STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, August 2026, pp 940–962.
Published online in International Academic Press (www.IAPress.org)

Using Stein's Loss Function on Bayesian Estimation of Kumaraswamy-Weibull Distribution Parameters with Application

Azhar Hussein Alwan [1], Jinan Abdullah Anber [2], Wadhah Sabri Ibrahim [3], Ahmed Mahdi Salih [4,*]

[1] *College of Basic Education, University of Diyala, Diyala, Iraq*
[2] *Department of Financial and Banking Sciences, Baghdad Technical College of Management, Middle Technical University, Baghdad, Iraq.*
[3] *Department of Statistics, College of Administration Economics, Mustansiriyah University, Baghdad, Iraq*
[4] *Department of Statistics, College of Administration and Economics, Wasit University, Kut, Iraq*
.

Abstract The paper deals with the joint estimation of all the parameters of the Kumaraswamy–Weibull (Kw–W) distribution, namely, the Kumaraswamy shape parameters and the Weibull shape and scale parameters. An objective inference is obtained by using Jeffreys' prior as a noninformative prior, and a Bayesian estimation framework based on Stein's loss function is developed. This work takes a different approach from previous studies that tend to estimate only part of the parameter vector and uses that to obtain partial Bayesian estimators, and derives Bayesian estimators for the entire vector of parameters. The suggested approach is tested via an elaborate simulation investigation and is contrasted with the classical methods such as the maximum likelihood estimation (MLE) and the method of moments (MoM) using the mean squared error (MSE) as a performance metric. In addition, two real lifetime data sets, namely, survival times of bacteria and failure times of iron bars, are analyzed by applying the methodology. The results show that the Bayesian method with Stein loss function yields a more consistent estimation of all parameters, including in moderate and large sample sizes, whereas the classical methods are competitive for small sample sizes. The results show the power of using Stein's loss function with full-parameter Bayesian estimation for flexible lifetime models.

Keywords Kumaraswamy-Weibull Distribution, Bayesian Estimation, Stein's loss function, Jeffreys prior, Life Distributions

AMS 2010 subject classifications 62Jxx

DOI: 10.19139/soic-2310-5070-3697

1. Introduction

Modeling lifetime and reliability data has been an important area of research in statistics, engineering, and applied sciences for many decades. The Weibull distribution, which was introduced by Weibull [23] is one of the earliest and most popular lifetime models. The Weibull distribution is widely used in reliability engineering and survival analysis due to its versatility in modeling changing or decreasing failure rates.

Despite its usefulness, the Weibull distribution may not adequately describe complex data sets with more complicated hazard rate behaviors. To address this limitation, researchers have proposed several generalized distributions. Among these distributions, the distribution that stands out most clearly is Kumaraswamy distribution, Which was presented and derived by Kumaraswamy [13], This has garnered significant attention from researchers due to the simplicity of its mathematical model and the smoothness of its cumulative distribution function.

*Correspondence to: Ahmed Mahdi Salih (Email: amahdi@uowasit.edu.iq). Department of Statistics, College of Administration and Economics, Wasit University, Kut, 50001, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Later, many researchers combined the Kumaraswamy distribution with several other distributions to ensure the flexibility of the new distribution. For example, Cordeiro et al. [4] presented a Kumaraswamy–Weibull distribution in innovative formats as an extension or generalization of the Weibull distribution. The proposed distribution is characterized by its high suitability for modeling lifetime data and for representing hazard and risk functions. including increasing, decreasing, and bathtub forms. Another work by Cordeiro et al. [3] they studied the statistical properties of the Kumaraswamy distribution, as well as the models derived from it, and their uses and applications in the field of reliability for various sciences, including engineering,Pereira et al [18] introduced the Kumaraswamy–Pareto distribution as a flexible heavy-tailed model extending the Pareto family for improved modeling of skewed lifetime and income data. Guven and Senoglu [9] developed statistical inference procedures for the Kumaraswamy–Weibull distribution under doubly Type II censored samples, focusing on parameter estimation and reliability analysis.

One of the earliest works in the field of Bayesian inference was by Jeffreys [11], who introduced a noninformative prior distribution based primarily on the Fisher information matrix. The characteristics of Jeffreys noninformative prior were discussed by Gelman [8] who highlights its general importance in the field of modern statistics, as well as its fundamental role in shaping Bayesian models.

Another interesting aspect of Bayesian estimation theory is the selection of an appropriate loss function. While the quadratic loss function is one of the most important loss functions used in many research studies. Zellner [25] He studied Bayes estimates under symmetric loss functions, discussing their statistical benefits in the field of decision theory. The same researcher further developed his study[24] presented balanced loss functions that combine Bayesian and classical estimation.

In addition to all of the above Diaconis and Freedman [5] A theoretical study in which they tested the consistency of Bayesian estimators They also studied the asymptotic properties and reached very important conclusions. These studies in general highlighted Bayesian estimation theory and encouraged many researchers to use Bayesian estimation in more complex models.

Finally, the researcher Mahmoud [14] studied the generalized Kumaraswamy distribution and highlighted the most important benefits of this distribution in modeling lifetime data. Developments in general have prompted researchers to combine two flexible distributions and obtain more flexible distributions such as Kumaraswamy–Weibull model and study Bayesian estimation techniques for it to improve reliability and survival data.

trhis research focuses on providing Bayesian estimates of shape parameters With the stability of other parameters such as the location parameter and the scale parameter of the Kumaraswamy–Weibull distribution using Jeffreys' prior and Stein's loss function. The behavior of Bayes estimators was then compared with that of classical estimators such as the maximum likelihood estimator MLE and the moment estimator MoM.

The Kumaraswamy–Weibull (Kw–W) distribution It is a four-parameter distribution that combines two distributions Kumaraswamy and Weibull distribution to obtain a more flexible distribution for representing lifetime data. decreasing, bathtub, and unimodal forms. Due of its flexibility, this distribution has wide-ranging applications in the fields of physics and engineering. A random variable X is said to follow a Kumaraswamy–Weibull distribution if its probability density function (PDF) is given by [6]:

$$\begin{aligned} f(x;a,b,\alpha,\beta) &= ab\alpha\beta x^{\alpha-1}\exp\left[-(\beta x)^{\alpha}\right]\left[1-\exp\left(-(\beta x)^{\alpha}\right)\right]^{a-1} \\ &\left[1-\left(1-\exp\left(-(\beta x)^{\alpha}\right)\right)^{a}\right]^{b-1}, \quad x>0, \end{aligned} \tag{1}$$

where $a>0$ and $b>0$ are the Kumaraswamy shape parameters, and $\alpha>0$ and $\beta>0$ are the Weibull shape and scale parameters, respectively. The cumulative distribution function (CDF) of the Kw–W distribution is [2]:

$$F(x;a,b,\alpha,\beta) = 1-\left[1-\left(1-\exp\left(-(\beta x)^{\alpha}\right)\right)^{a}\right]^{b} \tag{2}$$

Once the cumulative distribution function is obtained, we can define the reliability or survival function. In reliability analysis and lifetime modeling, the survival function measures the probability that a system or component continues functioning beyond a specific time. The survival function is simply the complement of the cumulative distribution function. Consequently, Equation 3 expresses the reliability function of the Kumaraswamy–Weibull

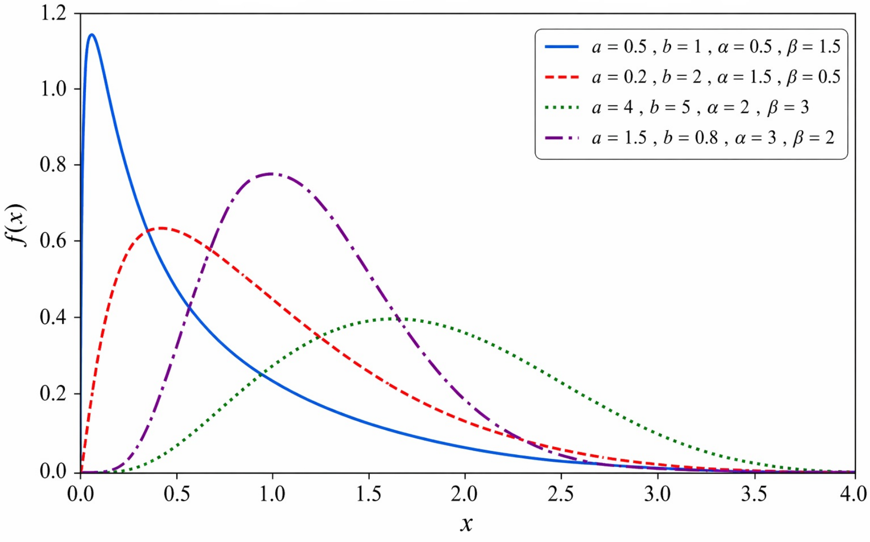

Figure 1. Kw-W distribution with different parameters values

distribution as $1 - F(x)$ [7].

$$R(x) = \left[1 - (1 - \exp\left(-(\beta x)^{\alpha}\right))^{a}\right]^{b}. \tag{3}$$

After defining the survival function, the next important quantity in reliability theory is the hazard rate function. The hazard function describes the instantaneous rate of failure at a particular time given that the system has survived up to that time. It is obtained by dividing the probability density function by the survival function. Thus, Equation 4 presents the hazard rate function of the Kumaraswamy–Weibull distribution, which can exhibit various shapes such as increasing, decreasing, or bathtub-shaped depending on the parameter values [9].

$$h(x) = \frac{f(x)}{R(x)} = \frac{ab\alpha\beta x^{\alpha-1}\exp\left[-(\beta x)^{\alpha}\right]\left[1 - \exp\left(-(\beta x)^{\alpha}\right)\right]^{a-1}}{1 - (1 - \exp\left(-(\beta x)^{\alpha}\right))^{a}}. \tag{4}$$

This function can take on increasing, decreasing, or bathtub shapes depending on parameter combinations, making Kw–W highly adaptable.

