Inference on the Kumaraswamy distribution

Bing Xing Wang*, Xiu Kun Wang

Zhejiang Gongshang University, China

Keming Yu

Brunel University, UK

Abstract

Many lifetime distribution models have successfully served as population models for risk analysis and reliability mechanisms. The Kumaraswamy distribution is one of these distributions which is particularly useful to many natural phenomena whose outcomes have lower and upper bounds or bounded outcomes in the biomedical and epidemiological research. This paper studies point estimation and interval estimation for the Kumaraswamy distribution. The inverse estimators for the parameters of the Kumaraswamy distribution are derived. Numerical comparisons with MLE and biased-corrected methods clearly indicate the proposed inverse estimators are promising. Confidence intervals for the parameters and reliability characteristics of interest are constructed using pivotal or generalized pivotal quantities. Then the results are extended to the stress-strength model involving two Kumaraswamy populations with different parameter values. Construction of confidence intervals for the stress-strength reliability is derived. Extensive simulations are used to demonstrate the performance of confidence intervals constructed using generalized pivotal quantities.

Key words: Kumaraswamy distribution, exact inference, confidence interval, stress-strength model, pivotal quantity, order statistics.

* *Address for correspondence:* Department of Statistics, Zhejiang Gongshang University, China. E-mail: wangbingxing@163.com

1 Introduction

The cumulative distribution function of the Kumaraswamy distribution Kum(α, β) is given by

$$F(x) = 1 - (1 - x^{\alpha})^{\beta}, \quad 0 < x < 1,$$

where $\alpha > 0, \beta > 0$ are unknown shape parameters.

Due to its beta-type and better than beta distribution by its explicit expression of quantile function, the Kumaraswamy distribution has received considerable attention in the literature and has also been used for other purposes (Sundar and Subbiah, 1989; Koutsoyiannis and Xanthopoulos, 1989; Fletcher and Ponnambalam, 1996; Seifi et al., 2000; Ponnambalam et al., 2001; Ganji et al., 2006; Courard-Hauri, 2007 and Sanchez et al., 2007). Jones (2009) provided the basic properties of the Kumaraswamy distribution and discussed some similarities and differences between the beta and Kumaraswamy distributions. Mitnik (2013) studied some new properties of the Kumaraswamy distribution. Lemonte (2011) derived modified maximum likelihood estimators that are bias-free to second order. Furthermore, Garg (2009) discussed generalized order statistics for the Kumaraswamy distribution. Nadar et al. (2013) considered statistical analysis procedures for the Kumaraswamy distribution based on record values.

Classical maximum likelihood estimate (MLE) is typically used in existing inference for the Kumaraswamy distribution. MLE is a popular method but not always the best one. In practice, most of the sample sizes of real data are not big enough, so that MLE-based large sample asymptotic intervals may be invalid for small sample (DiCiccio and Efron, 1996). For example, visual analogue scales (VAS), frequently used for the assessment of intensity of pain, are bounded within the interval 0-100 mm. But relatively small sample size is a moderately limiting factor on VAS based study (McCoy et al., 2005).

In this paper we propose exact inference of the Kumaraswamy distribution for the parameters from a univariate sample and for stress-strength reliability model from two independent samples. The inference includes exact confidence interval or generalized confidence interval. First, given a random sample, $X_1, X_2, ..., X_n$, from the Kumaraswamy distribution, we aim at an exact inference of parameters α and β. Second, we consider an exact inference of stress-strength model when both the stress and strength variables follow the Kumaraswamy distributions. The classical stress-strength reliability model involves two independent random variables X and Y, where X represents the strength variable of a unit and Y represents the stress variable to which the unit is subjected. The stress-strength reliability of the unit is defined as $R = P(Y < X)$. This model was introduced by Birnbaum (1956) and developed by Birnbaum and McCarty (1958). Since then, this model has been investigated under different distributions (Aminzadeh, 1997; Surles and Padgett, 2001; Kundu and Gupta, 2005, 2006; Baklizi, 2008; Krishnamoorthy and Lin, 2010; Lio and Tsai, 2012; Nadar et al., 2014). The stress-strength model has also been widely used in many other fields such as receiver operating characteristic curve analysis (Reiser, 2000) and clinical trial applications (Hauck et al., 2000). Kotz et al. (2003) provided some excellent information on the topic.

Because the Kumaraswamy distribution is the proportional hazard distribution family, we applied the procedures proposed by Wang et al. (2010) to this distribution. The point and interval estimation are detailed studied. We further study the biased-corrected estimation and BC_α bootstrap confidence intervals. Moreover, we extend the procedure to the stress-strength model. The paper is organized as follows. Section 2 gives the parameter estimation of α and β, including point estimators and interval estimation of these parameters and other distribution quantities such as quantiles. In addition, Section 2 carries out numerical comparisons of the proposed method with MLE-based method and bootstrap-based method. Section 3 discusses interval estimation for stress-strength model, including some numerical analyses. Section 4 illustrates the application by the analysis of a proportion of total capacity data. Some closing comments are briefed in Section 5.

2 Estimation of the Kumaraswamy distribution

In order to derive the parameter estimation of the Kumaraswamy distribution, the following results are needed. Lemma 1 can be found in Wang et al. (2010).

Lemma 1 *Supposed that $Z_{(1)}, Z_{(2)}, ..., Z_{(n)}$ are the order statistics from the exponential distribution with mean λ^{-1} and sample size n. Let $S_i = Z_{(1)} + ... + Z_{(i)} + (n-i)Z_{(i)}$, $Q_i = (S_i/S_{i+1})^i$, $i = 1, 2, ..., n-1$, $Q_n = S_n$ defined to be 1. Then*

(1) $Q_1, Q_2, ..., Q_n$ are independent;

(2) $Q_1, Q_2, ..., Q_{n-1}$ have the uniform distribution $U(0, 1)$;

(3) Q_n follows the gamma distribution with the probability density function

$$f(z) = \frac{\lambda^n}{\Gamma(n)} z^{n-1} e^{-\lambda z},\ z > 0.$$

Lemma 2 *Let*

$$g(x) = \frac{\log(1 - b^x)}{\log(1 - a^x)},\ x > 0,$$

where $0 < a < b < 1$ are constant. Then $g(x)$ is an increasing function.

