Journal of Statistics and Computer Science
Vol. 2, Issue 2, 2023, pp. 133-146
https://doi.org/10.47509/JSCS.2023.v02i02.05

Modified Extended Kumaraswamy Exponential Distribution: Model and Properties

[1]Arun Kumar Chaudhary, [2]Lal Babu Sah Telee, [3]Murari Karki, [4]Vijay Kumar

[1,2]Department of Statistics, Nepal Commerce campus, Tribhuvan University
[3]Department of Statistics, Saraswati Multiple College, Tribhuvan University
[4]Department of Mathematics and Statistics, DDU Gorakhpur University, Gorakhpur

Abstract

A continuous probability model with three parameters named Modified Extended Kumaraswamy Exponential distribution is established using Extended Kumaraswamy Exponential distribution as base distribution through adding one more scale parameter. Expressions for a number of functions, including the probability density function, skewness and kurtosis, survival function, hazard rate function, and distribution function, have been introduced in this context. Probability density curves and Hazard rate curves displayed. The hazard rate curves exhibit a monotonic increase, followed by a decrease, a period of constancy, constancy followed by an increase, and a J-shaped pattern across various parameter sets. To assess the effectiveness of the developed model, we employed a real dataset on daily COVID-19 death counts in Nepal during the initial wave from January 23, 2020, to December 24, 2020. The model parameters were established using the techniques of least squares, maximum likelihood, and Cramer's-von Mises. To validate the model, we used Corrected Akaike's, Akaike's, Bayesian, and Hannan-Quinn Information Criteria. Furthermore, we utilized Q-Q and P-P plots for validation purposes. The goodness of fit can be assessed using Cramer-von Mises, Anderson-Darling tests and Kolmogorov-Smirnov tests. It's worth noting that all of these analyses and assessments are carried out within the R programming language environment, leveraging its powerful statistical and graphical capabilities. This ensures a systematic and thorough exploration of the model's validity and fitness for the given data.

Keywords*: Corrected Akaike's Information; COVID-19; Goodness of fit; Hazard rate function; New Kw-G family.*

1 Introduction

Research and investigation are essential part of overall the fields. Development and updating is not possible without the research in all aspect. Today, research is being used social sciences. Data analysis, a vital research component, offers diverse techniques and tools found in literature for effective data interpretation. Statistical methods are widely used in data analysis for sampling, estimation, and drawing the inferences about the population parameters etc. Application of statistical tools in research has made easier to researchers getting quick, valid and reliable results. There are different tools available in statistics that have been used in research. One of the most important tools is the use of probability models. Numerous highly beneficial probability models in literature greatly aid data analysis, offering established and valuable tools for researchers. In many cases, the models available does not explain the all the properties of the data accurately. Sometimes we manipulate the data to make it suitable to fit a certain probability distribution. But manipulating data is one of the main causes of misleading results of the research. Researchers have introduced numerous novel probability models to enhance result accuracy. These new probability models try to analyze the data more precisely. In literature we can find different probability models.

New probability models can be formulated by different methods. One of the approaches of formulating novel probability model is by using the family of probability distribution. The Lindley distribution family (Cakmakyapan & Gamze, 2016), Power Lindley-G family of distributions (Hassan & Nassr, 2019), and others, represent various distribution families. Chaudhary et al. (2022) introduced Inverse Exponentiated Odd Lomax Exponential

Article History
Received : 16 October 2023; Revised : 24 November 2023, Accepted : 18 December 2023; Published : 30 December 2023

To cite this paper
Arun Kumar Chaudhary, Lal Babu Sah Telee, Murari Karki & Vijay Kumar (2023). Modified Extended Kumaraswamy Exponential Distribution: Model and Properties. *Journal of Statistics and Computer Science.* 2(2), 133-146.