2. Some Statistical Properties

The Kw–W distribution It possesses several statistical characteristics that can be studied by identifying parameters about this distribution.[12]: The first moment as follows:

$$\begin{aligned}
&E[X] = \int_0^{\infty} x f(x)\, dx \\
&= a\, b\, \alpha\, \beta \int_0^{\infty} x^{\alpha} \exp\left[-(\beta x)^{\alpha}\right]\left[1 - \exp\left(-(\beta x)^{\alpha}\right)\right]^{(a-1)} \\
&\left[1 - (1 - \exp\left(-(\beta x)^{\alpha}\right))^{a}\right]^{(b-1)}\, dx
\end{aligned} \tag{5}$$

To better understand the behavior of this distribution, it is necessary to study the moments of this distribution, Equation 5 represents the first moment, or the expected value (mean) [21], of the random variable. The mean represents one measure of the central tendency of the data, and naturally, solving above integral is complex and requires numerical methods. For this reason, a more general expression for the r-th moment is introduced. Equation 6 expresses the r-th moment using the Beta function and binomial coefficients, which allows numerical or analytical evaluation of higher-order moments such as variance, skewness, and kurtosis [16].

$$E[X^r] = (b/\beta^r) \sum_{j=0}^{\infty} (-1)^j C(b-1,j) B(1 + r/\alpha, a(j+1)) \tag{6}$$

where $B(\cdot,\cdot)$ denotes the Beta function and $C(\cdot,\cdot)$ represents the binomial coefficient. Variance and Skewness and Kurtosis can be computed from the central moments, providing insight into asymmetry and tail heaviness [19].

3. Classical Estimation Methods

Two common methods for estimating the parameters of the Kw–W distribution are the Maximum Likelihood Estimation (MLE) and the Method of Moments (MoM).

3.1. *Maximum Likelihood Estimation (MLE)*

Let $X_1, \ldots, X_n$ be i.i.d. observations from the Kumaraswamy–Weibull (Kw–W) distribution defined as the Kumaraswamy transform of a Weibull baseline. Weibull baseline CDF and PDF[22]:

$$G(x) = 1 - \exp(-t), \quad g(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp(-t),$$

where $t = \left(\frac{x}{\beta}\right)^{\alpha}$, $x > 0$, $\alpha > 0$, $\beta > 0$. The Kumaraswamy–G construction with parameters $a > 0$, $b > 0$ yields CDF

$$F(x) = 1 - (1 - G(x)^a)^b$$

and PDF

$$f(x) = abg(x)G(x)^{a-1}(1 - G(x)^a)^{b-1}.$$

For each observation i:

$$t_i = \left(\frac{x_i}{\beta}\right)^{\alpha}, \quad G_i = 1 - e^{-t_i}.$$

then the Log-Likelihood will be [20]

$$\begin{aligned} \ell(\psi) &= \sum \log f(x_i) = n\log a + n\log b + n\log\alpha - n\log\beta \\ &+(\alpha-1)\sum\log(x_i/\beta) - \sum t_i + (a-1)\sum\log G_i \\ &+(b-1)\sum\log(1-G_i^a) \end{aligned} \tag{7}$$

To estimate the parameters of the Kumaraswamy–Weibull distribution, the Maximum Likelihood Estimation method is applied. Equation 7 represents the log-likelihood function constructed from a random sample drawn from the distribution. Maximizing this function with respect to the parameters leads to the most likely parameter estimates. To find these estimates, we differentiate the log-likelihood function with respect to each parameter. Equations 8 - 11 present the partial derivatives of the log-likelihood function with respect to a, b, α, and β. Setting these derivatives equal to zero yields a system of nonlinear equations that must be solved numerically [10].

$$\frac{\partial\ell}{\partial a} = \frac{n}{a} + \Sigma\log G_i + (b-1)\Sigma\left[-\frac{G_i^a\log G_i}{1-G_i^a}\right] \tag{8}$$

$$\frac{\partial\ell}{\partial b} = \frac{n}{b} + \Sigma\log(1-G_i^a) \tag{9}$$

$$\begin{aligned} \tfrac{\partial\ell}{\partial\alpha} &= \tfrac{n}{\alpha} + \sum\log\left(\tfrac{x_i}{\beta}\right) - \sum t_i\log\left(\tfrac{x_i}{\beta}\right) \\ &+(a-1)\sum\left[e^{-t_i}t_i\log\left(\tfrac{x_i}{\beta}\right)/G_i\right] \\ &-(b-1)a\sum\left[G_i^{a-1}e^{-t_i}t_i\log\left(\tfrac{x_i}{\beta}\right)/(1-G_i^a)\right] \end{aligned} \tag{10}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} = -\left(\frac{\alpha n}{\beta}\right) + \left(\frac{\alpha}{\beta}\right)\sum t_i \\ -(a-1)\left(\frac{\alpha}{\beta}\right)\sum \left[e^{-t_i} t_i / G_i\right] \\ +(b-1)a\left(\frac{\alpha}{\beta}\right)\sum \left[G_i^{a-1} e^{-t_i} t_i / (1-G_i^a)\right] \end{aligned} \tag{11}$$

The MLEs of the parameters $(\hat{a}, \hat{b}, \hat{\alpha}, \hat{\beta})$ are obtained by solving the system of nonlinear equations formed by setting the partial derivatives of the log-likelihood with respect to each parameter equal to zero. These equations are generally solved numerically using iterative optimization algorithms such as the Newton–Raphson or Expectation–Maximization (EM) method [1].

3.2. *Method of Moments Estimation (MoM)*

The method of moments equates the theoretical moments $E[X^r]$ with the corresponding sample moments $M_r = \frac{1}{n}\sum x_i^r$. The resulting equations can be solved simultaneously to estimate the parameters. Although MoM is simpler, it is less efficient and less robust than MLE, especially for small samples [15].

4. Bayesian Estimation

Bayesian estimation provides an alternative to classical methods by incorporating prior information about the parameters of the Kumaraswamy–Weibull (Kw–W) distribution. This approach combines the prior distribution with the likelihood function derived from the observed data to obtain the posterior distribution of the parameters [27]. Let $X_1, X_2, \ldots, X_n$ be a random sample from the Kw–W distribution with parameters $\beta = (a, b, \alpha, \beta)$. The posterior distribution is given by:

$$\pi(\beta \mid x) \propto L(\beta) \times \pi(\beta),$$

where $L(\beta)$ is the likelihood function and $\pi(\beta)$ represents the prior distribution for the parameters. Jeffreys' prior is a principled noninformative prior derived from the Fisher information. For a parameter vector β, the Jeffreys prior is defined as [17].

$$\pi_J(\beta) \propto \sqrt{\det\left(I(\beta)\right)} \tag{12}$$

In the Bayesian framework, prior information about the parameters is incorporated through a prior distribution. Jeffreys' prior is commonly used as a non-informative prior because it is invariant under reparameterization. Equation 12 defines Jeffreys' prior in terms of the Fisher information matrix. The Fisher information measures the amount of information that the observed data provide about the unknown parameters. Equation 13 defines the elements of the Fisher information matrix using the expected value of the second derivatives of the log-likelihood function.

$$I_{jk}(\beta) = -E\left[\frac{\partial^2}{\partial \beta_j \partial \beta_k} \ell(\beta; X)\right] \tag{13}$$

and $\ell(\beta; X)$ is the log-likelihood function. Jeffreys' prior is invariant under reparameterization, making it a common choice for objective Bayesian analysis. For the Kumaraswamy–Weibull (Kw–W) distribution with parameters $\beta = (a, b, \alpha, \beta)$ the liklihood function 7, Analytic derivation of the Fisher information matrix for the Kw–W model is algebraically involved and does not lead to simple closed-form expressions. Consequently, Jeffreys' prior for Kw–W is typically computed numerically. A practical approach is to compute the observed information (negative Hessian of the log-likelihood) at each parameter value and use it as an approximation to the expected Fisher information. The observed information matrix is:

$$J(\beta) = -\begin{bmatrix} \frac{\partial^2 \ell}{\partial a^2} & \frac{\partial^2 \ell}{\partial a \partial b} & \cdots \\ \vdots & & \ddots \end{bmatrix}_{\beta}. \tag{14}$$

Then the Jeffreys-like prior can be approximated by:

$$\pi_J(\beta) \propto \sqrt{\det\left(J(\beta)\right)}.$$

This approximation is widely used in practice when the expectation is intractable.

5. Bayesian Approach for Kw-W Distribution

In this section, the Bayesian estimation procedure is extended to the full parameter vector $\boldsymbol{\theta} = (a, b, \alpha, \beta)$ of the Kumaraswamy–Weibull (Kw–W) distribution. Let $X_1, \dots, X_n$ be an i.i.d. sample from the Kw–W distribution with probability density function

$$f(x;\theta) = ab\alpha\beta x^{\alpha-1} e^{-(\beta x)^\alpha}\left[1 - e^{-(\beta x)^\alpha}\right]^{a-1}\left[1 - \left(1 - e^{-(\beta x)^\alpha}\right)^a\right]^{b-1}.$$

The likelihood function is therefore given by

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i;\theta), \tag{15}$$

and the log-likelihood can be written as

$$\begin{aligned}\ell(\theta) = n\log a + n\log b + n\log\alpha + n\log\beta + (\alpha-1)\sum\log x_i - \sum(\beta x_i)^\alpha \\ +(a-1)\sum\log\left(1 - e^{-(\beta x_i)^\alpha}\right) + (b-1)\sum\log\left[1 - \left(1 - e^{-(\beta x_i)^\alpha}\right)^a\right].\end{aligned} \tag{16}$$

The Fisher information matrix for the full parameter vector is defined as

$$I(\theta) = -E\left[\frac{\partial^2\ell(\theta)}{\partial\theta_i\partial\theta_j}\right], \quad i,j = 1,\dots,4. \tag{17}$$

Thus, $I(\theta)$ takes the form

$$I(\theta) = \begin{pmatrix} I_{aa} & I_{ab} & I_{a\alpha} & I_{a\beta} \\ I_{ba} & I_{bb} & I_{b\alpha} & I_{b\beta} \\ I_{\alpha a} & I_{\alpha b} & I_{\alpha\alpha} & I_{\alpha\beta} \\ I_{\beta a} & I_{\beta b} & I_{\beta\alpha} & I_{\beta\beta} \end{pmatrix}. \tag{18}$$