Proof. We shall first prove that

$$h(x) = \frac{x \log x}{(1-x)\log(1-x)}$$

is a decreasing function in (0, 1).

By the mean value theorem, we have, for any $x \in (0,1)$, $\log(1-x) - \log(1-0) = -x/(1-\xi_1)$, and $\log(1) - \log x = (1-x)/\xi_2$, where $\xi_1 \in (0,x)$, and $\xi_2 \in (x,1)$. Therefore, we have

$$\begin{aligned} h'(x) &= \frac{(\log x + 1 - x)\log(1-x) + x\log x}{[(1-x)\log(1-x)]^2} \\ &= \frac{1}{[(1-x)\log(1-x)]^2} \frac{x(1-x)(\xi_1 - \xi_2)}{\xi_2(1-\xi_1)} < 0. \end{aligned}$$

Thus, $h(x)$ is an decreasing function.

Notice that $g'(x) = g(x)[h(a^x) - h(b^x)]/x$, and $h(x)$ is an decreasing function, hence $g'(x) > 0$. Therefore, $g(x)$ is an increasing function.

2.1 Point estimation

Let $X_1, X_2, ..., X_n$ be a random sample from the Kumaraswamy distribution Kum(α, β) and $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are the corresponding order statistics. Then $\{F(X_{(i)})\}_{i=1}^n$ are the order statistics from the uniform distribution $U(0,1)$ with size n. Accordingly,

$$Z_i = -\log(1 - F(X_{(i)})) = -\beta \log(1 - X_{(i)}^\alpha),\ i = 1, 2, ..., n$$

are the order statistics from the standard exponential distribution $Exp(1)$ with size n.

Let

$$S_i = \sum_{j=1}^{i} \log(1 - X_{(j)}^\alpha) + (n-i)\log(1 - X_{(i)}^\alpha),\ i = 1, 2, ..., n.$$

Then we have from Lemma 1 that $S_1/S_2, (S_2/S_3)^2, ..., (S_{n-1}/S_n)^{n-1}$ are independent and have the uniform distribution $U(0,1)$. Therefore, we have

$$W(\alpha) = \sum_{i=1}^{n-1} \left[-2\log\left(\frac{S_i}{S_{i+1}}\right)^i \right] = 2\sum_{i=1}^{n-1} \log\frac{S_n}{S_i} \sim \chi^2(2n-2).$$

Because $W(\alpha)/(2n-4)$ converges to 1 with probability one, we can obtain a corresponding point estimator $\hat{\alpha}$ of α from $W(\alpha) = 2(n-2)$ or the following equation:

$$\sum_{i=1}^{n-1} \log\frac{S_n}{S_i} = n - 2. \tag{1}$$

Notice that

$$\frac{S_n}{S_i} = 1 + \frac{\sum_{j=i+1}^{n} V_{i,j} - (n-i)}{\sum_{j=1}^{i} V_{i,j} + (n-i)},$$

and that $V_{i,j}$ is an increasing function of α when $i < j$ and decreasing function of α when $i > j$, we have from Lemma 2 that $W(\alpha)$ is an increasing function of α, where

$$V_{i,j} = \log(1 - X_{(j)}^{\alpha}) / \log(1 - X_{(i)}^{\alpha}).$$

Moreover, it is clear that $W(\alpha)$ can take any positive value. Thus the equation (1) has a unique solution.

Similarly, since $-2\beta S_n \sim \chi^2(2n)$, we obtain estimator $\hat{\beta}$ of β from the following equation:

$$\hat{\beta} = -\frac{n-1}{\sum_{i=1}^{n} \log(1 - X_{(i)}^{\hat{\alpha}})}. \quad (2)$$

The estimators given by (1) and (2) are a type of inverse estimators (IE) of parameters (Wang et al., 2010). We shall study the finite sample properties of the proposed estimators in Section 2.3. They turn out to be very good and outperform MLEs.

Similar to Lemonte (2011), the proposed estimators ($\hat{\alpha}$, $\hat{\beta}$) can be bias-corrected based on bootstrap method. Let $(\hat{\alpha}_i^*, \hat{\beta}_i^*)$ be the inverse estimators of (α, β) based on the ith parametric bootstrap sample ($i = 1, 2, ..., B$). Then the bias-corrected estimators of (α, β) are given by

$$\bar{\alpha} = 2\hat{\alpha} - \hat{\alpha}_{(\cdot)}^*, \ \bar{\beta} = 2\hat{\beta} - \hat{\beta}_{(\cdot)}^*,$$

where $\hat{\alpha}_{(\cdot)}^* = \frac{1}{B}\sum_{i=1}^{B} \hat{\alpha}_i^*, \hat{\beta}_{(\cdot)}^* = \frac{1}{B}\sum_{i=1}^{B} \hat{\beta}_i^*$.

2.2 Interval estimation

First, we discuss interval estimation of the parameter α. Notice that the pivotal quantity, $W(\alpha)$, is a function of α only and does not depend on β, an exact confidence interval for α is thus given by the following theorem.