distribution using T-X family (Aljarrah et al., 2014) of distribution. Chaudhary et al. (2023) have also created the inverse exponential power distribution using the inverse transformation technique. Other method of formulating the new model of probability distribution is by modifying the existing probability model. Within the realm of literature, numerous adaptations of the Weibull distribution can be found, and one such adaptation is the Weibull distribution with two-parameter, which is articulated as follows:

$$\overline{F}(x,\alpha,\beta) = \exp[-(\alpha, x)]^{\beta} \quad (1)$$

The exponentiated Weibull distribution, presented by (Mudholkar & Srivastava,1993), emerges by modifying the Weibull distribution to yield various bathtub hazard rate functions. Modified Weibull (Lai et al., 2003) distribution is modified form of Weibull distribution as

$$\overline{F}(x) = \exp[ax^{b}.\exp(\lambda x)] \quad (2)$$

Telee and Kumar (2023) have new probability model termed as developed modified the Generalized Exponential distribution by adding on shape parameter in Generalized Exponential distribution introduced by (Gupta & Kundu, 2001). In this article, we have modified the Extended Kumaraswamy Exponential (EKwE) Distribution, originally proposed by (Chaudhary et al., 2023) to create a new probability model termed as the Modified Extended Kumaraswamy Exponential (MEKwE) Distribution. The EKwE distribution was initially introduced using the New Kumaraswamy Generalized Family of Distributions (NKwG) framework, as outlined by (Tahir et al., 2020). The EKwE model is characterized by two parameters applied to a single continuous variable. The cumulative distribution function (CDF) for the New Kw-G family (NKwG) is established as follows:

$$F(x;\lambda,\beta,\theta) = 1-\left(1-\left(1-\bar{G}(x;\eta)^{G(x;\eta)}\right)^{\beta}\right)^{\theta} \;;(\eta,\beta,\theta) > 0, x \geq 0 \quad (3)$$

The unique situation involving the New Kw-G family (NKwG) occurs when we assign the parameter θ a value of 1, like this:

$$F(x;\eta,\beta) = \left\{1-\bar{G}(x;\eta)^{G(x;\eta)}\right\}^{\beta} \;;(\eta,\beta) > 0, x \geq 0 \quad (4)$$

The exponential distribution serves as the foundational function in eq. (4) to establish the EKwE model and its CDF is

$$G(x;\lambda) = 1-e^{-\lambda x} \text{ where, } \bar{G}(x;\lambda) = e^{-\lambda x}; x \geq 0, \lambda > 0 \quad (5)$$

CDF of the EKwE model is defined in eq. (6)

$$F(x;\lambda,\beta) = \left(1-e^{-\lambda x(1-e^{-\lambda x})}\right)^{\beta} \;; x \geq 0, (\lambda,\beta) > 0 \quad (6)$$

To introduce the proposed model, we have modified the model (6) by adding one more scale parameter α and the resulting CDF of the MEKwE is provided as

$$F(x;\alpha,\ \lambda,\ \beta) = \left(1-\exp\left(-\lambda.xe^{\alpha x}(1-e^{-\lambda x})\right)\right)^{\beta} \;; x \geq 0, (\alpha,\lambda,\beta) > 0 \quad (7)$$

2 Model Analysis

The suggested model MEKwE is described by eq. (7) as its PDF is

$$f(x)=\beta\lambda e^{\alpha x}e^{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))}\{\lambda x\exp(-\lambda x)+(1+\alpha x)(1-\exp(-\lambda x))\}\{1-\exp(\lambda x e^{\alpha x}(1-\exp(-\lambda x)))\}^{\beta-1};x\ge 0,(\alpha,\lambda,\beta)>0 \quad (8)$$

2.1 Survival function

Model's survival function is given by expression (9) as

$$S(x)=1-\left(1-\exp\left[-\lambda x e^{\alpha x}(1-\exp\{-\lambda x\})\right]\right)^{\beta}\ ;\ x\ge 0,\ (\alpha,\lambda,\beta)>0 \quad (9)$$

2.2 Failure rate function

The expression (10) provides the hazard rate function for the suggested model which is crucial for assessing the instantaneous failure probability at any given time, offering valuable insights into the model's reliability and potential risks over time.

$$h(x)=\beta\lambda\exp(\alpha x)\left\{\lambda x\exp(-\lambda x)+(1+\alpha x)(1-\exp(-\lambda x))\right\}e\{-\lambda x e^{\alpha x}[1-\exp(-\lambda x)]\}$$
$$\left\{1-\left(1-e^{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))}\right)^{\beta}\right\}^{-1}\left\{1-e^{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))}\right\}^{\beta-1}\ ;x>0 \quad (10)$$

Figure 1 depicts probability density curves and hazard rate curves for various parameter sets. Different shapes of the curves show that the proposed model may be flexible corresponding to the different type of real data sets