Each element is obtained from the expected second derivatives:

$$I_{aa} = \frac{n}{a^2} + E\left[\sum\frac{(b-1)G_i^a(\log G_i)^2}{(1-G_i^a)^2}\right], \tag{19}$$

$$I_{bb} = \frac{n}{b^2}, \tag{20}$$

$$I_{ab} = E\left[\sum\frac{G_i^a\log G_i}{1-G_i^a}\right], \tag{21}$$

where $G_i = 1 - e^{-(\beta x_i)^\alpha}$.and for the Weibull parameters, we obtain:

$$I_{\alpha\alpha} = \frac{n}{\alpha^2} + E\left[\sum\left(\frac{\partial t_i}{\partial\alpha}\right)^2\right], \quad t_i = (\beta x_i)^\alpha, \tag{22}$$

$$I_{\beta\beta} = \frac{n\alpha}{\beta^2} + E\left[\sum\left(\frac{\partial t_i}{\partial\beta}\right)^2\right], \tag{23}$$

$$I_{\alpha\beta} = E\left[\sum \frac{\partial t_i}{\partial \alpha}\frac{\partial t_i}{\partial \beta}\right]. \tag{24}$$

The cross-information terms between (a, b) and (α, β) arise from the dependence of G_i on α, β:

$$I_{a\alpha} = E\left[\sum \frac{\partial}{\partial \alpha}\log\left(1 - e^{-t_i}\right)\right], \tag{25}$$

$$I_{a\beta} = E\left[\sum \frac{\partial}{\partial \beta}\log\left(1 - e^{-t_i}\right)\right], \tag{26}$$

$$I_{b\alpha} = E\left[\sum \frac{\partial}{\partial \alpha}\log\left(1 - G_i^a\right)\right], \tag{27}$$

$$I_{b\beta} = E\left[\sum \frac{\partial}{\partial \beta}\log\left(1 - G_i^a\right)\right]. \tag{28}$$

Due to the analytical intractability of these expectations, the Fisher information matrix is typically approximated using the observed information matrix:

$$J(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^T}. \tag{29}$$

The Jeffreys prior for the full parameter vector is given by

$$\pi_J(\theta) \propto \sqrt{\det I(\theta)} \tag{30}$$

Due to the complexity of $I(\theta)$, the determinant is computed numerically:

$$\pi_J(a, b, \alpha, \beta) \propto \sqrt{\det J(\theta)}.$$

Assuming independence of parameters a priori, the posterior distribution is

$$\pi(\theta|x) \propto L(\theta|x)\,\pi_J(\theta),$$

that is

$$\pi(a, b, \alpha, \beta|x) \propto \left[\prod_{i=1}^{n} f(x_i;\theta)\right]\sqrt{\det J(\theta)}. \tag{31}$$

Using Stein's loss function, Stein's loss function is an asymmetric loss function commonly used in Bayesian estimation for positive parameters such as shape and scale parameters. Unlike squared error loss, it penalizes overestimation and underestimation differently, making it more suitable for reliability and lifetime data analysis. In this paper, it is applied to estimate all parameters of the Kumaraswamy–Weibull distribution within a Bayesian framework. One key advantage is that it leads to estimators based on the reciprocal of parameters, which often improves stability and reduces bias. It is also invariant under scale transformations, making it appropriate for positive-valued parameters. The use of Stein's loss function enhances estimation accuracy, especially for moderate and large samples. Overall, it provides more reliable and efficient estimates compared to symmetric loss functions [26].

$$L(\theta, \delta) = \frac{\delta}{\theta} - \log\left(\frac{\delta}{\theta}\right) - 1, \tag{32}$$

the Bayes estimators for each parameter are obtained as

$$\begin{aligned} \hat{a}_{\text{Bayes}} &= \tfrac{1}{E[1/a|x]}, & \hat{b}_{\text{Bayes}} &= \tfrac{1}{E[1/b|x]}, \\ \hat{\alpha}_{\text{Bayes}} &= \tfrac{1}{E[1/\alpha|x]}, & \hat{\beta}_{\text{Bayes}} &= \tfrac{1}{E[1/\beta|x]}. \end{aligned} \tag{33}$$

These expectations are computed numerically as

$$E\left[\frac{1}{a} \mid x\right] = \frac{\int\int\int\int \frac{1}{a} L(\theta|x)\sqrt{\det J(\theta)}\, d\theta}{\int\int\int\int L(\theta|x)\sqrt{\det J(\theta)}\, d\theta}, \tag{34}$$

and similarly for b, α, β. Due to the high dimensionality and nonlinearity, numerical techniques such as Markov Chain Monte Carlo (MCMC) or Newton–Raphson integration are required.

6. Simulation Results

By generating data for the Kw-W distribution using inverse transform by CDF function, evaluating the Maximum Likelihood Estimator (MLE), the Moment Estimator (MoM), and the Bayesian Estimator (BE) for the following set of parameters values Table 1 and repeating the experiment 3000 times to obtain the following results using Mean Square Errors MSE.

	a	b	α	β
Case I	0.5	1.0	0.5	1.5
Case II	0.2	2.0	1.5	0.5
Case III	4.0	5.0	2.0	3.0
Case IV	1.5	0.8	3.0	2.0

Table 1. Sets of Parameters values

n	MLE	ME	BE
5	1.77393	1.42846	1.57302
10	1.41227	1.26011	1.49331
20	1.15027	1.12434	1.26112
30	0.95848	1.02634	1.10027
40	0.82646	0.94762	0.86502
60	0.72905	0.89511	0.61446
80	0.66551	0.86502	0.51788
100	0.62151	0.84552	0.46208
150	0.59380	0.83593	0.43648
200	0.58042	0.82646	0.42664

Table 2. MSE for methods Case I

Table 2 presents the Mean Squared Error (MSE) values for the three estimation methods (MLE, MoM, and Bayesian estimation) under Case I, and it is clear that all methods exhibit a consistent decrease in MSE as the sample size increases, indicating improved estimation accuracy with larger samples. In small samples ($n = 5$ and 10), the Method of Moments shows relatively lower MSE compared to MLE and Bayesian estimation, while the Bayesian estimator initially performs less efficiently. However, as the sample size increases beyond $n = 40$, the Bayesian estimator begins to outperform both MLE and MoM, achieving the smallest MSE values. They are asymptotically efficient, for large samples. Table 3 supports this conclusion by showing the bias of the parameter estimates: As the sample size increases, the parameter estimates (a, b, α, β) become less biased. For large n, the bias values approach zero, which shows consistency, and better reliability is true especially for the Bayesian estimator. Moreover, the Bayesian estimates also have smaller magnitudes of bias in moderate and large samples than the other estimates are, further justifying their superiority. Figure 2 supports the trends of Table 2, showing that the MSE decreases as N increases for all estimation methods. The figure indicates that the Bayesian curve is initially higher in value but drops rapidly and becomes lower than the MLE curve and MoM curve for higher n. The MoM

n	MLE				ME				BE			
	a	b	α	β	a	b	α	β	a	b	α	β
5	0.587	1.829	0.842	1.861	0.414	1.414	0.601	1.343	0.587	1.588	0.851	1.761
10	0.406	1.395	0.589	1.318	0.331	1.212	0.482	1.091	0.547	1.492	0.795	1.641
20	0.275	1.081	0.405	0.925	0.262	1.049	0.387	0.887	0.431	1.213	0.633	1.292
30	0.179	0.851	0.271	0.638	0.213	0.932	0.318	0.741	0.351	1.021	0.521	1.051
40	0.113	0.692	0.179	0.441	0.174	0.837	0.263	0.621	0.233	0.738	0.356	0.698
60	0.065	0.575	0.111	0.294	0.148	0.774	0.227	0.543	0.107	0.437	0.181	0.322
80	0.033	0.499	0.066	0.198	0.133	0.738	0.206	0.498	0.059	0.321	0.113	0.177
100	0.011	0.446	0.035	0.132	0.123	0.715	0.192	0.468	0.031	0.254	0.073	0.093
150	-0.003	0.413	0.016	0.091	0.118	0.703	0.185	0.454	0.018	0.224	0.056	0.055
200	-0.011	0.397	0.006	0.071	0.113	0.692	0.179	0.441	0.013	0.212	0.049	0.041

Table 3. Bias for the Parameters Case I

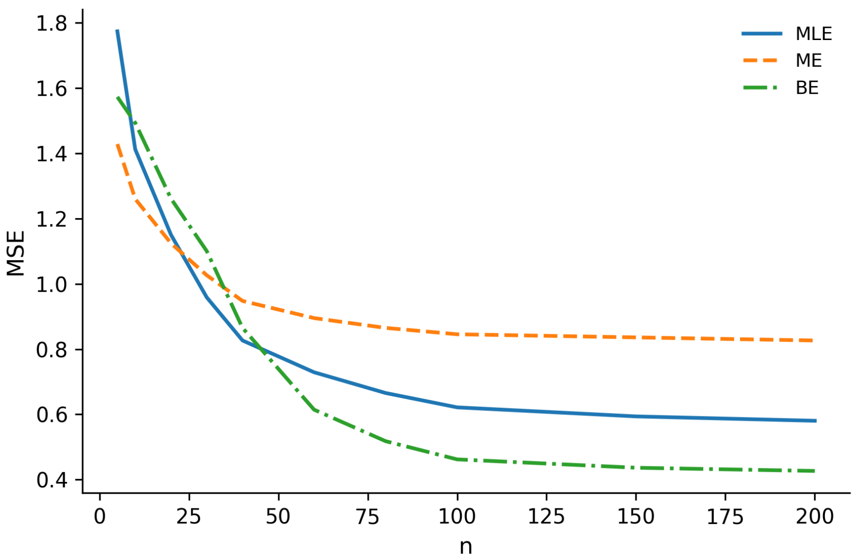

Figure 2. MSE for methods Case I

curve, on the other hand, declines more gradually and stays higher than the rest which means its performance isn't that great. In large samples, the MLE curve is a moderate improvement but not as efficient as the Bayesian estimator. In general, the results given in Table 2, Table 3 and Figure 2 show that for large sample sizes the method with Stein's loss function seems more accurate, less biased and more stable than classical method, although for small sample sizes the classical method may be better.