Theorem 1 *Suppose that $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are the order statistics from the Kumaraswamy distribution* $\mathrm{Kum}(\alpha, \beta)$ *with sample size n. Then, for any $0 < \gamma < 1$,*

$$\left[W^{-1}\{\chi_{1-\gamma/2}^2(2n-2)\}, \ W^{-1}\{\chi_{\gamma/2}^2(2n-2)\}\right]$$

is an $1-\gamma$ confidence interval for the parameter α, where $\chi_{\gamma}^2(v)$ is the upper γ percentile of the χ^2 distribution with v degrees of freedom and, for $t > 0$, $W^{-1}(t)$ is the solution in α of the equation $W(\alpha) = t$.

We now derive generalized confidence intervals for the parameter β and some important quantities of the Kumaraswamy distribution, such as its mean, quantiles and reliability function.

Let $g(W, \mathbf{X})$ is the unique solution of $W(\alpha) = W$, where $\mathbf{X} = (X_{(1)}, X_{(2)}, ..., X_{(n)})$. Because $V = -2\beta S_n \sim \chi^2(2n)$, we have that $\beta = -V/(2S_n)$. Based on the substitution method given by Weerahandi (2004), we substitute $g(W, \mathbf{X})$ for α in the expression for β and obtain the following generalized pivotal quantity for the parameter β:

$$
\begin{aligned}
Y_1 &= -\frac{V}{2\sum_{i=1}^{n} \log(1 - x_{(i)}^{g(W,\mathbf{X})})} && (3)\\
&= \frac{\beta \sum_{i=1}^{n} \log(1 - X_{(i)}^{g(W,\mathbf{X})})}{\sum_{i=1}^{n} \log(1 - x_{(i)}^{g(W,\mathbf{X})})}, && (4)
\end{aligned}
$$

where $\mathbf{x} = (x_{(1)}, x_{(2)}, ..., x_{(n)})$ is the observed value of $\mathbf{X}$. It is clear from (3) that the distribution of Y_1 is free of any unknown parameters. It is also clear from (4) that Y_1 reduces to β when $\mathbf{X} = \mathbf{x}$. Thus Y_1 is a generalized pivotal quantity. Let $Y_{1,\gamma}$ is the upper γ percentile of Y_1, then $[Y_{1,1-\gamma/2}, Y_{1,\gamma/2}]$ is an $1-\gamma$ generalized confidence interval for β. The values $Y_{1,1-\gamma/2}$ and $Y_{1,\gamma/2}$ can be obtained by using the following steps.

Step 1: For a given data set $(n, \mathbf{x})$, generate $W \sim \chi^2(2n-2)$ and $V \sim \chi^2(2n)$, independently. Using these values, compute $g(W, \mathbf{x})$ from the equation $W(\alpha) = W$.

Step 2: Compute value of Y_1 using (3).

Step 3: Repeat the steps 1-2 $m(\geq 10000)$ times. Then $Y_{1,\gamma}$ can be estimated by the $100(1-\gamma)$th percentile of these m generated Y_1.

Notice that the mean, pth quantile $(0 < p < 1)$ and reliability function of the Kumaraswamy distribution are given by $\mu = \beta B(1 + 1/\alpha, \beta)$, $x_p = [1 - (1-p)^{1/\beta}]^{1/\alpha}$ and $R(x_0) = (1 - x_0^{\alpha})^{\beta}$ respectively, where $B(\cdot, \cdot)$ is the beta function. Similar to the derivation of Y_1 for the parameter β, we obtain the following generalized pivotal quantities Y_2, Y_3 and Y_4 for μ, x_p and $R(x_0)$ respectively:

$$
\begin{aligned}
Y_2 &= Y_1 \cdot B\left(1 + \frac{1}{g(W,\mathbf{x})}, Y_1\right), && (5)\\
Y_3 &= [1 - (1-p)^{1/Y_1}]^{1/g(W,\mathbf{X})}, && (6)\\
Y_4 &= (1 - x_0^{g(W,\mathbf{X})})^{Y_1}. && (7)
\end{aligned}
$$

Let $Y_{2,\gamma}, Y_{3,\gamma}, Y_{4,\gamma}$ denote the upper γ percentiles of Y_2, Y_3, Y_4, respectively. Then $Y_{2,\gamma}, Y_{3,\gamma}, Y_{4,\gamma}$ are the $1-\gamma$ lower confidence limits for μ, x_p and $R(x_0)$, respectively. Just as in the case of Y_1, the percentiles of Y_2, Y_3, Y_4 can be obtained by Monte Carlo simulations.

Remark: Notice that $X_{(i)}^{g(W,\mathbf{X})} = (X_{(i)}^{\alpha})^{g(W,\mathbf{X})/\alpha}$ and that $X_{(1)}^{\alpha}, X_{(2)}^{\alpha}, ..., X_{(n)}^{\alpha}$ are the order statistics from the Kum$(1, \beta)$ with sample size n, thus $g(W, \mathbf{x})/\alpha$ does not depend on α. It is observed from (3) that the generalized confidence interval of β does not depend on the parameter α. Similarly, notice that Y_3^{α} does not depend on α, thus the coverage probability of the generalized confidence interval for x_p does not depend on α, but its interval length depends on α.

The confidence intervals of the parameters β, μ, x_p and $R(x_0)$ can be derived by the bootstrap procedure — BC_a method below.

Let θ be the parameter of interest and $\hat{\theta}_1^*, ..., \hat{\theta}_B^*$ are the MLEs of θ based on parametric bootstrap samples. The bias correction value z_0 is given by

$$z_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta}_i^* < \hat{\theta})}{B}\right),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. The value a, which measures skewness of the data, is given by

$$a = \frac{\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6\left(\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2\right)^{3/2}},$$

where $\hat{\theta}_{(i)}$ is the MLE of the sample without the ith observation, $\hat{\theta}_{(\cdot)}$ is the mean of the $\hat{\theta}_{(i)}$ values. Hence an $100(1-\gamma)\%$ BC_a bootstrap confidence interval is given by $[\hat{\theta}_{(B\gamma_1)}^*, \hat{\theta}_{(B\gamma_2)}^*]$, where

$$\gamma_1 = \Phi\left(z_0 + \frac{z_0 + \Phi^{-1}(\gamma/2)}{1 - a[z_0 + \Phi^{-1}(\gamma/2)]}\right)$$

and

$$\gamma_2 = \Phi\left(z_0 + \frac{z_0 + \Phi^{-1}(1-\gamma/2)}{1 - a[z_0 + \Phi^{-1}(1-\gamma/2)]}\right).$$

Because the coverage probabilities of these generalized confidence intervals and bootstrap confidence intervals may depend on nuisance parameters, a simulation is conducted to study the performance of coverage probabilities of these confidence intervals. These simulation results are reported in Section 2.3.