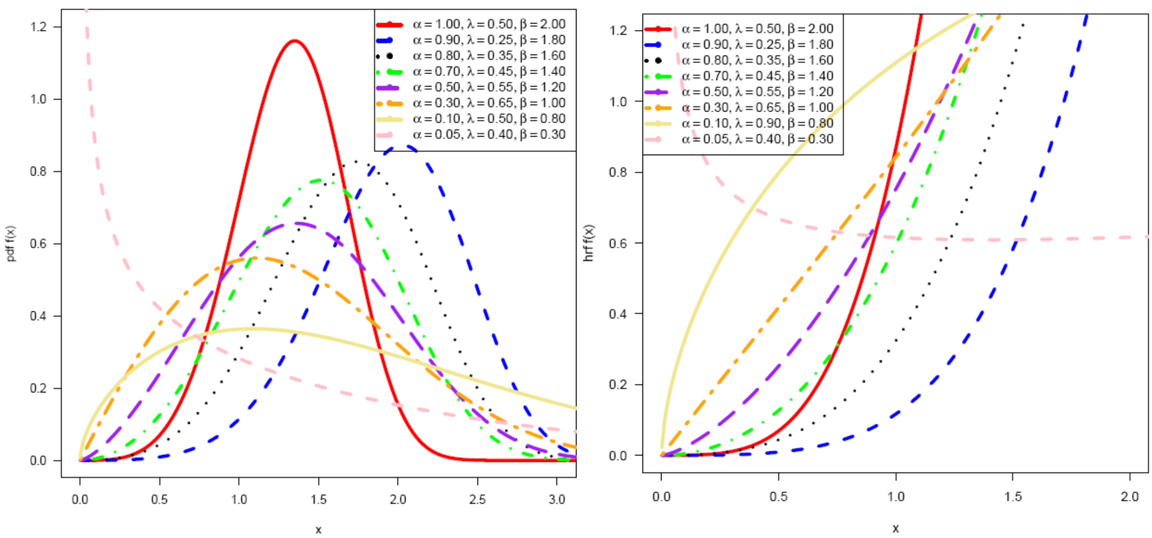

Figure 1: Density curve (on the left) and hazard curve (on the right) for MEKwE

The hazard rate curves exhibit a monotonic increase, followed by a decrease, a period of constancy, constancy followed by an increase, and a J-shaped pattern across various parameter sets. Understanding these diverse hazard rate behaviors is crucial for accurate risk assessment and decision-making in different contexts.

2.3 Reversed hazard rate function

The suggested model's reversed failure rate function is articulated as follows:

$$h_{rev}(x)=\beta\lambda\exp(\alpha x)e^{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))}\left\{\lambda x\exp(-\lambda x)+\{1-\exp(-\lambda x)\}(1+\alpha x)\right\}$$
$$\left(1-\exp\{-\lambda x e^{\alpha x}(1-\exp(-\lambda x)\}\right)^{-\beta}\left\{1-\exp\{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))\}\right\}^{\beta-1};x>0,(\alpha,\lambda,\beta)>0 \quad (11)$$

2.4 Cumulative failure rate function

Given is the cumulative failure rate function for the suggested model which illustrates the accumulated failure rate of the model over time. It provides a representation of how the failure rate increases as time progresses.

$$H(x) = -\ln(S(x)) = -\ln\left[1-\left(1-e^{-\lambda x e^{\alpha x}(1-e^{-\lambda x})}\right)^{\beta}\right]; x \geq 0, (\alpha, \lambda, \beta) > 0 \quad (12)$$

2.5 Quantile function

The quantile function serves as an alternative to the distribution function and can be employed to calculate various descriptive measures of the model. The model's quantile function is presented as

$$\lambda x e^{\alpha x}\left(1-e^{-\lambda x}\right)+\ln\left(1-p^{1/\beta}\right)=0; \quad 0 \leq p \leq 1 \quad (13)$$

2.6 Asymptotic properties

Through confirming that $\underset{x\to 0}{\text{limit}} f(x) = \underset{x\to\infty}{\text{limit}} f(x)$,the density function's asymptotic behavior can be explored. If the model conforms to asymptotic properties, a modal value will be present. This can be verified by taking limits at the endpoints.

