alpha} + 1\right)\Gamma(\beta)}, \quad (24)$$

with $(\hat{\alpha}_{mom}$ *and* $\hat{\beta}_{mom})$ represent Moment Estimators for $(\alpha$ *and* $\beta)$ respectively. To obtain the estimators from the previous formulas, one of the numerical methods is adopted (Newton Raphson)

4.3 Mixed Method (MM)

This method depends on (p, q) estimators through the following formulas [10]-[13]:

$$\hat{\alpha}_{mixed} = p.\,\hat{\alpha}_{mle} + q.\,\hat{\alpha}_{mom}, \quad (25)$$

$$\hat{\beta}_{mixed} = p.\,\hat{\beta}_{mle} + q.\,\hat{\beta}_{mom}, \quad (26)$$

with $0 \le p, q \le 1$, $p + q = 1$, and $(\hat{\alpha}_{mixed}$ *and* $\hat{\beta}_{mixed})$ represent Mixed Estimators for $(\alpha$ *and* $\beta)$ respectively, it is known that when $(p = 1, q = 0)$ then $\hat{\alpha}_{mixed} = \hat{\alpha}_{mle}$, $\hat{\beta}_{mixed} = \hat{\beta}_{mle}$. And when $(p = 0, q = 1)$ then $\hat{\alpha}_{mixed} = \hat{\alpha}_{mom}$, $\hat{\beta}_{mixed} = \hat{\beta}_{mom}$$(p$ *and* $q)$The following algorithm can option estimators (a-start, b-propose (p) to be as small as possible and near to (0), c-calculate initial mixed estimators $(\hat{\alpha}_{mixed}, \hat{\beta}_{mixed})$, d-calculate (p_n) such that $(p_n = p + \delta)$ and (δ) small as possible, e-re calculate new mixed estimators, f-finding (ε) such that (ε) represents the absolute difference between estimators in step (c) and (e), respectively, g-if $(\varepsilon < 0.05)$, go to step (j), h-make new mixed estimators equal to initial mixed estimators, i-go to step (d), j-make final mixed estimators equal to the new mixed estimators, k-end).

5 EXPERIMENTAL RESULTS

Many simulation experiments for Kumaraswamy distribution were done with various experimental parameters such that sample size ($n_1 = 60, n_2 = 100, n_3 = 200$), polluted ratio ($\varpi_1 = 5, \varpi_2 = 10, \varpi_3 = 15$) first parameter values ($\alpha_1 = 1, \alpha_2 = 1.5, \alpha_3 = 2$) and second parameter values ($\beta_1 = 0.25, \beta_2 = 0.5, \beta_3 = 0.75$). The total number of simulation experiments is equal to (108) with (1000) iterations for each experiment. Mean square error (Mse) can be:

$$Mse = \frac{\sum_{i=1}^{T}(\hat{E}_i - E)^2}{T}, \qquad (27)$$

with T represents the number of iterations, $\hat{E}_i$ represent entropy estimator for each (i) represent estimation method, E represents true entropy. Numerical simulation experiments result for (27) unpolluted samples in Table 1.

By observing the simulation results of the unpolluted data according to Table 2 and the total number of best estimation methods for each experiment Figure 3, it becomes clear that the estimation method is affected by the change in the sample size and actual values of the distribution parameters, which may make the mean squares error the smallest sometimes or the largest at other times. The total number of best estimation methods shows that the best estimation method for unpolluted samples was (MlE). Numerical simulation experiments resulted in (81) polluted samples in Tables 2-4 with (5%, 10%, and 15%) polluted ratios.

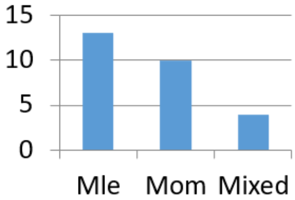

Figure 3: Total number of best estimation methods for unpolluted data.

Table 1: Entropy estimators for unpolluted samples.