2.3 Simulation study

In order to assess the finite sample properties of the proposed procedures, a simulation study is conducted to compare the performance of the proposed point estimators with MLEs and study the coverage probabilities of the proposed generalized confidence intervals.

Notice that $X_{(j)}^{\hat{\alpha}} = (X_{(j)}^{\alpha})^{\hat{\alpha}/\alpha}$ and that $X_{(1)}^{\alpha}, X_{(2)}^{\alpha}, ..., X_{(n)}^{\alpha}$ are the order statistics from the Kumaraswamy distribution Kum$(1, \beta)$ with sample size n, thus we have from (1) and (2) that the relative biases and relative MSEs of the estimators $\hat{\alpha}$ and $\hat{\beta}$ do not depend on the parameter α. Therefore, without loss of generality, we take $\alpha = 1$ in our simulation study and consider different values of β. For different choices of sample sizes, we generated random samples from the Kumaraswamy distribution Kum$(1, \beta)$.

We report the average relative biases and average relative mean square errors (MSEs) in point estimation of α and β over 10,000 replications for the same cases. The results are presented in Tables 1–3.

Table 1: The relative biases and relative MSEs of the estimators $\hat{\alpha}$ and $\hat{\beta}$ in Kum$(1, 0.5)$

	Relative bias				Relative MSE			
	α		β		α		β	
n	IE	MLE	IE	MLE	IE	MLE	IE	MLE
10	0.0650	0.4495	0.0185	0.3135	0.6182	1.2131	0.3062	0.7427
15	0.0418	0.2705	0.0082	0.1794	0.2966	0.4712	0.1413	0.2429
20	0.0234	0.1851	0.0017	0.1224	0.1887	0.2650	0.0919	0.1366
30	0.0191	0.1224	0.0027	0.0793	0.1124	0.1410	0.0531	0.0691
50	0.0095	0.0691	0.0010	0.0452	0.0613	0.0689	0.0306	0.0357
80	0.0074	0.0440	0.0007	0.0278	0.0363	0.0385	0.0183	0.0201
100	0.0065	0.0355	0.0005	0.0221	0.0289	0.0299	0.0147	0.0158

Note: The relative bias and relative MSE of the estimation $\hat{\theta}$ for θ are defined as $bias(\hat{\theta})/\theta$ and $MSE(\hat{\theta})/\theta^2$ respectively.

Table 2: The relative biases and relative MSEs of the estimators $\hat{\alpha}$ and $\hat{\beta}$ in Kum$(1, 1)$

	Relative bias				Relative MSE			
	α		β		α		β	
n	IE	MLE	IE	MLE	IE	MLE	IE	MLE
10	0.0114	0.2941	0.0478	0.4318	0.2680	0.4843	0.5952	1.9073
15	0.0096	0.1834	0.0224	0.2335	0.1546	0.2327	0.2101	0.4078
20	0.0028	0.1277	0.0102	0.1564	0.1042	0.1412	0.1249	0.2045
30	0.0052	0.0858	0.0081	0.0995	0.0648	0.0801	0.0684	0.0950
50	0.0019	0.0488	0.0038	0.0561	0.0367	0.0415	0.0386	0.0470
80	0.0025	0.0314	0.0026	0.0344	0.0221	0.0239	0.0227	0.0257
100	0.0025	0.0255	0.0021	0.0274	0.0177	0.0188	0.0181	0.0200

Table 3: The relative biases and relative MSEs of the estimators $\hat{\alpha}$ and $\hat{\beta}$ in Kum$(1, 2)$

	Relative bias				Relative MSE			
	α		β		α		β	
n	IE	MLE	IE	MLE	IE	MLE	IE	MLE
10	-0.0065	0.2288	0.1066	0.6532	0.1707	0.2935	1.7255	8.5038
15	-0.0021	0.1441	0.0487	0.3221	0.1035	0.1506	0.3839	0.8625
20	-0.0047	0.1010	0.0258	0.2098	0.0711	0.0941	0.1906	0.3487
30	-0.0002	0.0682	0.0174	0.1303	0.0449	0.0547	0.0967	0.1444
50	-0.0010	0.0387	0.0088	0.0722	0.0258	0.0290	0.0528	0.0670
80	0.0005	0.0251	0.0057	0.0442	0.0156	0.0169	0.0303	0.0354
100	0.0008	0.0204	0.0048	0.0352	0.0125	0.0134	0.0241	0.0272

Table 4: The relative biases and relative MSEs of the estimators $(\bar{\alpha}, \bar{\beta})$ and CBC when $n = 15$

	Relative bias				Relative MSE			
	α		β		α		β	
(α, β)	$\bar{\alpha}$	CBC	$\bar{\beta}$	CBC	$\bar{\alpha}$	CBC	$\bar{\beta}$	CBC
(1, 0.5)	-0.0063	-0.0489	-0.0064	-0.0611	0.2876	0.2688	0.1291	0.1028
(1, 1)	-0.0024	-0.0222	-0.0115	-0.1162	0.1567	0.1538	0.1671	0.1073
(1, 2)	-0.0017	-0.0148	-0.0270	-0.2670	0.1059	0.1052	0.2092	1.4468
(1, 3)	-0.0016	-0.0130	-0.0505	-0.5132	0.0900	0.0894	0.3132	30.9501