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} \beta\lambda e^{\alpha x} e^{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))} \left\{\lambda x \exp(-\lambda x)+(1+\alpha x)(1-\exp(-\lambda x))\right\}\left\{1-\exp(\lambda x e^{\alpha x}(1-\exp(-\lambda x)))\right\}^{\beta-1} = 0$$

$$\lim_{x\to \infty} f(x) = \lim_{x\to \infty} \beta\lambda e^{\alpha x} e^{-\lambda x e^{\alpha x}(1-\exp(-\lambda x))} \left\{\lambda x \exp(-\lambda x)+(1+\alpha x)(1-\exp(-\lambda x))\right\}\left\{1-\exp(\lambda x e^{\alpha x}(1-\exp(-\lambda x)))\right\}^{\beta-1} = 0$$

Ensuring that the density function maintains its characteristics under these conditions is crucial for confirming the reliability of the model. The existence of a modal value, supported by the fulfillment of asymptotic properties and verified through endpoint limits, enhances the robustness of the analysis and underscores the validity of the model's representation.

In this context, $\underset{x\to 0}{\text{limit}} f(x) = \underset{x\to\infty}{\text{limit}} f(x)$. Consequently, the proposed model exhibits a modal value. To find the mode of the distribution, solve the equation $f^{'}(x)=0$ under the condition that $f''(x)<0$. This process helps identify the critical points where the slope is zero and ensures that these points correspond to a maximum on the curve, indicating the mode.

2.7 Skewness and kurtosis

In this research, Al-saiary et al. (2019) employed Bowley's skewness coefficient using quantiles as

$$\text{SK}(\text{B}) = \{\text{Q}(1/4)+\text{Q}(3/4) - 2\text{Q}(1/2)\}\{\text{Q}(3/4)-\text{Q}(1/4)\}^{-1}$$

The formula for determining Octiles Kurtosis coefficients, as outlined by (Moors, 1988), can be applied as

$$\text{K}_\text{u} = \frac{\text{Q}(0.375) - \text{Q}(0.625) + \text{Q}(0.875) - \text{Q}(0.125)}{\text{Q}(0.75) - \text{Q}(0.25)}$$

3 Parameter estimation

The recommended model's parameters are established using Least Squares, Maximum Likelihood, and Cramer's-von Mises approaches. These comprehensive approaches enhance the accuracy and reliability of the model, ensuring that it effectively captures the underlying patterns and relationships in the data.

3.1 Maximum likelihood estimation

Maximum likelihood function maximizes the likelihood function to estimate the parameters. Suppose a random sample $\underline{x} = (x_1, \ldots, x_n)$ having size 'n' elements, drawn from MEKwE. In this instance, the log likelihood function can be formulated as follow

$$\ell(x;\alpha,\beta,\lambda) = n\ln\beta + n\ln\lambda + \alpha\sum_{i=1}^{n} x + (\beta-1)\sum_{i=1}^{n}\ln\{1-e^{-\lambda x e^{\alpha x}(1-e^{-\lambda x})}\} - \lambda\sum_{i=1}^{n} x e^{\alpha x(1-e^{-\lambda x})} + \sum_{i=1}^{n}\ln\{\lambda x e^{-\lambda x} + (1+\alpha x)(1-e^{\lambda x})\} \tag{14}$$

With respect to the parameters $\alpha, \lambda,$ and β ,we calculate the partial derivatives of equation (14). Next, by setting the first-order partial derivatives to zero and solving for the values, we obtain the estimated parameters. When differentiating equation (14) with respect to $\alpha, \lambda,$ and β , the following expressions are derived:

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} x_i + (\beta-1)\sum_{i=1}^{n}\left[\frac{e^{-\lambda x_i e^{\alpha x_i}(1-e^{-\lambda x_i})}\lambda x^2 e^{\alpha x}(1-e^{-\lambda x_i})\}}{1-e^{-\lambda x e^{\alpha x}(1-e^{-\lambda x_i})}}\right] - \lambda\sum_{i=1}^{n} x_i^2 e^{\alpha x_i(1-e^{-\lambda x_i})}(1-e^{-\lambda x_i}) + \sum_{i=1}^{n}\left[\frac{x(1-e^{-\lambda x})}{\lambda x e^{-\lambda x_i} + (1+\alpha x)(1-e^{-\lambda x_i})}\right]$$