Simulation Parameter			$True_{entropy}$	$Mle_{entropy}$	$Mom_{entropy}$	$Mixed_{entropy}$
α_1	β_1	n_1	-1.61371	-1.4991	-1.80051	-1.62311
		n_2	-1.61371	-1.54643	-1.76379	-1.64148
		n_3	-1.61371	-1.57986	-1.67154	-1.61987
	β_2	n_1	-0.30685	-0.27878	-0.34248	-0.29616
		n_2	-0.30685	-0.28731	-0.32811	-0.29918
		n_3	-0.30685	-0.29654	-0.29848	-0.29428
	β_3	n_1	-4.57E-02	-2.80E-02	-7.65E-02	-3.82E-02
		n_2	-4.57E-02	-3.38E-02	-2.54E-02	-0.01918
		n_3	-4.57E-02	-3.89E-02	-3.19E-02	-3.08E-02
α_2	β_1	n_1	-0.45921	-0.35863	-0.60703	-0.46215
		n_2	-0.29004	-0.23117	-0.41054	-0.30942
		n_3	-5.98E-02	-2.85E-02	-0.09746	-5.91E-02
	β_2	n_1	0.847639	0.868272	0.791913	0.838153
		n_2	1.016808	1.030126	0.980281	1.009547
		n_3	1.247026	1.254692	1.214741	1.237172
	β_3	n_1	1.10884	1.116132	1.093353	1.111268
		n_2	1.278009	1.283563	1.28402	1.28791
		n_3	1.508227	1.511365	1.507502	1.511734
α_3	β_1	n_1	3.31E-02	0.127667	-0.13251	1.76E-02
		n_2	0.286836	0.349941	0.133287	0.254183
		n_3	0.632163	0.664067	0.587775	0.630069
	β_2	n_1	1.339935	1.356927	1.282683	1.327616
		n_2	1.593689	1.605592	1.548785	1.582499
		n_3	1.939015	1.945189	1.919115	1.934115
	β_3	n_1	1.601137	1.607716	1.570156	1.595583
		n_2	1.85489	1.858722	1.854498	1.860042
		n_3	2.200217	2.20224	2.191239	2.198156

Table 2: Mean square error for unpolluted samples.

Simulation Parameter			Mse_{Mle}	Mse_{Mom}	Mse_{Mixed}	$Best$
α_1	β_1	n_1	0.920968	0.741731	0.914256	0.741731
		n_2	0.893928	0.795912	0.888796	0.795912
		n_3	0.873797	0.822762	0.872583	0.822762
	β_2	n_1	3.37E-02	2.66E-02	3.36E-02	0.026567
		n_2	0.033018	0.030727	0.032804	0.030727
		n_3	0.031994	3.22E-02	3.19E-02	0.031942
	β_3	n_1	1.09E-03	8.53E-04	1.06E-03	0.000853
		n_2	9.12E-04	1.04E-03	9.06E-04	0.000906
		n_3	7.96E-04	7.77E-04	7.96E-04	0.000777
α_2	β_1	n_1	8.76E-02	5.36E-02	8.54E-02	0.053609
		n_2	3.45E-02	2.18E-02	3.37E-02	0.021773
		n_3	2.14E-03	1.19E-03	2.00E-03	0.001189
	β_2	n_1	0.230563	0.244471	0.230868	0.230563
		n_2	0.335017	0.347399	0.335246	0.335017
		n_3	0.507428	0.529239	0.507681	0.507428
	β_3	n_1	0.400966	0.419673	0.400952	0.400952
		n_2	0.534321	0.536047	0.534414	0.534321
		n_3	0.745134	0.756962	0.745149	0.745134
α_3	β_1	n_1	2.25E-03	0.012091	1.82E-03	0.001822
		n_2	2.26E-02	0.046823	2.29E-02	0.022564
		n_3	0.125153	0.140557	0.125638	0.125153
	β_2	n_1	0.581917	0.620785	0.582396	0.581917
		n_2	0.827493	0.855322	0.827878	0.827493
		n_3	1.230498	1.243514	1.230689	1.230498
	β_3	n_1	0.83851	0.854783	0.838597	0.83851
		n_2	1.12749	1.130324	1.127665	1.12749
		n_3	1.587792	1.595005	1.587815	1.587792

Table 3: Entropy estimators for (5%) polluted samples.