Table 5: The coverage probabilities and average lengths (in parentheses) of the generalized confidence intervals and BC_a bootstrap confidence intervals

		$n = 10$				$n = 20$			
		GCI		BC_a		GCI		BC_a	
(α, β)	parameter	0.90	0.95	0.90	0.95	0.90	0.95	0.90	0.95
(0.5, 0.7)	β	0.9038	0.9527	0.8896	0.9382	0.9017	0.9481	0.8991	0.9467
		(1.2554)	(1.5289)	(1.3185)	(1.6707)	(0.7335)	(0.8809)	(0.7438)	(0.9037)
	$x_{0.1}$	0.9028	0.9513	0.8588	0.9148	0.9003	0.9497	0.8814	0.9363
		(0.1297)	(0.1552)	(0.1282)	(0.1561)	(0.0827)	(0.0990)	(0.0814)	(0.0982)
	μ	0.8975	0.9493	0.8863	0.9351	0.8968	0.9472	0.8939	0.9434
		(0.3054)	(0.3622)	(0.3142)	(0.3752)	(0.2211)	(0.2628)	(0.2253)	(0.2689)
	$R(0.2)$	0.9009	0.9507	0.9127	0.9525	0.9006	0.9509	0.9116	0.9588
		(0.3604)	(0.4236)	(0.4084)	(0.4939)	(0.2618)	(0.3097)	(0.2796)	(0.3364)
(2, 0.7)	β	0.9038	0.9527	0.8896	0.9382	0.9017	0.9481	0.8991	0.9467
		(1.2553)	(1.5289)	(1.3185)	(1.6707)	(0.7335)	(0.8809)	(0.7438)	(0.9037)
	$x_{0.1}$	0.9028	0.9513	0.8599	0.9159	0.9003	0.9497	0.8831	0.9370
		(0.4374)	(0.5076)	(0.3876)	(0.4513)	(0.3218	0.3803)	(0.3010)	(0.3542)
	μ	0.8985	0.9488	0.8664	0.9179	0.8991	0.9493	0.8844	0.9360
		(0.2378)	(0.2867)	(0.2234)	(0.2672)	(0.1670)	(0.2003)	(0.1622)	(0.1939)
	$R(0.2)$	0.9024	0.9516	0.8668	0.9158	0.9006	0.9502	0.8912	0.9379
		(0.1659)	(0.2103)	(0.1331)	(0.1618)	(0.1030)	(0.1284)	(0.0922)	(0.1120)
(0.7, 2)	β	0.9045	0.9527	0.8856	0.9357	0.8997	0.9489	0.8957	0.9439
		(5.9294)	(7.5926)	(6.4576)	(8.7720)	(2.7925)	(3.3929)	(2.8693)	(3.5472)
	$x_{0.1}$	0.9012	0.9503	0.8696	0.9250	0.9006	0.9502	0.8858	0.9386
		(0.0622)	(0.0746)	(0.0608)	(0.0735)	(0.0408)	(0.0487)	(0.0401)	(0.0481)
	μ	0.8985	0.9487	0.8694	0.9183	0.8955	0.9461	0.8843	0.9349
		(0.2281)	(0.2753)	(0.2146)	(0.2564)	(0.1606)	(0.1927)	(0.1566)	(0.1870)
	$R(0.2)$	0.8983	0.9498	0.9262	0.9647	0.8971	0.9485	0.9169	0.9604
		(0.3810)	(0.4465)	(0.4410)	(0.5313)	(0.2776)	(0.3278)	(0.2994)	(0.3593)
(3, 2)	β	0.9045	0.9527	0.8856	0.9357	0.8997	0.9489	0.8957	0.9439
		(5.9294)	(7.5926)	(6.4576)	(8.7719)	(2.7925)	(3.3929)	(2.8693)	(3.5472)
	$x_{0.1}$	0.9012	0.9503	0.8693	0.9246	0.9006	0.9502	0.8867	0.9400
		(0.3151)	(0.3744)	(0.2838)	(0.3344)	(0.2224)	(0.2652)	(0.2122)	(0.2517)
	μ	0.8994	0.9492	0.8572	0.9127	0.8981	0.9485	0.8815	0.9339
		(0.2006)	(0.2425)	(0.1842)	(0.2195)	(0.1407)	(0.1689)	(0.1351)	(0.1610)
	$R(0.2)$	0.9016	0.9508	0.8781	0.9253	0.9007	0.9498	0.8936	0.9428
		(0.1205)	(0.1565)	(0.0994)	(0.1228)	(0.0693)	(0.0880)	(0.0639)	(0.0790)

It is quite clear from the Tables 1–3 that as sample size n increases the average relative biases and average relative MSEs decrease as expected. It is also observed from the Tables 1–3 that as the parameter β increases the average relative biases and average relative MSEs of

$\hat{\alpha}$ decrease, but reverse for $\hat{\beta}$. In all, the simulation results show that the proposed estimators outperform the MLEs for all cases in terms of bias and MSE. But the difference becomes small with large sample size n.

Now let us compare the bias-corrected estimators considered in Section 2.1 with that of Lemonte (2011). Lemonte (2011) considered three bias correction estimators: the bias-corrected estimators (BCE) based on Cox and Snell (1968), preventive bias-corrected estimators (PBC) based on Firth (1993) and constant bias-corrected estimators (CBC) based on bootstrap method. He had concluded that the BCE and PBC estimators cannot be recommended when $1 \leq \beta \leq 3$. Hence we compare $(\bar{\alpha}, \bar{\beta})$ with the CBC proposed by Lemonte (2011). The simulation results are given in Table 4 with 10,000 Monte Carlo replications and $B = 500$. Both the bias-corrected estimators for α are comparable, but the Lemonte's bias-corrected estimator for β suffers stability, even fails in the range of $\beta \in (1, 3)$.