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n}\ln\left\{1-e^{\lambda x_i e^{\alpha x_i}(1-e^{-\lambda x_i})}\right\} \quad \text{and}$$

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} + (\beta-1)\sum_{i=1}^{n} e^{-\lambda x_i e^{\alpha x_i}(1-e^{-\lambda x_i})} x_i e^{\alpha x_i}\{(1-e^{-\lambda x_i}) + \lambda^2 e^{-\lambda x_i}\}\left\{1-e^{-\lambda x_i e^{\alpha x_i}(1-e^{-\lambda x_i})}\right\}^{-1} - \sum_{i=1}^{n} x_i e^{\alpha x_i}(1-e^{-\lambda x_i}) - \lambda\left\{\sum_{i=1}^{n} x_i e^{\alpha x_i}(1-e^{-\lambda x_i}) - \alpha x_i \lambda(1-e^{-\lambda x_i})\right\} + \sum_{i=1}^{n} e^{-\lambda x_i}\{x_i - \lambda^2 x_i + (1+\alpha x_i)\lambda\}\left\{\lambda x_i e^{-\lambda x_i} + (1+\alpha x_i)(1-e^{-\lambda x_i})\right\}$$

Finding an analytical solution for these first-order derivatives is not feasible. Therefore, R programming language is employed for their solution.

If we define $\hat{\Theta} = (\hat{\alpha}, \hat{\lambda}, \hat{\beta})$ and $\Theta = (\alpha, \lambda, \beta)$ with Θ being the parameter vector and $\hat{\Theta}$ being the estimated constants, then the resultant asymptotic normality can be expressed as $\left(\hat{\Theta} - \Theta\right) \rightarrow N_3\left[0, \left(I(\Theta)\right)^{-1}\right]$. The expression for the Fisher's information matrix, symbolized as $I(\Theta)$, can be formulated in the following manner:

$$I(\Theta) = -\begin{pmatrix} E\left(\frac{\partial^2 l}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 l}{\partial \alpha \partial \lambda}\right) & E\left(\frac{\partial^2 l}{\partial \alpha \partial \beta}\right) \\ E\left(\frac{\partial^2 l}{\partial \lambda . \partial \alpha}\right) & E\left(\frac{\partial^2 l}{\partial \lambda^2}\right) & E\left(\frac{\partial^2 l}{\partial \lambda \partial \beta}\right) \\ E\left(\frac{\partial^2 l}{\partial \beta . \partial \alpha}\right) & E\left(\frac{\partial^2 l}{\partial \beta . \partial \lambda}\right) & E\left(\frac{\partial^2 l}{\partial \beta^2}\right) \end{pmatrix}$$

The Maximum Likelihood Estimator's asymptotic variance, denoted as $\left(I(\Theta)\right)^{-1}$,is of limited value since Θ cannot be determined. Instead, we can utilize $O(\hat{\Theta})$ to symbolize the observed Fisher information matrix. We can derive the Hessian matrix ∇^2 by estimating $O(\hat{\Theta})$ based on $I(\Theta)$ in the following manner.

$$O(\hat{\Theta}) = -\begin{pmatrix} \left(\frac{\partial^2 l}{\partial \alpha^2}\right) & \left(\frac{\partial^2 l}{\partial \alpha \partial \lambda}\right) & \left(\frac{\partial^2 l}{\partial \alpha \partial \beta}\right) \\ \left(\frac{\partial^2 l}{\partial \lambda . \partial \alpha}\right) & \left(\frac{\partial^2 l}{\partial \lambda^2}\right) & \left(\frac{\partial^2 l}{\partial \lambda \partial \beta}\right) \\ \left(\frac{\partial^2 l}{\partial \beta . \partial \alpha}\right) & \left(\frac{\partial^2 l}{\partial \beta . \partial \lambda}\right) & \left(\frac{\partial^2 l}{\partial \beta^2}\right) \end{pmatrix} = -\nabla^2 (\Theta)\Big|_{(\Theta=\hat{\Theta})}$$