Simulation Parameter			$True_{entropy}$	$Mle_{entropy}$	$Mom_{entropy}$	$Mixed_{entropy}$
α_1	β_1	n_1	-1.61371	-1.51677	-0.49661	-1.51742
		n_2	-1.61371	-1.55894	-0.44386	-1.55894
		n_3	-1.61371	-1.58539	-0.38032	-1.58539
	β_2	n_1	-0.30685	-0.2852	6.58E-02	-0.28673
		n_2	-0.30685	-0.29216	0.119968	-0.2925
		n_3	-0.30685	-0.30062	0.141053	-0.30062
	β_3	n_1	-4.57E-02	-3.25E-02	0.303576	-0.03292
		n_2	-4.57E-02	-3.66E-02	0.315768	-3.69E-02
		n_3	-4.57E-02	-0.04162	0.366511	-4.17E-02
α_2	β_1	n_1	-0.45921	-0.36434	0.156473	-0.36519
		n_2	-0.29004	-0.23074	0.368025	-0.23077
		n_3	-5.98E-02	-2.76E-02	0.615932	-2.76E-02
	β_2	n_1	0.847639	0.86667	0.951885	0.865382
		n_2	1.016808	1.028374	1.131685	1.027654
		n_3	1.247026	1.253488	1.374999	1.25345
	β_3	n_1	1.10884	1.115613	1.175867	1.115221
		n_2	1.278009	1.282501	1.388351	1.282456
		n_3	1.508227	1.511324	1.624471	1.511309
α_3	β_1	n_1	3.31E-02	0.133091	5.30E-02	0.117785
		n_2	0.286836	0.347973	0.462327	0.341624
		n_3	0.632163	0.660967	0.794484	0.659436
	β_2	n_1	1.339935	1.358462	1.28979	1.35725
		n_2	1.593689	1.604614	1.556856	1.604238
		n_3	1.939015	1.945038	1.93175	1.944723
	β_3	n_1	1.601137	1.607396	1.570131	1.60712
		n_2	1.85489	1.858945	1.855732	1.858806
		n_3	2.200217	2.202913	2.200576	2.202914

Table 4: Mean square error for (5%) polluted samples.

Simulation Parameter			Mse_{Mle}	Mse_{Mom}	Mse_{Mixed}	$Best$
α_1	β_1	n_1	0.910359	1.876019	0.910042	0.910042
		n_2	0.885931	1.945804	0.885931	0.885931
		n_3	0.870779	2.029506	0.870779	0.870779
	β_2	n_1	3.34E-02	0.118286	3.32E-02	0.033221
		n_2	0.032476	0.140177	3.24E-02	0.032398
		n_3	3.16E-02	0.138519	3.16E-02	0.031597
	β_3	n_1	9.90E-04	4.23E-02	9.75E-04	0.000975
		n_2	8.21E-04	4.49E-02	8.14E-04	0.000814
		n_3	7.84E-04	6.48E-02	7.83E-04	0.000783
α_2	β_1	n_1	0.087477	0.289353	8.72E-02	0.087241
		n_2	3.46E-02	0.234739	3.46E-02	0.034574
		n_3	2.18E-03	0.17217	2.18E-03	0.002176
	β_2	n_1	0.231168	0.204755	0.231586	0.204755
		n_2	0.335426	0.307817	0.335682	0.307817
		n_3	0.507939	0.459671	0.507966	0.459671
	β_3	n_1	0.400965	0.378318	0.401125	0.378318
		n_2	0.534229	0.494337	0.53424	0.494337
		n_3	0.745239	0.69989	0.745245	0.69989
α_3	β_1	n_1	2.57E-03	3.48E-04	1.83E-03	0.000348
		n_2	0.02234	2.12E-02	2.26E-02	0.021189
		n_3	0.125578	0.105337	0.125839	0.105337
	β_2	n_1	0.581534	0.602495	0.581966	0.581534
		n_2	0.828016	0.853932	0.828161	0.828016
		n_3	1.230804	1.241422	1.230957	1.230804
	β_3	n_1	0.838505	0.863061	0.83865	0.838505
		n_2	1.127503	1.132021	1.127549	1.127503
		n_3	1.587461	1.58678	1.587463	1.58678

Table 5: Entropy estimators for (10%) polluted samples.