Because the confidence interval for α is exact, we only report the coverage probabilities and average lengths of the generalized confidence intervals (GCI) and BC_a bootstrap confidence intervals at 0.9 and 0.95 confidence levels for β, $x_{0.1}$, μ and $R(0.2)$ in Table 5. These were computed over 10,000 replications for each different case using $m = 10,000$ and $B = 10,000$.

The simulation results show that the simulated probabilities of the GCIs for 0.9 and those for 0.95 are quite close to 0.9 and 0.95 respectively. However, BC_a bootstrap confidence intervals at 0.9 and 0.95 confidence levels at least for $x_{0.1}$ and μ do not perform well for $n = 10$ and not always improve under $n = 20$. Moreover, the interval lengths of the BC_a bootstrap confidence intervals are larger than ones of the proposed GCIs. Therefore, according to these simulation results, we would recommend the proposed GCIs for practical application with small and moderate sample sizes.

3 Estimation for the stress-strength model

Let both the stress and strength variables X and Y be independent and follow the Kumaraswamy distributions with the parameters (α_1, β_1) and (α_2, β_2) respectively. Then the reliability of the stress-strength model is given by

$$R = P(X < Y) = \int_0^1 f_X(x) P(Y > X | X = x) dx$$
$$= \int_0^1 \alpha_1 \beta_1 x^{\alpha_1 - 1} (1 - x^{\alpha_1})^{\beta_1 - 1} (1 - x^{\alpha_2})^{\beta_2} dx$$

$$= \int_0^1 \beta_1(1-t)^{\beta_1-1}(1-t^{\alpha_2/\alpha_1})^{\beta_2}dt.$$

In particular, when $\alpha_1 = \alpha_2$, we have

$$R = \frac{\beta_1}{\beta_1+\beta_2}.$$

3.1 Interval estimation of R

Let $X_{i,1}, X_{i,2}, ..., X_{i,n_i}$ be a random sample from the Kumaraswamy distribution Kum(α_i, β_i), and $\mathbf{X}_i = (X_{i,(1)}, X_{i,(2)}, ..., X_{i,(n_i)})$ are the corresponding order statistics $(i = 1, 2)$. Moreover, let

$$S_{i,j} = \sum_{k=1}^{j} \log(1 - X^{\alpha}_{i,(k)}) + (n_i - j)\log(1 - X^{\alpha}_{i,(j)}),\ j = 1, 2, ..., n_i,$$

$$W_i(\alpha_i) = 2\sum_{j=1}^{n_i-1} \log\frac{S_{i,n_i}}{S_{i,j}},\ i = 1, 2.$$

Similar to the discussion in Section 2, we have the following results:

(1) $W_1(\alpha_1), W_2(\alpha_2), S_{1,n_1}, S_{2,n_2}$ are independent;

(2) $W_i(\alpha_i) \sim \chi^2(2n_i - 2), -2\beta_i S_{i,n_i} \sim \chi^2(2n_i),\ i = 1, 2.$

Since it is easy to obtain MLE for R, the main purpose of this section is to obtain the generalized confidence interval for R.

We first consider interval estimation for R when $\alpha_1 = \alpha_2 \hat{=} \alpha$. In this case, the reliability of the stress-strength model is $R = \beta_1/(\beta_1 + \beta_2)$, and we have that

$$W_3(\alpha) = W_1(\alpha) + W_2(\alpha) \sim \chi^2(2n_1 + 2n_2 - 4).$$

Let $g(W_3, \mathbf{X}_1, \mathbf{X}_2)$ be the solution of the equation $W_3(\alpha) = W_3$, where $W_3 \sim \chi^2(2n_1 + 2n_2 - 4)$. Notice that $V_i = -2\beta_i S_{i,n_i} \sim \chi^2(2n_i)$, we have

$$\beta_i = -\frac{V_i}{2S_{i,n_i}},\ i = 1, 2.$$

Similar to the derivation of Y_1 in Section 2.2, the generalized pivotal quantity for R is given by

$$Y_5 = \frac{V_1/\sum_{j=1}^{n_1}\log(1 - x_{1,(j)}^{g(W_3,\mathbf{X}_1,\mathbf{X}_2)})}{V_1/\sum_{j=1}^{n_1}\log(1 - x_{1,(j)}^{g(W_3,\mathbf{X}_1,\mathbf{X}_2)}) + V_2/\sum_{j=1}^{n_2}\log(1 - x_{2,(j)}^{g(W_3,\mathbf{X}_1,\mathbf{X}_2)})},$$

where $\mathbf{x}_i = (x_{i,(1)}, x_{i,(2)}, ..., x_{i,(n_i)})$ is the observed value of $\mathbf{X}_i$.

Now we consider interval estimation for R when $\alpha_1 \neq \alpha_2$. In this case, let $g_i(W_i, \mathbf{X}_i)$ is the solution of the equation $W_i(\alpha) = W_i$, where $W_i \sim \chi^2(2n_i - 2), i = 1, 2$. Notice that

$V_i = -2\beta_i S_{i,n_i} \sim \chi^2(2n_i)$, using the substitution method, the generalized pivotal quantity for R is given by

$$Y_6 = \int_0^1 T_1(1-t)^{T_1-1}(1-t^{g_2(W_2,\mathbf{X}_2)/g_1(W_1,\mathbf{X}_1)})^{T_2}dt,$$

where

$$T_i = -\frac{V_i}{2\sum_{j=1}^{n_i}\log(1-x_{i,(j)}^{g_i(W_i,\mathbf{X}_i)})},\ i=1,2.$$

Let $Y_{5,\gamma}, Y_{6,\gamma}$ denote the γ percentiles of Y_5, Y_6 respectively. Then $[Y_{5,\gamma/2}, Y_{5,1-\gamma/2}]$ and $[Y_{6,\gamma/2}, Y_{6,1-\gamma/2}]$ are the $1-\gamma$ generalized confidence intervals for R when $\alpha_1 = \alpha_2$ or $\alpha_1 \neq \alpha_2$ respectively. Just as in the case of Y_1, the percentiles of Y_5, Y_6 can be obtained by Monte Carlo simulations.