Table 4 presents the Mean Squared Error (MSE) results for Case II and shows a clear downward trend in MSE values for all estimation methods (MLE, MoM, and Bayesian estimation) as the sample size increases, indicating With larger samples, the performance of the estimators improves. The MLE method has the smallest MSE at small sample size ($n = 5$ or 10) but the largest MSE values for the Bayesian estimator, which shows initial instability for small samples. But when $n > 30$, the MSE of the Bayesian estimator shows a significant decrease and eventually becomes smaller than that of the MLE and MoM for moderate and large sample sizes. The Method of Moments generally has higher MSE values for most sample sizes, which is not as efficient as the other methods. Table 5 offers additional evidence by performing a bias analysis on the parameter estimates (a, b, α, β) for all methods, which show that the bias is decreasing with increasing sample size, which is consistent. Of particular interest, the

n	MLE	ME	BE
5	1.027984	3.043475	4.030194
10	0.829609	2.206483	3.610793
20	0.687701	1.665291	2.059382
30	0.585552	1.290977	1.143932
40	0.519026	1.191246	0.852143
60	0.466265	1.000800	0.601457
80	0.424518	0.875290	0.472555
100	0.386509	0.796920	0.413292
150	0.381364	0.755330	0.364352
200	0.376288	0.735356	0.356337

Table 4. MSE for methods Case II

n	MLE				ME				BE			
	a	b	α	β	a	b	α	β	a	b	α	β
5	0.314	1.556	0.522	0.836	1.222	3.352	1.73	3.765	1.815	4.536	2.571	5.445
10	0.215	1.159	0.364	0.578	0.803	2.348	1.145	2.51	1.605	4.033	2.278	4.816
20	0.144	0.875	0.25	0.394	0.533	1.698	0.766	1.698	0.83	2.171	1.192	2.489
30	0.093	0.671	0.168	0.261	0.345	1.249	0.504	1.136	0.372	1.073	0.551	1.116
40	0.06	0.538	0.115	0.175	0.296	1.129	0.434	0.987	0.226	0.723	0.347	0.678
60	0.033	0.433	0.073	0.106	0.2	0.901	0.301	0.701	0.101	0.422	0.171	0.302
80	0.012	0.349	0.04	0.052	0.138	0.75	0.213	0.513	0.036	0.267	0.081	0.109
100	-0.007	0.273	0.009	0.002	0.098	0.656	0.158	0.395	0.007	0.196	0.039	0.02
150	-0.009	0.263	0.005	-0.004	0.078	0.606	0.129	0.333	-0.018	0.137	0.005	-0.053
200	-0.012	0.253	0.001	-0.011	0.068	0.582	0.115	0.303	-0.022	0.128	-0.001	-0.065

Table 5. Bias for Parameters Case II

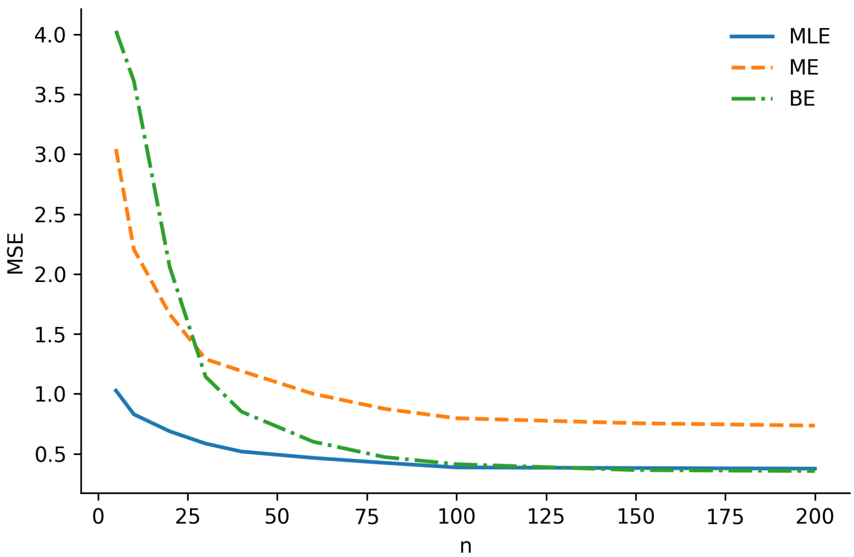

Figure 3. MSE for methods Case II

Bayesian estimator exhibits relatively high bias in small samples but exhibits substantially reduced bias values as n increases until it has smaller values of bias than MLE and MoM. MLE is more stable with moderate bias for various sample sizes, whereas MoM has larger and consistent bias, particularly for some of the parameters. These results are further confirmed visually, in figure 3, by showing the MSE behaviour with respect to the sample size for Case II where all curves are decreasing as the sample size increases. The Bayesian curve is already high at the beginning but falls very quickly and turns to be the minimum one of the three for more samples, implying better long-range performance. The MLE curve continues to rise steadily but is still above. The MoM curve falls less steeply and remains at a higher level, indicating its lower efficiency, compared to the Bayesian curve in large samples. Overall, the interpretation of the results in Table 4, 5 and Figure 3 illustrates that MLE is best in small samples but that the Bayesian estimation based on Stein's loss function gains in accuracy, lack of bias and efficiency in this case as sample sizes grow.

n	MLE	ME	BE
5	0.967152	4.380663	7.949062
10	0.684409	2.402717	4.787098
20	0.502279	1.443398	2.317848
30	0.382281	0.932581	0.924877
40	0.301737	0.648042	0.561968
60	0.246992	0.484325	0.238163
80	0.213525	0.389302	0.174785
100	0.191436	0.336553	0.143073
150	0.177995	0.307279	0.128272
200	0.171632	0.296295	0.123687

Table 6. MSE for methods Case III

n	MLE				ME				BE			
	a	b	α	β	a	b	α	β	a	b	α	β
5	0.674	1.634	0.474	1.634	6.271	4.957	2.866	4.857	5.464	9.439	6.259	11.029
10	0.448	1.069	0.248	1.069	3.304	2.583	1.482	2.483	3.251	5.645	3.73	6.602
20	0.302	0.705	0.102	0.705	1.865	1.432	0.81	1.332	1.522	2.681	1.754	3.145
30	0.206	0.465	0.006	0.465	1.099	0.819	0.453	0.719	0.547	1.01	0.64	1.195
40	0.141	0.303	-0.059	0.303	0.672	0.478	0.254	0.378	0.293	0.574	0.35	0.687
60	0.098	0.194	-0.102	0.194	0.426	0.281	0.139	0.181	0.067	0.186	0.091	0.233
80	0.071	0.127	-0.129	0.127	0.284	0.167	0.073	0.067	0.022	0.11	0.04	0.145
100	0.053	0.083	-0.147	0.083	0.205	0.104	0.036	0.004	0.001	0.072	0.014	0.1
150	0.042	0.056	-0.158	0.056	0.161	0.069	0.015	-0.031	-0.012	0.054	0.003	0.08
200	0.037	0.043	-0.163	0.043	0.144	0.056	0.007	-0.044	-0.013	0.048	-0.001	0.073

Table 7. Bias for Parameters Case III

Table 6 displays the Mean Squared Error (MSE) results of Case III and displays a clear and steady decrease in MSE values as the sample size increases with each estimation technique (MLE, MoM, and Bayesian estimation), suggesting an increase in accuracy of the estimation with an increase in sample size. From the result of the MSE in the table above, the MLE has the smallest MSE value for very small sample size ($n = 5$ and 10) which show better short-sample performance and the BE has a significantly larger value of MSE which shows initial instability for a limited amount of data. As the sample size grows to moderate size ($n = 30$–40), however, the Bayesian estimator starts to perform well and the difference in performance between the methods decreases. The Bayesian estimator obviously has the smallest MSE values for larger sample sizes ($n = 60$), which shows the advantage in large samples in both efficiency and having the smallest MSE value as compared to MLE and MoM. The Method

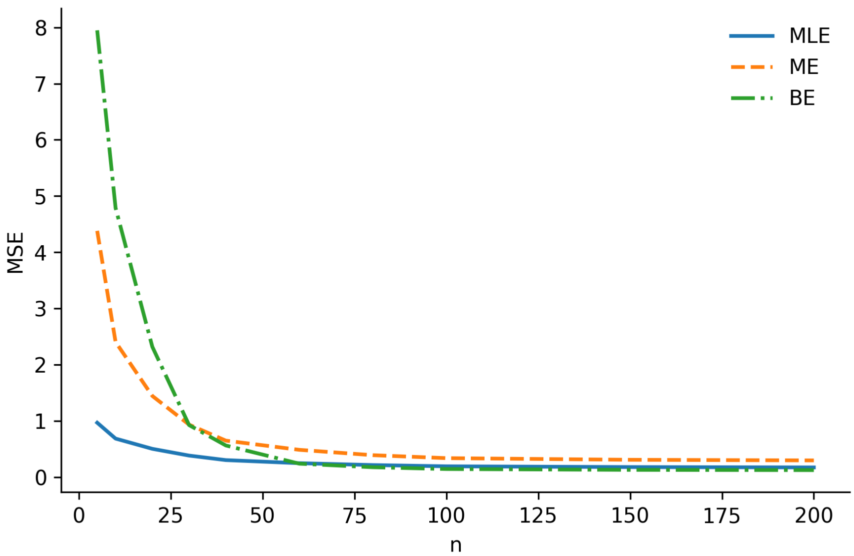

Figure 4. MSE for methods Case III

of Moments always has the highest MSE values for all of the sizes, which further shows its relatively poorer performance. The bias values for the parameters (a, b, α, β) both are given in table 7 and it can be seen that the bias is monotonically decreasing with the sample size for all the methods and is consistent. The MLE is moderately biased while the MoM has larger and more persistent bias especially for small and moderate samples. The Bayesian estimator is biased initially but this bias decreases quickly with sample size and for large sample sizes the bias is the smallest of the three given estimators. It demonstrates the power of Bayesian estimation in reducing systematic error with increasing data, to achieve this. Figure 4 The above observations are visually reinforced in the above graph which displays the decrease in the mean square error with the increase in sample size for all the methods. The Bayesian curve is initially high, but then falls rapidly, and for large sample sizes is the lowest curve, showing good asymptotic performance. In large samples, the MLE curve falls steadily but has a higher peak than the Bayesian curve, whereas the MoM curve falls more slowly, and has a higher peak and is less efficient. The overall conclusion from Table 6, Table 7 and Figure 4 is that MLE is better in small samples, whereas the Bayesian under the Stein loss function is better in large samples in terms of being more accurate, less biased, and more efficient in estimating the parameters in practical large-sample applications.

n	MLE	ME	BE
5	0.711201	2.250607	8.669531
10	0.399796	1.340175	5.384048
20	0.267135	0.861741	2.845354
30	0.189077	0.554106	0.892193
40	0.141763	0.384735	0.274171
60	0.112591	0.282974	0.119266
80	0.094723	0.220469	0.081236
100	0.084416	0.181954	0.062088
150	0.078175	0.162155	0.053248
200	0.075231	0.153079	0.050267