Simulation Parameter			$True_{entropy}$	$Mle_{entropy}$	$Mom_{entropy}$	$Mixed_{entropy}$
α_1	β_1	n_1	-1.61371	-1.52776	-0.17959	-1.52776
		n_2	-1.61371	-1.56413	-0.11677	-1.56413
		n_3	-1.61371	-1.59034	-3.95E-02	-1.59034
	β_2	n_1	-0.30685	-0.28771	0.235725	-0.28856
		n_2	-0.30685	-0.29522	0.296676	-0.29531
		n_3	-0.30685	-0.30144	0.375599	-0.30144
	β_3	n_1	-4.57E-02	-3.61E-02	0.366934	-3.62E-02
		n_2	-4.57E-02	-0.04135	0.50985	-4.15E-02
		n_3	-4.57E-02	-4.14E-02	0.556392	-4.15E-02
α_2	β_1	n_1	-0.45921	-0.37531	0.399822	-0.37542
		n_2	-0.29004	-0.23483	0.572279	-0.23483
		n_3	-5.98E-02	-3.28E-02	0.863922	-3.28E-02
	β_2	n_1	0.847639	0.865293	0.967567	0.864099
		n_2	1.016808	1.0288	1.165374	1.027987
		n_3	1.247026	1.252918	1.450341	1.252773
	β_3	n_1	1.10884	1.116381	1.199827	1.116143
		n_2	1.278009	1.282445	1.429242	1.282444
		n_3	1.508227	1.510293	1.66147	1.510223
α_3	β_1	n_1	3.31E-02	0.130564	0.211526	0.115545
		n_2	0.286836	0.345741	0.504413	0.339342
		n_3	0.632163	0.666728	0.89133	0.664084
	β_2	n_1	1.339935	1.359449	1.246503	1.35745
		n_2	1.593689	1.604329	1.552303	1.604013
		n_3	1.939015	1.945065	1.911959	1.944864
	β_3	n_1	1.601137	1.606513	1.572402	1.606345
		n_2	1.85489	1.858795	1.846517	1.858778
		n_3	2.200217	2.202338	2.195873	2.202258

Table 6: Mean square error for (10%) polluted samples.

Simulation Parameter			Mse_{Mle}	Mse_{Mom}	Mse_{Mixed}	$Best$
α_1	β_1	n_1	0.905221	2.317955	0.905221	0.905221
		n_2	0.882861	2.400418	0.882861	0.882861
		n_3	0.867977	2.529457	0.867977	0.867977
	β_2	n_1	3.27E-02	0.181931	3.26E-02	0.032557
		n_2	3.20E-02	0.216071	3.20E-02	0.032005
		n_3	3.16E-02	0.249715	3.16E-02	0.031583
	β_3	n_1	8.82E-04	6.41E-02	8.78E-04	0.000878
		n_2	7.48E-04	0.116952	7.45E-04	0.000745
		n_3	7.87E-04	0.122997	7.86E-04	0.000786
α_2	β_1	n_1	8.41E-02	0.454718	8.41E-02	0.084052
		n_2	3.45E-02	0.368339	3.45E-02	0.034473
		n_3	1.86E-03	0.304289	1.86E-03	0.001857
	β_2	n_1	0.231306	0.205026	0.231555	0.205026
		n_2	0.335683	0.299644	0.336036	0.299644
		n_3	0.508286	0.44135	0.508423	0.44135
	β_3	n_1	0.400978	0.370606	0.401054	0.370606
		n_2	0.534222	0.482366	0.534223	0.482366
		n_3	0.745724	0.678435	0.74578	0.678435
α_3	β_1	n_1	2.31E-03	8.91E-03	1.66E-03	0.001661
		n_2	2.25E-02	2.25E-02	2.31E-02	0.022456
		n_3	0.124496	9.66E-02	0.124926	0.096567
	β_2	n_1	0.580856	0.630687	0.582009	0.580856
		n_2	0.828363	0.853707	0.82852	0.828363
		n_3	1.230586	1.248955	1.230777	1.230586
	β_3	n_1	0.838926	0.844608	0.839022	0.838926
		n_2	1.127433	1.143146	1.127433	1.127433
		n_3	1.588087	1.592386	1.588154	1.588087

Table 7: Entropy estimators for (15%) polluted sample.