The variance covariance matrix is,

$$\left[-\nabla^2 (\underline{\Theta})\Big|_{(\Theta=\hat{\Theta})}\right]^{-1} = \begin{pmatrix} \mathrm{var}(\hat{\alpha}) & \mathrm{cov}(\hat{\alpha},\hat{\lambda}) & \mathrm{cov}(\hat{\alpha},\hat{\beta}) \\ \mathrm{cov}(\hat{\lambda},\hat{\alpha}) & \mathrm{var}(\hat{\lambda}) & \mathrm{cov}(\hat{\lambda},\hat{\beta}) \\ \mathrm{cov}(\hat{\beta},\hat{\alpha}) & \mathrm{cov}(\hat{\beta},\hat{\lambda}) & \mathrm{var}(\hat{\beta}) \end{pmatrix}$$

In this context, the confidence intervals for α, λ, and β at the 100(1-γ) % level are $\hat{\alpha} \pm Z_{\cdot\gamma/2}\sqrt{Var(\hat{\alpha})}$, $\hat{\lambda} \pm Z_{\gamma/2}\sqrt{Var(\hat{\lambda})}$, and $\hat{\beta} \pm Z_{\gamma/2}\sqrt{Var(\hat{\beta})}$, respectively.

3.2 Estimating through the Least-Squares Method (LSE)

We arrange a series of ordered random variables as $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ and create a random sample $\{X_1, X_2, \ldots, X_n\}$ from the probability model described by the function F(.).We then create a function $K(x_{(i)};\alpha,\lambda,\beta)$ by utilizing the cumulative distribution function (CDF) of ordered statistics, denoted as $F(X_{(i)})$, as described in equation (15).

$$K(x;\alpha,\lambda,\beta) = \sum_{i=1}^{n}\left[F(X_{(i)}) - \frac{i}{n+1}\right]^2$$

$$K(x_{(i)};\alpha,\lambda,\beta) = \sum_{i=1}^{n}[\{1-e^{-\lambda x e^{\alpha x(i)}(1-e^{-\lambda x(i)})}\}^{\beta} - \frac{i}{n+1}]^2; \quad x_{(i)} \geq 0, (\alpha,\lambda,\beta) \geq 0 \tag{15}$$

To ascertain the MEKwE model's parameters, we can accomplish this by minimizing function (15). This process involves calculating the partial derivatives of function $K(x_{(i)};\alpha,\lambda,\beta)$ with respect to its parameters up to the second order. The first order derivatives are

$$\frac{\partial A}{\partial \alpha} = -2\alpha\beta\lambda e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})} x_{(i)} e^{\alpha x_{(i)}} (1-e^{-\lambda x_{(i)}}) \left\{1-e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta-1} \sum_{i=1}^{n}\left[\left\{1-e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta} - \frac{i}{n+1}\right]$$

$$\frac{\partial K}{\partial \beta} = 2\sum_{i=1}^{n}\left[\left\{1-e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta} - \frac{i}{n+1}\right]\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta} \ln\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}; \text{ and}$$

$$\frac{\partial K}{\partial \lambda} = 2\beta\sum_{i=1}^{n} x_{(i)} e^{\alpha x_{(i)}}\left[1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\}^{\beta} - \frac{i}{n+1}\right]\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta-1} e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\left(1+\lambda^2 e^{-\lambda x_{(i)}} - e^{-\lambda x_{(i)}}\right)$$

Alternatively, we can obtain the parameters by minimizing function E through weighted least squares estimation (LSE).

$$E(X;\ \alpha,\ \lambda,\ \beta) = \sum_{i=1}^{n} w_i\left[\left\{1-e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})\beta}\right\} - \frac{i}{n+1}\right]^2 \quad \text{Where, } w_i = \frac{1}{Var(X_{(i)})} = \frac{(n+2)(n+1)^2}{i(n-i+1)}$$

We can obtain expression (17) by utilizing the CDF for order statistics and the weight w_i from the preceding equation, as demonstrated below.

$$E(X;\alpha,\lambda,\beta) = \sum_{i=1}^{n}\frac{(n+2)(n+1)^2}{i(n-i+1)}\left[\left\{1-e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta} - \frac{i}{n+1}\right]^2 \tag{17}$$

3.3 Cramer-Von-Mises (CVM) method

Minimizing function (18) estimates parameters α, λ, and β using this method for estimation.