Table 8. MSE for methods Case IV

n	MLE				ME				BE			
	a	b	α	β	a	b	α	β	a	b	α	β
5	0.753	1.557	1.934	1.678	3.276	2.501	1.505	2.551	5.48	9.439	6.269	11.029
10	0.38	0.81	0.999	0.899	1.91	1.408	0.868	1.458	3.267	5.645	3.74	6.602
20	0.221	0.491	0.601	0.568	1.193	0.834	0.533	0.884	1.538	2.681	1.764	3.145
30	0.127	0.304	0.367	0.373	0.731	0.465	0.318	0.515	0.563	1.01	0.65	1.195
40	0.07	0.19	0.225	0.254	0.477	0.262	0.199	0.312	0.309	0.574	0.36	0.687
60	0.035	0.12	0.138	0.181	0.324	0.14	0.128	0.19	0.083	0.186	0.101	0.233
80	0.014	0.077	0.084	0.137	0.231	0.065	0.084	0.115	0.038	0.11	0.05	0.145
100	0.001	0.053	0.053	0.111	0.173	0.018	0.057	0.068	0.016	0.072	0.024	0.1
150	-0.006	0.038	0.035	0.095	0.143	-0.005	0.044	0.045	0.006	0.054	0.013	0.08
200	-0.01	0.031	0.026	0.088	0.13	-0.016	0.037	0.034	0.003	0.048	0.009	0.073

Table 9. Bias for Parameters Case IV

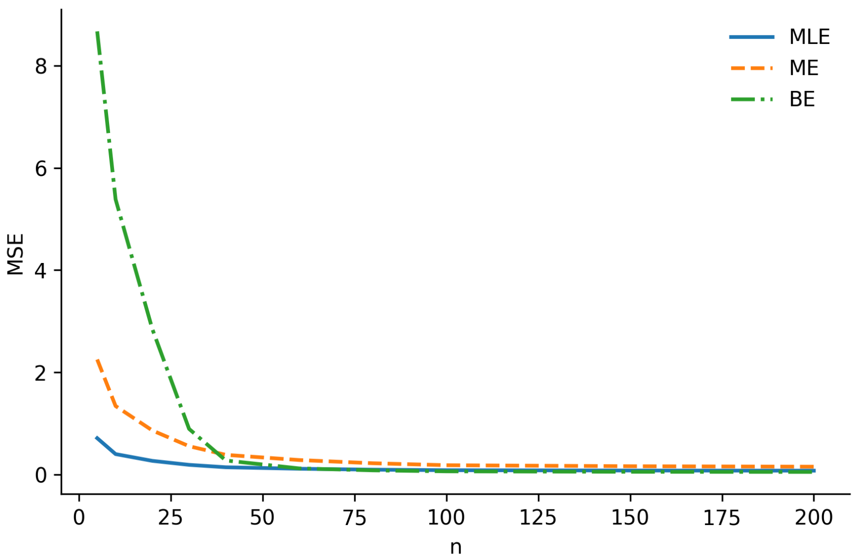

Figure 5. MSE for methods Case IV

Table 8 presents the Mean Squared Error (MSE) results for Case IV and demonstrates a clear decreasing pattern in MSE values for all estimation methods (MLE, MoM, and Bayesian estimation) as the sample size increases, confirming improved estimation accuracy with larger samples. For very small sample sizes ($n = 5$ and 10), the Maximum Likelihood Estimator (MLE) shows the lowest MSE values, indicating better short-sample efficiency, while the Bayesian estimator (BE) exhibits significantly higher MSE due to initial variability and sensitivity to limited data. As the sample size increases to moderate levels ($n =$ 30–40), the performance gap between the estimators begins to narrow, and the Bayesian estimator improves substantially. For larger sample sizes ($n \geq 60$), It is clear that the Bayesian estimator is clearly the best estimator of all the three with lower MSE values than MLE and MoM. The Method of Moments results in larger MSE values for all sample sizes, indicating less good estimation ability. The results of the biases of the parameters (a, b, α, β) are given in Table 9 and are consistent in that the biases are decreasing with increasing sample size for all methods. The MLE is somewhat less biased, and the MoM is somewhat more biased, particularly in small sample sizes. The Bayesian estimator has relatively high bias at the small sample size, but improves very quickly as the sample size increases and this estimator has the smallest bias at large samples. This indicates how well Bayesian estimation works in minimizing systematic error when there is enough data. The findings are visually confirmed in Figure 5, which can be seen that all curves are decreasing as n increases. The Bayesian curve is high at low sample sizes, then drops off precipitously and is the lowest curve for large sample sizes, so it is highly asymptotically efficient. The MLE curve will initially fall gradually but will never be below the Bayesian curve, whereas the MoM curve will fall slower and will always be above the Bayesian curve, which means it is less efficient. In summary, the results in Table 8, Table 9 and Figure 5 indicate that MLE is superior for small samples, but the Bayesian estimation under Stein loss function is more accurate, less biased and more

7. Real Data

In this section, two sets of real data were obtained. The first set included the death times of a group of tuberculosis bacteria, measured in hours, when exposed to a specific type of antibiotic. The second set included the fracture times of a group of iron bars when subjected to a specific pressure. These two sets of data were used to cover most of the cases in which life distributions are used.

Data Set I

29	25	50	18	5	8
27	13	1	21	19	15
18	36	15	14	39	6
15	7	15	12	39	14
7	21	15	14	70	44

Table 10. Death times of tuberculosis bacteria

Data Set II

201	166	146	116	188	153	133	86
208	171	148	118	191	155	136	100
211	175	148	122	191	157	136	107
213	179	148	126	192	166	136	113
213	183	148	126	192	166	144	114

Table 11. Iron bars pressure.

The outcome is shown in the histogram of Data Set I 6 (death times of tuberculosis bacteria exposed to a particular antibiotic) and the fitted distribution of the Kumaraswamy–Weibull (Kw–W). The histogram illustrates the empirical distribution of the observed lifetimes and the theoretical one is overlaid, which is the Kw–W distribution for the estimated parameters. This can be seen to be generally an appropriate shape for the data, with the majority of data points occurring in the lower and middle portions and the data points in the tail following a reasonable curve. This suggests that the Kumaraswamy–Weibull distribution can describe the variability and skewness of the data on bacterial lifetime. The fit of the proposed model to empirical data is supported by the visual agreement between the empirical histogram and the fitted curve. Figure. 7 shows the histogram for Data Set II, which is a set of data for fracture. The fitted Kumaraswamy–Weibull distribution and times of iron bars subjected to a specific pressure. The empirical frequency distribution is displayed by the histogram of the observed failure times and the curve overlaid is a theoretical density function based on the estimated parameters of Kw and W. The fitted curve follows the general pattern of the histogram, especially in the middle, where most of the observations are. This agreement suggests that the Kumaraswamy–Weibull model is a good approximation of the underlying distribution of fracture-time data. Therefore, the graphical fit demonstrates the flexibility of the Kw–W distribution in modeling engineering reliability data and supports its application in lifetime analysis. Table 12 presents the model fitting criteria (AIC, BIC, and HQIC) for Data Set I, which represents the death times of

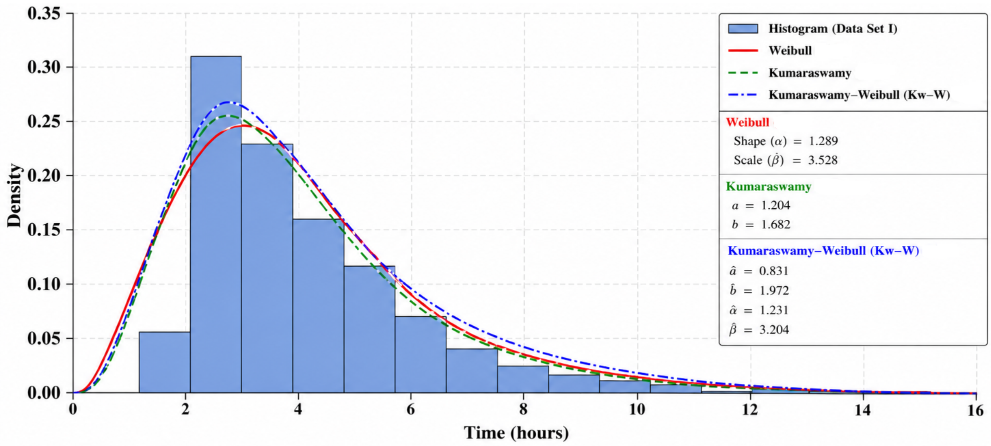

Figure 6. Data Set I Fitting Histogram

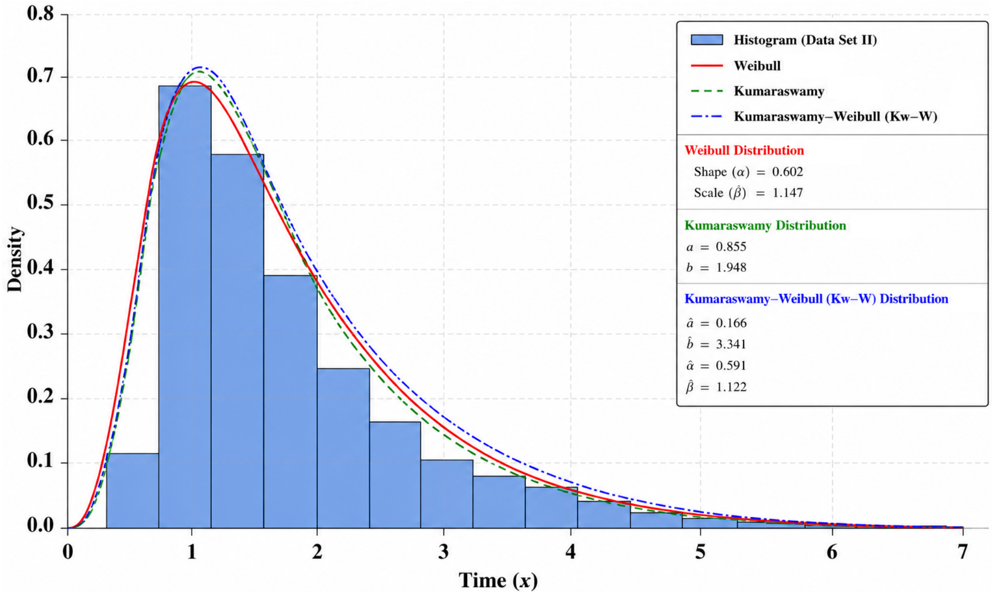

Figure 7. Data Set II Fitting Histogram

Distribution	AIC	BIC	HQIC
Kw–W	**239.12**	244.73	**240.91**
Weibull	240.17	**242.97**	241.06
Kumaraswamy	252.86	255.67	253.76