Simulation Parameter			$True_{entropy}$	$Mle_{entropy}$	$Mom_{entropy}$	$Mixed_{entropy}$
α_1	β_1	n_1	-1.61371	-1.58678	0.956204	-1.58678
		n_2	-1.61371	-1.60359	1.127636	-1.60359
		n_3	-1.61371	-1.61185	1.350412	-1.61185
	β_2	n_1	-0.30685	-0.32004	1.051619	-0.32004
		n_2	-0.30685	-0.31337	1.223889	-0.31337
		n_3	-0.30685	-0.31062	1.485089	-0.31062
	β_3	n_1	-4.57E-02	-5.39E-02	1.168763	-0.05427
		n_2	-4.57E-02	-5.26E-02	1.334221	-5.27E-02
		n_3	-4.57E-02	-4.92E-02	1.577179	-4.92E-02
α_2	β_1	n_1	-0.45921	-0.37435	1.038057	-0.37435
		n_2	-0.29004	-0.24327	1.268302	-0.24327
		n_3	-5.98E-02	-3.72E-02	1.587408	-3.72E-02
	β_2	n_1	0.847639	0.861197	1.207057	0.858905
		n_2	1.016808	1.023201	1.44239	1.02255
		n_3	1.247026	1.250272	1.776981	1.249852
	β_3	n_1	1.10884	1.113756	1.382758	1.113756
		n_2	1.278009	1.280313	1.622865	1.280313
		n_3	1.508227	1.509365	1.931998	1.509365
α_3	β_1	n_1	3.31E-02	0.12758	0.500406	0.119359
		n_2	0.286836	0.340469	0.744165	0.34026
		n_3	0.632163	0.661074	1.140866	0.661074
	β_2	n_1	1.339935	1.356115	1.057572	1.356015
		n_2	1.593689	1.603637	1.383846	1.603637
		n_3	1.939015	1.944362	1.77236	1.944362
	β_3	n_1	1.601137	1.606745	1.442743	1.606745
		n_2	1.85489	1.858391	1.714245	1.858391
		n_3	2.200217	2.202193	2.131986	2.202193

Table 8: Mean square error for (15%) polluted sample.

Simulation Parameter			Mse_{Mle}	Mse_{Mom}	Mse_{Mixed}	$Best$
α_1	β_1	n_1	0.87009	4.464725	0.87009	0.87009
		n_2	0.86023	4.87518	0.86023	0.86023
		n_3	0.855916	5.417952	0.855916	0.855916
	β_2	n_1	3.00E-02	0.791039	0.030038	0.030038
		n_2	3.03E-02	0.981239	3.03E-02	0.030312
		n_3	3.06E-02	1.294964	3.06E-02	0.03065
	β_3	n_1	5.79E-04	0.522284	5.77E-04	0.000577
		n_2	5.78E-04	0.661871	5.77E-04	0.000577
		n_3	6.75E-04	0.91522	6.75E-04	0.000675
α_2	β_1	n_1	8.39E-02	1.060256	8.39E-02	0.083896
		n_2	3.30E-02	1.003138	3.30E-02	0.032971
		n_3	1.76E-03	0.94891	1.76E-03	0.001762
	β_2	n_1	0.232077	0.176739	0.232689	0.176739
		n_2	0.337223	0.259288	0.337351	0.259288
		n_3	0.509123	0.385804	0.509294	0.385804
	β_3	n_1	0.401695	0.326406	0.401695	0.326406
		n_2	0.535022	0.433511	0.535022	0.433511
		n_3	0.746388	0.595247	0.746388	0.595247
α_3	β_1	n_1	2.36E-03	7.28E-02	1.94E-03	0.001936
		n_2	2.28E-02	5.33E-02	2.28E-02	0.022837
		n_3	0.125697	0.113573	0.125697	0.113573
	β_2	n_1	0.582051	0.724558	0.582174	0.582051
		n_2	0.828687	0.960917	0.828687	0.828687
		n_3	1.231053	1.354227	1.231053	1.231053
	β_3	n_1	0.838435	0.932324	0.838435	0.838435
		n_2	1.127518	1.229791	1.127518	1.127518
		n_3	1.588113	1.635419	1.588113	1.588113

By observing the simulation results of the polluted data according to Tables 3-8 and the total number of best estimation methods for each experiment Figure 4, it becomes clear that the estimation method is affected by the change of the polluted ratios, sample size and actual values of the distribution parameters, which may make the mean squares error the smallest sometimes or the largest at other times.

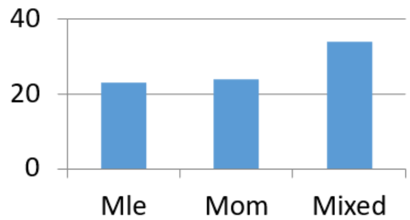

Figure 4: Total number of best estimation methods for polluted data.

The total number of best estimation methods shows that the best method for contaminated samples was (Mixed). Comparing Figure 1 to Figure 2 shows the effect of pollution ratios, and the (Mixed)method instead of the (Mle) method has the minimum mean square error for most of the simulation experiments.