3.2 Simulation study

In this subsection, a simulation study is conducted to assess the coverage probabilities and average lengths of the proposed generalized confidence intervals for R.

Similar to the discussion in Section 2.4, the distribution of the generalized pivotal quantity Y_5 for R does not depend on the parameter α_1, α_2 when $\alpha_1 = \alpha_2$. Hence, the generalized confidence interval of R based on Y_5 does not depend on the parameter α_1, α_2 when $\alpha_1 = \alpha_2$. Without loss of generality, we take $\alpha_1 = \alpha_2 = 1$ in our simulation study and consider different values of β_1, β_2 when $\alpha_1 = \alpha_2$. For given different values of $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ and different choices of sample sizes, we generated random samples from the Kumaraswamy distributions Kum(α_1, β_1) and Kum(α_2, β_2) respectively.

Table 6: The coverage probabilities and average lengths (in parentheses) of the generalized confidence interval for R

	$(n_1 = 10, n_2 = 10)$		$(n_1 = 10, n_2 = 15)$		$(n_1 = 15, n_2 = 10)$		$(n_1 = 15, n_2 = 15)$	
$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	0.9	0.95	0.9	0.95	0.9	0.95	0.9	0.95
(1, 1, 1, 1)	0.9033	0.9528	0.9011	0.9532	0.9011	0.9478	0.9003	0.9504
	(0.3476)	(0.4091)	(0.3200)	(0.3773)	(0.3200)	(0.3773)	(0.2888)	(0.3412)
(1, 1, 3, 2)	0.9014	0.9513	0.9010	0.9531	0.9010	0.9471	0.8992	0.9487
	(0.3394)	(0.4000)	(0.3133)	(0.3702)	(0.3112)	(0.3670)	(0.2815)	(0.3329)
(1, 1, 2, 1)	0.9018	0.9511	0.9026	0.9528	0.9005	0.9479	0.9008	0.9480
	(0.3222)	(0.3807)	(0.2977)	(0.3528)	(0.2943)	(0.3477)	(0.2664)	(0.3156)
(1, 1, 1.5, 0.5)	0.9021	0.9525	0.9069	0.9535	0.8995	0.9491	0.8991	0.9484
	(0.2868)	(0.3408)	(0.2649)	(0.3157)	(0.2597)	(0.3082)	(0.2352)	(0.2799)
(1, 1, 4, 1)	0.9040	0.9517	0.9087	0.9530	0.9003	0.9480	0.9006	0.9487
	(0.2620)	0.3128)	(0.2415)	(0.2890)	(0.2355)	(0.2805)	(0.2134)	(0.2548)
(3, 2, 2, 1)	0.9185	0.9626	0.9168	0.9616	0.9150	0.9606	0.9154	0.9596
	(0.3798)	(0.4453)	(0.3415)	(0.4018)	(0.3582)	(0.4209)	(0.3162)	(0.3725)
(1.5, 2, 3, 1)	0.9192	0.9621	0.9147	0.9605	0.9121	0.9589	0.9157	0.9590
	(0.3036)	(0.3604)	(0.2713)	(0.3225)	(0.2838)	(0.3378)	(0.2467)	(0.2933)

We report the coverage probabilities and average lengths of the generalized confidence intervals at 0.9 and 0.95 confidence levels for R in Table 6. These were computed over 10,000 replications for each different case using $m = 10,000$. The simulation results show that the simulated probabilities for 0.9 and those for 0.95 are quite close to 0.9 and 0.95 respectively.

4 A proportion data analysis

We illustrate the exact inference of the Kumaraswamy distribution by the analysis of the monthly water capacity data from the Shasta reservoir in California, USA, during the month of February from 1991 to 2010 ($http://cdec.water.ca.gov/reservoir_map.html$). The 20 values of proportions of year capacity are also available from the Table 1 in Nadar et al. (2013). Nadar et al. (2013) showed that these observations follow the Kumaraswamy distribution.

The proposed point estimators ($\hat{\alpha} = 5.7878, \hat{\beta} = 3.6913$) of (α, β) is different from their MLEs ($\hat{\alpha}_M = 6.3476, \hat{\beta}_M = 4.4894$). According to the analysis in Section 2.1 and the simulation results in Section 2.3, the proposed estimators are preferable in terms of biases and MSEs.

After the point estimation, confidence intervals for the parameters can be constructed. The 95% exact confidence interval and asymptotic confidence interval for α are (3.4778, 9.2419) and (3.6433, 9.0518), respectively, without much difference. But the 95% generalized confidence interval and asymptotic confidence interval for β are quite different with (1.7161, 9.7315) and (0.9693, 8.0095) respectively.

5 Conclusion

In this study, we have systematically explored statistical inference procedures for the Kumaraswamy distribution. The inverse estimators was derived. The simulation results in Section 2.3 showed that the biases and MSEs of the proposed estimators are much smaller than the MLEs. A disadvantage of the proposed IEs is that there is not formula for the variances of the proposed IEs, but they can be estimated by the bootstrap method. The pivotal quantity $W(\alpha)$ enables construction of confidence intervals for α. To construct confidence intervals for the mean μ, quantilex_p and the reliability $R(x)$, the method of generalized pivotal quantities was used. The simulation results in Section 2.3 validated the satisfactory performance of the generalized pivotal method.