Table 12. Data Set I Fitting Measures Criteria

tuberculosis germs. The results clearly showed that the Kumaraswamy–Weibull (Kw–W) distribution is the most suitable distribution among the various competing models. Specifically, the Kw–W model is the one with the lowest AIC (239.12), BIC (244.73) and HQIC (240.91) values, indicating superior performance of the model. The following criteria are commonly applied to assess the compromise between goodness-of-fit and model complexity and lower values represent a better model. The Weibull distribution yields slightly higher values, meaning that it is somewhat more successful at modeling the data, but is not as flexible as the Kw–W model. In contrast, the Kumaraswamy distribution shows much larger values for all the criteria which indicates a poorer fit. This is due to its restricted capacity of capturing skewness and tail behaviour of the data. The uniformity of the Kw–W model on all 3 The robustness of criteria. Also, there are only small differences between AIC and HQIC, which suggests stability in model selection. The more expensive the BIC the more complex the model, and the Kw–W model is still the best. These findings confirm that the Kw–W distribution effectively captures the underlying characteristics of biological lifetime data. The results are also consistent with the graphical analysis shown in Figure 6. Overall, Table 12 strongly supports the suitability of the Kw–W model for Data Set I. Table 13 reports the Fitting the

Distribution	AIC	BIC	HQIC
Kw–W	**395.44**	402.20	**397.88**
Weibull	398.38	**401.76**	399.61
Kumaraswamy	462.31	465.69	463.54

Table 13. Data Set II Fitting Measures Criteria

parameters for Data Set II, consisting of fracture times of iron bars under pressure. The best model performance is observed in the Kumaraswamy–Weibull (Kw–W) distribution, like Table 12. The Kw–W model has the lowest AIC (395.44), BIC (402.20) and HQIC (397.88), which means that it is the most accurate model for the data. The Weibull distribution gives slightly higher values, indicating that it is less flexible in the underlying data structure, and thus may not be suitable. The mean of the Kumaraswamy distribution is much higher, however, suggesting a poor fit for this distribution. The result is that simpler models are not able to describe sufficiently the distributional shape and variability of engineering reliability data. The agreement of the three groups (AIC, BIC, and HQIC) makes the model selection more reliable. Although the penalty for any extra parameters, the Kw–W model still outperforms the other models. This demonstrates that it is very important to be flexible when modelling real world data. The results also verify that the Kw–W distribution can be used for moderately symmetric data structures. Additionally, the results are similar to the histogram presented in figure 7 and the curve is very close to the empirical distribution. The Kw–W model is observed to have a superior performance, which shows its viability in reliability analysis. Table 13 substantiates the effectiveness of the proposed distribution for engineering data overall.

8. Conclusions

In this study, a complete model for the simultaneous estimation of the parameters of the Kumaraswamy–Weibull (Kw–W) distribution by classical and Bayesian approaches is provided. The results show that the Kw–W model provides a significant model flexibility for the modeling of complex lifetime data, and that the model is able to capture the various distributional shapes and hazard rate behaviors that are not captured by simpler models. One of the most important results of this work is the extension of the Bayesian estimation to the entire vector of parameters under Stein's loss function, which has been proved to dramatically improve the estimation performance.

The simulation results confirm that the classical estimators (MLE and MoM) can be efficient in small samples but not in larger samples. On the other hand, the Bayesian estimators show clear superiority in moderate and large samples, where they have less bias than their competitors, and they have less mean squared error for all parameters. This improvement is due to the combination of Jeffreys' noninformative prior leading to objectiveness, and to Stein's loss function, which provides an asymmetric penalty system and makes the estimator more stable.

Moreover, empirical applications with real data sets demonstrate the applicability of the proposed approach. The Kw–W distribution is consistently the best performer in comparison to other distributions, based on standard model selection criteria (AIC, BIC and HQIC) and hence successfully represents lifetime phenomena in the real world. These results demonstrate the need to use flexible distributions in conjunction with sophisticated estimation methods.

Finally, it is worth noting that, on the whole, the simultaneous use of the full-parameter Bayesian estimation with Stein's loss function offers a powerful and effective inferential approach to complex lifetime models. The proposed approach has the potential both to enhance estimation accuracy and to provide a reliable tool for practical applications in reliability engineering, survival analysis and beyond, Analysis, related areas. Further development of the model for extensions to censored data, hierarchical Bayesian structures and advances in computation could be explored in future research to further improve applicability.

Appendix A. Derivation of the Fisher Information Matrix for the Kumaraswamy–Weibull Model

In this appendix, we provide the detailed derivation of the observed and expected Fisher information matrices for the Kumaraswamy–Weibull (Kw–W) distribution considered in Section 5. Let

$$\boldsymbol{\theta} = (a, b, \alpha, \beta)^\top,$$

where $a > 0$ and $b > 0$ are the Kumaraswamy shape parameters, and $\alpha > 0$, $\beta > 0$ are the Weibull shape and scale parameters, respectively. Suppose that $X_1, \ldots, X_n$ is a random sample from the Kw–W distribution with density

$$f(x; \boldsymbol{\theta}) = ab\alpha\beta x^{\alpha-1} \exp[-(\beta x)^\alpha] \left(1 - \exp[-(\beta x)^\alpha]\right)^{a-1} \left[1 - (1 - \exp[-(\beta x)^\alpha])^a\right]^{b-1}, \qquad x > 0. \tag{35}$$

For notational convenience, define

$$t_i = (\beta x_i)^\alpha, \qquad G_i = 1 - e^{-t_i}, \qquad H_i = 1 - G_i^a, \qquad i = 1, \ldots, n. \tag{36}$$

Using (36), the log-likelihood function may be written as

$$\ell(\boldsymbol{\theta}) = n \log a + n \log b + n \log \alpha + n \log \beta + (\alpha - 1) \sum_{i=1}^n \log x_i - \sum_{i=1}^n t_i + (a - 1) \sum_{i=1}^n \log G_i + (b - 1) \sum_{i=1}^n \log H_i. \tag{37}$$

The observed information matrix is defined by

$$J(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top}, \tag{38}$$

while the Fisher information matrix is

$$I(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^\top}\right] = E[J(\boldsymbol{\theta})]\,. \tag{39}$$

A.1. Preliminary derivatives

Let

$$s_i = \log(\beta x_i). \tag{40}$$

Then, from $t_i = (\beta x_i)^\alpha$, we obtain

$$\frac{\partial t_i}{\partial \alpha} = t_i s_i, \qquad \frac{\partial t_i}{\partial \beta} = \frac{\alpha}{\beta} t_i, \tag{41}$$

and the corresponding second-order derivatives are

$$\frac{\partial^2 t_i}{\partial \alpha^2} = t_i s_i^2, \qquad \frac{\partial^2 t_i}{\partial \beta^2} = \frac{\alpha(\alpha - 1)}{\beta^2} t_i, \qquad \frac{\partial^2 t_i}{\partial \alpha\, \partial \beta} = \frac{t_i}{\beta}\big(\alpha s_i + 1\big). \tag{42}$$

Next, since $G_i = 1 - e^{-t_i}$, it follows that

$$\frac{\partial G_i}{\partial p} = e^{-t_i} \frac{\partial t_i}{\partial p}, \qquad p \in \{\alpha, \beta\}, \tag{43}$$

and

$$\frac{\partial^2 G_i}{\partial p\,\partial q} = e^{-t_i}\left(\frac{\partial^2 t_i}{\partial p\,\partial q} - \frac{\partial t_i}{\partial p}\frac{\partial t_i}{\partial q}\right), \qquad p,q \in \{\alpha,\beta\}. \tag{44}$$

Therefore,

$$\frac{\partial G_i}{\partial \alpha} = e^{-t_i} t_i s_i, \qquad \frac{\partial G_i}{\partial \beta} = e^{-t_i}\frac{\alpha}{\beta} t_i, \tag{45}$$

$$\frac{\partial^2 G_i}{\partial \alpha^2} = e^{-t_i} t_i s_i^2 (1-t_i), \tag{46}$$

$$\frac{\partial^2 G_i}{\partial \beta^2} = e^{-t_i}\left(\frac{\alpha(\alpha-1)}{\beta^2} t_i - \frac{\alpha^2}{\beta^2} t_i^2\right), \tag{47}$$

and

$$\frac{\partial^2 G_i}{\partial \alpha\,\partial \beta} = e^{-t_i}\left[\frac{t_i}{\beta}(\alpha s_i + 1) - \frac{\alpha}{\beta} t_i^2 s_i\right]. \tag{48}$$

A.2. Derivatives of $\log G_i$

Define

$$A_i = \log G_i. \tag{49}$$

Then

$$\frac{\partial A_i}{\partial p} = \frac{1}{G_i}\frac{\partial G_i}{\partial p}, \qquad p \in \{\alpha,\beta\}, \tag{50}$$

and

$$\frac{\partial^2 A_i}{\partial p\,\partial q} = \frac{1}{G_i}\frac{\partial^2 G_i}{\partial p\,\partial q} - \frac{1}{G_i^2}\frac{\partial G_i}{\partial p}\frac{\partial G_i}{\partial q}, \qquad p,q \in \{\alpha,\beta\}. \tag{51}$$

Substituting from (45)–(48), we obtain

$$\frac{\partial A_i}{\partial \alpha} = \frac{e^{-t_i} t_i s_i}{G_i}, \qquad \frac{\partial A_i}{\partial \beta} = \frac{e^{-t_i}}{G_i}\frac{\alpha}{\beta} t_i, \tag{52}$$

$$\frac{\partial^2 A_i}{\partial \alpha^2} = \frac{e^{-t_i}}{G_i} t_i s_i^2 - \frac{e^{-t_i}}{G_i^2} t_i^2 s_i^2, \tag{53}$$

$$\frac{\partial^2 A_i}{\partial \beta^2} = \frac{e^{-t_i}}{G_i}\frac{\alpha(\alpha-1)}{\beta^2} t_i - \frac{e^{-t_i}}{G_i^2}\frac{\alpha^2}{\beta^2} t_i^2, \tag{54}$$

and

$$\frac{\partial^2 A_i}{\partial \alpha\,\partial \beta} = \frac{e^{-t_i}}{G_i}\frac{t_i}{\beta}(\alpha s_i + 1) - \frac{e^{-t_i}}{G_i^2}\frac{\alpha}{\beta} t_i^2 s_i. \tag{55}$$