6 CONCLUSIONS

The results of the simulation experiments conducted on both polluted and unpolluted samples indicate that the performance of the estimation methods, namely Maximum Likelihood Estimation (MLE), Moment Method (MOM), and Mixed Method (MM), is significantly influenced by changes in sample size, distribution parameter values, and pollution ratios. The findings demonstrate that the entropy estimator of the Kumaraswamy distribution is sensitive to variations in these factors, particularly in the presence of contaminated observations. In the case of unpolluted samples, the Maximum Likelihood Estimation method provided more accurate parameter estimates with lower mean square error values. However, when pollution ratios were introduced into the data, the Mixed Method showed superior performance compared to the other estimation techniques, achieving the minimum mean square error in most simulation scenarios.

These results highlight the effectiveness of the Mixed Method in improving entropy estimation under polluted data conditions and its ability to provide more stable estimates in the presence of outliers or contaminated observations. The study also suggests that further investigation using alternative estimation methods such as Bayesian or shrinkage techniques may provide additional improvements in entropy estimation for the Kumaraswamy distribution under varying sample conditions.

REFERENCES

[1] M. Mohie Eldin, “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, p. 217, doi: 10.11648/j.ajtas.20140306.17.

[2] S.-J. Chen and C.-L. Hwang, “Fuzzy Multiple Attribute Decision Making Methods,” Fuzzy Multiple Attribute Decision Making, 1992, pp. 289-486, doi: 10.1007/978-3-642-46768-4_5.

[3] H. G. Simbolon, I. Fithriani, and S. Nurrohmah, “Estimation of shape β parameter in Kumaraswamy distribution using Maximum Likelihood and Bayes method,” AIP Conference Proceedings, vol. 1862, 2017, p. 030160, doi: 10.1063/1.4991264.

[4] George and S. Thobias, “Marshall-Olkin Kumaraswamy Distribution,” International Mathematical Forum, vol. 12, no. 2, 2017, pp. 47-69, [Online]. Available: https://doi.org/10.12988/imf.2017.611151.

[5] A. S. Hassan and S. A. Atia, “Statistical inference and data analysis for inverted Kumaraswamy distribution based on maximum ranked set sampling with unequal samples,” Scientific Reports, vol. 14, no. 1, Oct. 2024, doi: 10.1038/s41598-024-74468-4.

[6] E.-S. El-Sherpieny and M. Ahmed, “On the Kumaraswamy Kumaraswamy distribution,” International Journal of Basic and Applied Sciences, vol. 3, no. 4, Sep. 2014, doi: 10.14419/ijbas.v3i4.3182.

[7] B. E. Mohammed, “Statistical Properties of Kumaraswamy-Generalized Exponentiated Exponential Distribution,” International Journal of Computer Applications, vol. 94, no. 4, May 2014, pp. 1-8, doi: 10.5120/16328-5602.

[8] M. S. Khan, R. King, and I. L. Hudson, “Transmuted Kumaraswamy Distribution,” Statistics in Transition New Series, vol. 17, no. 2, Jun. 2016, pp. 183-210, doi: 10.21307/stattrans-2016-013.

[9] L. M. Nassir, H. A. Rasheed, and A. A. R. Rasheed, “The effect of polluted samples on Bayesian Estimators of Burr type –XII distribution,” Periodicals of Engineering and Natural Sciences (PEN), vol. 10, no. 2, Apr. 2022, p. 337, doi: 10.21533/pen.v10i2.2862.

[10] A. J. Lemonte, W. Barreto-Souza, and G. M. Cordeiro, “The exponentiated Kumaraswamy distribution and its log-transform,” Brazilian Journal of Probability and Statistics, vol. 27, no. 1, Feb. 2013, doi: 10.1214/11-bjps149.

[11] R. I. Kadhim, et al., “AI-enhanced Cybersecurity in Fluid Dynamics Simulation Software,” 2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS), pp. 1-8, Dec. 2024, doi: 10.1109/isas64331.2024.10845531.

[12] A. A. Hadi, et al., “Machine Learning Techniques for Turbulence Analysis and Simulation,” 2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS), pp. 1-7, Dec. 2024, doi: 10.1109/isas64331.2024.10845730.

[13] M. Abdualwhab, J. Gareed, A. J. Al-Shimary, H. B. Furaijl, S. R. Ahmed, and S. Algburi, “Leveraging IoT Data for Real-Time Business Decision-Making,” 2024 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1-8, Nov. 2024, doi: 10.1109/ismsit63511.2024.10757297.