We then developed the inference procedures for the stress-strength model. When the parameters α_1 and α_2 are equal or unequal, confidence intervals were constructed based on the generalized pivotal quantities Y_5 and Y_6 respectively. The good performance of the generalized confidence intervals was validated in Section 3.2.

6 References

Aminzadeh, M. (1997) Estimation of reliability for exponential stress-strength models with explanatory variables, *Applied Mathematics and Computation*, 84, 269–274.

Baklizi, A. (2008) Likelihood and Bayesian estimation of $Pr(X < Y)$ using lower record values from generalized exponential distribution. *Computational Statistics and Data Analysis*, 52, 3468-3473.

Birnbuam, Z.M. (1956) On a use of the Mann-Whitney statistics. Proceedings of the Third Berkeley Symposium in Mathematics, *Statistics and Probability*, 1, 13-17.

Birnbaum, Z. W. and McCarty, R. C. (1958) A distribution-free upper confidence bound for $Pr(Y < X)$, based on independent samples of X and Y. *The Annals of Mathematical Statistics,* 29(2), 558-562.

Cox, D.R. and Snell, E. (1968) A general definition of residuals (with discussion), *Journal of the Royal Statistical Society. Series B*, 30, 248–275.

Courard-Hauri, D. (2007) Using Monte Carlo analysis to investigate the relationship between overconsumption and uncertain access to one's personal utility function. *Ecological Economics,* 64(1), 152–162.

DiCiccio, T.J. and Efron, B. (1996) Bootstrap confidence intervals. *Statistical Science*, 8, 189–212.

Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.

Fletcher, S.G. and Ponnambalam, K. (1996) Estimation of reservoir yield and storage distribution using moments analysis. Journal of Hydrology, 182(1-4), 259–275.

Ganji, A., Ponnambalam, K. and Khalili, D. (2006) Grain yield reliability analysis with crop water demand uncertainty. *Stochastic Environmental Research and Risk Assessment,* 20(4), 259–277.

Garg, M. (2009) On generalized order statistics from Kumaraswamy distribution, *Tamsui Oxford Journal of Mathematical Sciences,* 25(2), 153–166.

Hauck, W.W., Hyslop, T., Andweson, S. (2000) Generalized treatment effects for clinical trials. *Statistics in Medicine,* 19, 887-899.

Jones, M. C. (2009) Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology,* 6(1), 70–81.
Kotz, S., Lumelskii, Y. and Pensky, M. (2003) *The stress-strength model and its generalizations.* World Scientific Press, Singapore.
Koutsoyiannis, D. and Xanthopoulos, T. (1989) On the parametric approach to unit hydrography identification. Water Resources Management, 3(2), 107–128.
Krishnamoorthy, K. and Lin, Y. (2010) Confidence limits for stress-strength reliability involving Weibull models. *Journal of Statistical Planning and Inference*, 140(7), 1754-1764.
Kundu, D. and Gupta, R.D. (2005) Estimation of $R = P(Y < X)$ for the generalized exponential distribution. *Metrika,* 61(3), 291-308.
Kundu, D. and Gupta, R.D. (2006) Estimation of $R = P(Y < X)$ for Weibull distribution. *IEEE Transaction on Reliability,* 55(2), 270-280.
Lemonte, A.J. (2011) Improved point estimation for the Kumaraswamy distribution. *Journal of Statistical Computation and Simulation,* 81(12), 1971–1982.
Lio, Y.L. and Tsai, T.R. (2012) Estimation of $\delta = P(X < Y)$ for Burr XII distribution based on the progressively first failure-censored samples. *Journal of Applied Statistics*, 39, 309-322.
McCoy, S.J., Beal, J.M., Payton, M.E., Stewart, A.L., DeMers, A.M. and Watson, G.H. (2005) Correlations of visual analog scales with Edinburgh Postnatal Depression Scale, *J. Affect. Disord.*, 86, 295–297.
Mitnik, P.A. (2013) New Properties of the Kumaraswamy Distribution, *Communications in Statistics – Theory and Methods,* 42(5), 741–755.
Nadar, M, Kizilaslan, F. and Alexander Papadopoulos, A. (2014) Classical and Bayesian estimation of $P(Y < X)$ for Kumaraswamy's distribution.*Journal of Statistical Computation and Simulation,* 84, 1505–1529.
Nadar, M., Papadopoulos, A. and Kizilaslan, F. (2013) Statistical analysis for Kumaraswamys distribution based on record data. *Statistical Papers,* 54(2), 355–369.
Ponnambalam, K., Seifi, A. and Vlach, J. (2001) Probabilistic design of systems with general distributions of parameters. *International Journal of Circuit Theory and Applications,* 29(6), 527–536.
Reiser, B. (2000) Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of ROC curves. *Statistics in Medicine,* 19 2115-2129.
Sanchez, S., Ancheyta, J., McCaffrey, W. C. (2007) Comparison of probability distribution functions for fitting distillation curves of petroleum. *Energy & Fuels,* 21(5), 2955–2963.

Seifi, A., Ponnambalam, K. and Vlach, J. (2000) Maximization of manufacturing yield of systems with arbitrary distributions of component values. *Annals of Operations Research*, 99(1-4), 373–383.

Sundar, V.and Subbiah, K. (1989) Application of double bounded probability density-function for analysis of ocean waves. *Ocean Engineering,* 16(2), 193–200.

Surles, J.G. and Padgett, W.J. (2001) Inference for reliability and stress-strength for a scaled Burr Type X distribution. *Lifetime Data Analysis,* 7, 187-200.

Wang, B.X., Yu, K. and Jones, M.C. (2010) Inference under progressively type II right-censored sampling for certain lifetime distributions. *Technometrics,* 52, 453-460.

Weerahandi, S. (2004) *Generalized inference in repeated measures: exact methods in MANOVA and mixed models,* New Jersey: Wiley.