A.3. Derivatives of $\log H_i$

Let

$$B_i = \log H_i = \log(1 - G_i^a). \tag{56}$$

The derivative of B_i with respect to a is

$$\frac{\partial B_i}{\partial a} = -\frac{G_i^a \log G_i}{H_i}, \tag{57}$$

and differentiating once more gives

$$\frac{\partial^2 B_i}{\partial a^2} = -\frac{G_i^a (\log G_i)^2}{H_i^2}. \tag{58}$$

For $p \in \{\alpha, \beta\}$, the first derivative with respect to p is

$$\frac{\partial B_i}{\partial p} = -\frac{a G_i^{a-1}}{H_i} \frac{\partial G_i}{\partial p}, \tag{59}$$

and the mixed derivative with respect to a and p becomes

$$\frac{\partial^2 B_i}{\partial a \, \partial p} = -\frac{G_i^{a-1} \dfrac{\partial G_i}{\partial p} \left(1 + a \log G_i - G_i^a\right)}{H_i^2}. \tag{60}$$

Similarly, for $p, q \in \{\alpha, \beta\}$,

$$\frac{\partial^2 B_i}{\partial p \, \partial q} = -\frac{a G_i^{a-1}}{H_i} \frac{\partial^2 G_i}{\partial p \, \partial q} - \frac{a G_i^{a-2} \left(a - 1 + G_i^a\right)}{H_i^2} \frac{\partial G_i}{\partial p} \frac{\partial G_i}{\partial q}. \tag{61}$$

Hence,

$$\frac{\partial B_i}{\partial \alpha} = -\frac{a G_i^{a-1} e^{-t_i} t_i s_i}{H_i}, \qquad \frac{\partial B_i}{\partial \beta} = -\frac{a G_i^{a-1} e^{-t_i}}{H_i} \frac{\alpha}{\beta} t_i, \tag{62}$$

$$\frac{\partial^2 B_i}{\partial a \, \partial \alpha} = -\frac{G_i^{a-1} e^{-t_i} t_i s_i \left(1 + a \log G_i - G_i^a\right)}{H_i^2}, \tag{63}$$

$$\frac{\partial^2 B_i}{\partial a \, \partial \beta} = -\frac{G_i^{a-1} e^{-t_i} \dfrac{\alpha}{\beta} t_i \left(1 + a \log G_i - G_i^a\right)}{H_i^2}, \tag{64}$$

$$\frac{\partial^2 B_i}{\partial \alpha^2} = -\frac{a G_i^{a-1}}{H_i} \frac{\partial^2 G_i}{\partial \alpha^2} - \frac{a G_i^{a-2} (a - 1 + G_i^a)}{H_i^2} \left(\frac{\partial G_i}{\partial \alpha}\right)^2, \tag{65}$$

$$\frac{\partial^2 B_i}{\partial \beta^2} = -\frac{a G_i^{a-1}}{H_i} \frac{\partial^2 G_i}{\partial \beta^2} - \frac{a G_i^{a-2} (a - 1 + G_i^a)}{H_i^2} \left(\frac{\partial G_i}{\partial \beta}\right)^2, \tag{66}$$

and

$$\frac{\partial^2 B_i}{\partial \alpha \, \partial \beta} = -\frac{a G_i^{a-1}}{H_i} \frac{\partial^2 G_i}{\partial \alpha \, \partial \beta} - \frac{a G_i^{a-2} (a - 1 + G_i^a)}{H_i^2} \frac{\partial G_i}{\partial \alpha} \frac{\partial G_i}{\partial \beta}. \tag{67}$$

A.4. Score functions

Differentiating (37), the score components are obtained as

$$\frac{\partial \ell}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \log G_i - (b - 1) \sum_{i=1}^{n} \frac{G_i^a \log G_i}{H_i}, \tag{68}$$

$$\frac{\partial \ell}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} \log H_i, \tag{69}$$

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} t_i s_i + (a - 1) \sum_{i=1}^{n} \frac{e^{-t_i} t_i s_i}{G_i} - (b - 1) \sum_{i=1}^{n} \frac{a G_i^{a-1} e^{-t_i} t_i s_i}{H_i}, \tag{70}$$

and

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^{n} \frac{\alpha}{\beta} t_i + (a-1) \sum_{i=1}^{n} \frac{e^{-t_i}}{G_i} \frac{\alpha}{\beta} t_i - (b-1) \sum_{i=1}^{n} \frac{a G_i^{a-1} e^{-t_i}}{H_i} \frac{\alpha}{\beta} t_i. \tag{71}$$

A.5. Second derivatives of the log-likelihood

The second-order derivatives required for the Hessian matrix are as follows. First,

$$\frac{\partial^2 \ell}{\partial a^2} = -\frac{n}{a^2} - (b-1) \sum_{i=1}^{n} \frac{G_i^a (\log G_i)^2}{H_i^2}, \tag{72}$$

$$\frac{\partial^2 \ell}{\partial a\, \partial b} = -\sum_{i=1}^{n} \frac{G_i^a \log G_i}{H_i}, \tag{73}$$

and

$$\frac{\partial^2 \ell}{\partial b^2} = -\frac{n}{b^2}. \tag{74}$$

Next, the mixed derivatives involving a and (α, β) are

$$\frac{\partial^2 \ell}{\partial a\, \partial \alpha} = \sum_{i=1}^{n} \frac{e^{-t_i} t_i s_i}{G_i} - (b-1) \sum_{i=1}^{n} \frac{G_i^{a-1} e^{-t_i} t_i s_i (1 + a \log G_i - G_i^a)}{H_i^2}, \tag{75}$$

$$\frac{\partial^2 \ell}{\partial a\, \partial \beta} = \sum_{i=1}^{n} \frac{e^{-t_i}}{G_i} \frac{\alpha}{\beta} t_i - (b-1) \sum_{i=1}^{n} \frac{G_i^{a-1} e^{-t_i} \frac{\alpha}{\beta} t_i (1 + a \log G_i - G_i^a)}{H_i^2}. \tag{76}$$

Likewise, the mixed derivatives involving b and (α, β) are

$$\frac{\partial^2 \ell}{\partial b\, \partial \alpha} = -\sum_{i=1}^{n} \frac{a G_i^{a-1} e^{-t_i} t_i s_i}{H_i}, \tag{77}$$

$$\frac{\partial^2 \ell}{\partial b\, \partial \beta} = -\sum_{i=1}^{n} \frac{a G_i^{a-1} e^{-t_i}}{H_i} \frac{\alpha}{\beta} t_i. \tag{78}$$

For the Weibull parameters, we have

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{n}{\alpha^2} - \sum_{i=1}^{n} t_i s_i^2 + (a-1) \sum_{i=1}^{n} \frac{\partial^2 A_i}{\partial \alpha^2} + (b-1) \sum_{i=1}^{n} \frac{\partial^2 B_i}{\partial \alpha^2}, \tag{79}$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = -\frac{n}{\beta^2} - \sum_{i=1}^{n} \frac{\alpha(\alpha-1)}{\beta^2} t_i + (a-1) \sum_{i=1}^{n} \frac{\partial^2 A_i}{\partial \beta^2} + (b-1) \sum_{i=1}^{n} \frac{\partial^2 B_i}{\partial \beta^2}, \tag{80}$$

and

$$\frac{\partial^2 \ell}{\partial \alpha\, \partial \beta} = -\sum_{i=1}^{n} \frac{t_i}{\beta} (\alpha s_i + 1) + (a-1) \sum_{i=1}^{n} \frac{\partial^2 A_i}{\partial \alpha\, \partial \beta} + (b-1) \sum_{i=1}^{n} \frac{\partial^2 B_i}{\partial \alpha\, \partial \beta}. \tag{81}$$

Substituting from (53)–(55) and (65)–(67), these may be expanded explicitly as

$$\frac{\partial^2 \ell}{\partial \alpha^2} = -\frac{n}{\alpha^2} - \sum_{i=1}^{n} t_i s_i^2 + (a-1) \sum_{i=1}^{n} \left(\frac{e^{-t_i}}{G_i} t_i s_i^2 - \frac{e^{-t_i}}{G_i^2} t_i^2 s_i^2 \right) + (b-1) \sum_{i=1}^{n} \frac{\partial^2 B_i}{\partial \alpha^2}, \tag{82}$$

$$\frac{\partial^2 \ell}{\partial \beta^2} = -\frac{n}{\beta^2} - \sum_{i=1}^{n} \frac{\alpha(\alpha-1)}{\beta^2} t_i + (a-1)\sum_{i=1}^{n}\left(\frac{e^{-t_i}}{G_i}\frac{\alpha(\alpha-1)}{\beta^2} t_i - \frac{e^{-t_i}}{G_i^2}\frac{\alpha^2}{\beta^2} t_i^2\right) + (b-1)\sum_{i=1}^{n}\frac{\partial^2 B_i}{\partial \beta^2}, \tag{83}$$

and

$$\frac{\partial^2 \ell}{\partial \alpha\,\partial \beta} = -\sum_{i=1}^{n}\frac{t_i}{\beta}(\alpha s_i + 1) + (a-1)\sum_{i=1}^{n}\left(\frac{e^{-t_i}}{G_i}\frac{t_i}{\beta}(\alpha s_i + 1) - \frac{e^{-t_i}}{G_i^2}\frac{\alpha}{\beta} t_i^2 s_i\right) + (b-1)\sum_{i=1}^{n}\frac{\partial^2 B_i}{\partial \alpha\,\partial \beta}. \tag{84}$$

A.6. Observed information matrix

The observed information matrix $J(\boldsymbol{\theta})$ is obtained by negating the Hessian entries in (72)–(84). Thus,

$$J(\boldsymbol{\theta}) = \begin{pmatrix} J_{aa} & J_{ab} & J_{a\alpha} & J_{a\beta} \\ J_{ab} & J_{bb} & J_{b\alpha} & J_{b\beta} \\ J_{a\alpha} & J_{b\alpha} & J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{a\beta} & J_{b\beta} & J_{\alpha\beta} & J_{\beta\beta} \end{pmatrix}, \tag{85}$$

where

$$J_{aa} = \frac{n}{a^2} + (b-1)\sum_{i=1}^{n}\frac{G_i^a(\log G_i)^2}{H_i^2}, \tag{86}$$

$$J_{ab} = J_{ba} = \sum_{i=1}^{n}\frac{G_i^a \log G_i}{H_i}, \tag{87}$$

$$J_{bb} = \frac{n}{b^2}, \tag{88}$$

$$J_{a\alpha} = J_{\alpha a} = -\sum_{i=1}^{n}\frac{e^{-t_i} t_i s_i}{G_i} + (b-1)\sum_{i=1}^{n}\frac{G_i^{a-1}e^{-t_i}t_i s_i(1 + a\log G_i - G_i^a)}{H_i^2}, \tag{89}$$

$$J_{a\beta} = J_{\beta a} = -\sum_{i=1}^{n}\frac{e^{-t_i}}{G_i}\frac{\alpha}{\beta}t_i + (b-1)\sum_{i=1}^{n}\frac{G_i^{a-1}e^{-t_i}\frac{\alpha}{\beta}t_i(1 + a\log G_i - G_i^a)}{H_i^2}, \tag{90}$$

$$J_{b\alpha} = J_{\alpha b} = \sum_{i=1}^{n}\frac{aG_i^{a-1}e^{-t_i}t_i s_i}{H_i}, \tag{91}$$

$$J_{b\beta} = J_{\beta b} = \sum_{i=1}^{n}\frac{aG_i^{a-1}e^{-t_i}}{H_i}\frac{\alpha}{\beta}t_i, \tag{92}$$

$$J_{\alpha\alpha} = \frac{n}{\alpha^2} + \sum_{i=1}^{n} t_i s_i^2 - (a-1)\sum_{i=1}^{n}\frac{\partial^2 A_i}{\partial \alpha^2} - (b-1)\sum_{i=1}^{n}\frac{\partial^2 B_i}{\partial \alpha^2}, \tag{93}$$

$$J_{\beta\beta} = \frac{n}{\beta^2} + \sum_{i=1}^{n}\frac{\alpha(\alpha-1)}{\beta^2}t_i - (a-1)\sum_{i=1}^{n}\frac{\partial^2 A_i}{\partial \beta^2} - (b-1)\sum_{i=1}^{n}\frac{\partial^2 B_i}{\partial \beta^2}, \tag{94}$$

and

$$J_{\alpha\beta} = J_{\beta\alpha} = \sum_{i=1}^{n}\frac{t_i}{\beta}(\alpha s_i + 1) - (a-1)\sum_{i=1}^{n}\frac{\partial^2 A_i}{\partial \alpha\,\partial \beta} - (b-1)\sum_{i=1}^{n}\frac{\partial^2 B_i}{\partial \alpha\,\partial \beta}. \tag{95}$$

Appendix B: Newton–Raphson Integration

$$E\left[\frac{1}{a} \mid x\right] = \frac{\int \frac{1}{a} L(\theta|x) \sqrt{\det J(\theta)}\, d\theta}{\int L(\theta|x) \sqrt{\det J(\theta)}\, d\theta}, \tag{96}$$

where $\theta = (a, b, \alpha, \beta)$ and $J(\theta)$ is the observed information matrix.

Step 1: Log-Posterior Function

Define the log-posterior function as:

$$\ell^*(\theta) = \log L(\theta|x) + \frac{1}{2} \log \det J(\theta). \tag{97}$$

Step 2: Newton–Raphson Iteration

Let $\theta^{(k)}$ denote the parameter vector at iteration k. The Newton–Raphson update is given by:

$$\theta^{(k+1)} = \theta^{(k)} - H^{-1}(\theta^{(k)})\, \nabla \ell^*(\theta^{(k)}), \tag{98}$$

where $\nabla \ell^*(\theta)$ is the gradient vector and $H(\theta)$ is the Hessian matrix of second derivatives. The iterations are repeated until convergence:

$$\|\theta^{(k+1)} - \theta^{(k)}\| < \varepsilon, \tag{99}$$

for a small tolerance $\varepsilon > 0$.

Step 3: Covariance Approximation

After convergence, let $\hat{\theta}$ denote the maximizer of $\ell^*(\theta)$. The covariance matrix is approximated by:

$$\Sigma = \left[-H(\hat{\theta})\right]^{-1}. \tag{100}$$

Step 4: Laplace Approximation

Using the Laplace approximation, the posterior distribution is approximated by a multivariate normal distribution:

$$\theta \mid x \approx \mathcal{N}(\hat{\theta}, \Sigma). \tag{101}$$

Thus, the expectation in Equation (34) is approximated by:

$$E\left[\frac{1}{a} \mid x\right] \approx \frac{1}{\hat{a}} + \frac{\mathrm{Var}(a)}{\hat{a}^3}, \tag{102}$$

where $\mathrm{Var}(a)$ is obtained from the covariance matrix Σ.

Algorithm Summary

1. Initialize $\theta^{(0)} = (a^{(0)}, b^{(0)}, \alpha^{(0)}, \beta^{(0)})$ using MLE estimates.
2. Compute $\ell^*(\theta^{(k)})$, $\nabla \ell^*(\theta^{(k)})$, and $H(\theta^{(k)})$.
3. Update parameters:
$$\theta^{(k+1)} = \theta^{(k)} - H^{-1}(\theta^{(k)}) \nabla \ell^*(\theta^{(k)}).$$
4. Repeat until convergence.
5. Compute $\Sigma = (-H(\hat{\theta}))^{-1}$.

6. Approximate $E[1/a \mid x]$ using the Laplace approximation.

7. Obtain the Bayes estimator:

$$\hat{a}_{\text{Bayes}} = \frac{1}{E[1/a \mid x]}.$$

REFERENCES

1. Al-Doori, A. M., Salih, A., Mohammed, S. M., & Abdelfattah, A. M. (2025). Regression model for MG Gamma Lindley with application. *Journal of Applied Probability & Statistics*, 20(2).
2. Behairy, S. M., Al-Dayian, G. R., and El-Helbawy, A. A. (2016). The Kumaraswamy-Burr type III distribution: properties and estimation. *British Journal of Mathematics and Computer Science*, 14, 1–21.
3. Cordeiro, G. M., Braga Junior, A. C. R., Demétrio, C. G. B., Ortega, E. M. M., and Pescim, R. R. (2014). Some new results for the Kumaraswamy modified Weibull distribution. *Journal of Statistical Theory and Applications*, 13, 86–104.
4. Cordeiro, G. M., Ortega, E. M. M., and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, 347, 1399–1429.
5. Diaconis, P., and Freedman, D. (1986). On the consistency of Bayes estimates. *The Annals of Statistics*, 14, 1–26.
6. Ergenç, C., and Şenoğlu, B. (2023). Comparison of estimation methods for the Kumaraswamy Weibull distribution. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, 72, 1–21.
7. Ferreira, S. S., and Ferreira, D. (2025). Odd Generalized Exponential Kumaraswamy–Weibull Distribution. *Mathematics*, 13.
8. Gelman, A. (2009). Bayes, Jeffreys, prior distributions and the philosophy of statistics. *Statistical Science*, 24, 176–178.
9. Guven, G., and Senoglu, B. (2023). Statistical inference for the Kumaraswamy Weibull distribution under doubly type II censoring. *International Studies in Science and Mathematics*, 219–233.
10. Hussein, W. J., Salih, A., & Abdullah, M. (2025). A deep neural network approach for estimating time-varying parameters in ordinary differential equation models. *Journal of Applied Probability & Statistics*, 20(2).
11. Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, 186, 453–461.
12. Khan, M. S., King, R., and Hudson, I. L. (2016). Transmuted Kumaraswamy distribution. *Statistics in Transition New Series*, 17, 183–210.
13. Kumaraswamy, P. (1980), A generalized probability density function for double-bounded random processes, *Journal of Hydrology*, 46 , 79–88.
14. Mahmoud, M. A. A. (2012). *On the Generalized Kumaraswamy Distribution*. PhD Dissertation, Cairo University.
15. Mdlongwa, P., Oluyede, B. O., Amey, A. K. A., Fagbamigbe, A. F., and Makubate, B. (2019). Kumaraswamy log-logistic Weibull distribution: model, theory and application to lifetime and survival data. *Heliyon*, 5.
16. Mead, M. E., Afify, A. Z., and Butt, N. S. (2020). The Modified Kumaraswamy Weibull Distribution: Properties and Applications in Reliability and Engineering Sciences. *Pakistan Journal of Statistics and Operation Research*, 16.
17. O'Hagan, A., and Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201–203.
18. Pereira, M. B., Silva, R. B., Zea, L. M., and Cordeiro, G. M. (2012). The Kumaraswamy Pareto distribution. *arXiv preprint arXiv:1204.1389*.
19. Ramos, M. W. A., Marinho, P. R. D., Cordeiro, G. M., da Silva, R. V., and Hamedani, G. G. (2015). The Kumaraswamy-G Poisson family of distributions. *Journal of Statistical Theory and Applications*, 14, 222–239.
20. Salih, A., & Hussein, W. J. (2025). Quasi Lindley Regression Model residual analysis for biomedical data. *Statistics, Optimization & Information Computing*, 14(2), 956–969.
21. Salih, A., Husien, W., & Abdulah, M. (2025). High order statistics from Lambert–Topp–Leone distribution: Statistical properties and applications. *Statistics, Optimization & Information Computing*, 14(3), 1584–1597.
22. Shehata, W. A. M., Abdullah, M. M., and Refaie, M. K. A. (2022). A novel four-parameter log-logistic model: mathematical properties and applications to breaking stress, survival times and leukemia data. *Pakistan Journal of Statistics and Operation Research*, 18.
23. Weibull, W.(1951), A statistical distribution function of wide applicability, *Journal of Applied Mechanics*, 18 , 293–297.
24. Zellner, A. (1994). Bayesian and non-Bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Related Topics V*, 377–390. Springer.
25. Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81, 446–451.
26. Zhang, S., and Xie, L. (2023). Leader learning loss function in neural network classification. *Neurocomputing*, 557, 126735.
27. Zhang, Y. Y. (2019). The Bayesian posterior estimators under six loss functions for unrestricted and restricted parameter spaces. In *Bayesian Inference on Complicated Data*. IntechOpen.