$$Z(X;\alpha,\lambda,\beta) = \frac{1}{12n} + \sum_{i=1}^{n}\left[\left\{1-e^{-\lambda x e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta} - \frac{2i-1}{2n}\right]^2 \tag{18}$$

Calculating the partial derivatives of function Z, with respect to α, λ, and β, we find its partial derivatives up to second order and solving $\frac{\partial Z}{\partial \alpha} = 0, \frac{\partial Z}{\partial \lambda} = 0, \text{and } \frac{\partial Z}{\partial \beta} = 0$, this process can be allowed to obtain CVM estimates. First order partial derivatives are;

$$\frac{\partial Z}{\partial \alpha} = -2\alpha\beta\lambda\sum_{i=1}^{n}\left[\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\}^{\beta} - \frac{2i-1}{2n}\right]^2\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta-1} e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})} x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}}).$$

$$\frac{\partial Z}{\partial \beta} = 2\sum_{i=1}^{n}\left[\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\}^{\beta} - \frac{2i-1}{2n}\right]\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta} \ln\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}; \text{ and}$$

$$\frac{\partial Z}{\partial \lambda} = 2\beta\sum_{i=1}^{n}\left[\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\}^{\beta} - \frac{2i-1}{2n}\right] e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\left\{1-e^{-\lambda x_{(i)} e^{\alpha x_{(i)}}(1-e^{-\lambda x_{(i)}})}\right\}^{\beta-1} x_{(i)} e^{\alpha x_{(i)}}(1+\lambda^2 e^{-\lambda x_{(i)}} - e^{-\lambda x_{(i)}})$$

4 Real data analysis

To test the applicability of the data, we have taken a real data set. The dataset comprises daily COVID-19 death counts in Nepal during the initial wave from January 23, 2020, to December 24, 2020, as reported by the (Worldometer, 2023). During first wave, there were 1808 deaths at the end of 24 December 2020. In this study we have used the daily deaths more than 1 death. The sample contains the data of 153 days with total deaths of 1777.

2, 2, 2, 2, 2, 2, 2, 3, 2, 3, 3, 4, 2, 5, 5, 3, 2, 4, 4, 8, 4, 4, 3, 2, 3, 7, 6, 6, 11 , 9, 3, 8, 7, 11, 8, 12, 12, 14, 7, 11, 12, 6, 14, 9, 9, 11, 6, 6, 5, 5, 14, 9, 15, 11, 8, 4, 7, 11, 10, 16, 2,7, 17, 6, 8, 10, 4, 10, 7, 11, 11, 8, 7, 19, 9, 15, 12, 10, 14, 22, 9, 18, 12, 19, 21, 12, 12, 18, 8, 26, 21, 17, 13, 5, 15, 14, 11, 17, 16, 17, 23, 24, 20, 30, 18, 18, 17, 21, 18, 22, 26, 15, 13, 13, 6, 9, 17, 12, 17, 22, 7, 16, 16, 24, 28, 23, 23, 19, 25, 29, 21, 9, 13, 16, 10, 17, 20, 23, 14, 12, 11, 15 ,9, 18, 14, 13, 6, 16, 12, 11, 7, 3, 5, 5.

Exploratory data analysis: It provides insights into the structure and characteristics of the data. For getting some information about nature of the data and hazard rate curve, we have plotted boxplot and TTT plot of the dada considered. The empirical expression for the TTT plot is provided below.

$$T\left(\frac{r}{n}\right)=(n-r)y_{i:n}\left(\sum_{i=1}^{n} y_{(i:n)}\right)^{-1}+\sum_{i=1}^{n} y_{(i:n)}$$

In this scenario, $y_{(i:n)}\,(i=1,2,...,r)$ represents sample order statistics, where 'r' varies from 1 to 'n', the 'total time on test' graphic can take on various forms. Aarset (1987) demonstrated that when the curve tends towards a straight diagonal function, a constant failure rate is applicable. If the curve is convex or concave, the failure rate function is consistently increasing or decreasing, respectively, and is considered appropriate. In cases where the failure rate function exhibits both convex and concave characteristics, the U-shaped format is suitable; otherwise, a unimodal failure rate function is more appropriate. The TTT data plot's concave form indicates an increasing hazard rate for the proposed model. This information aids in selecting an appropriate failure rate function, ensuring a more accurate representation of the underlying dynamics of the system under consideration.

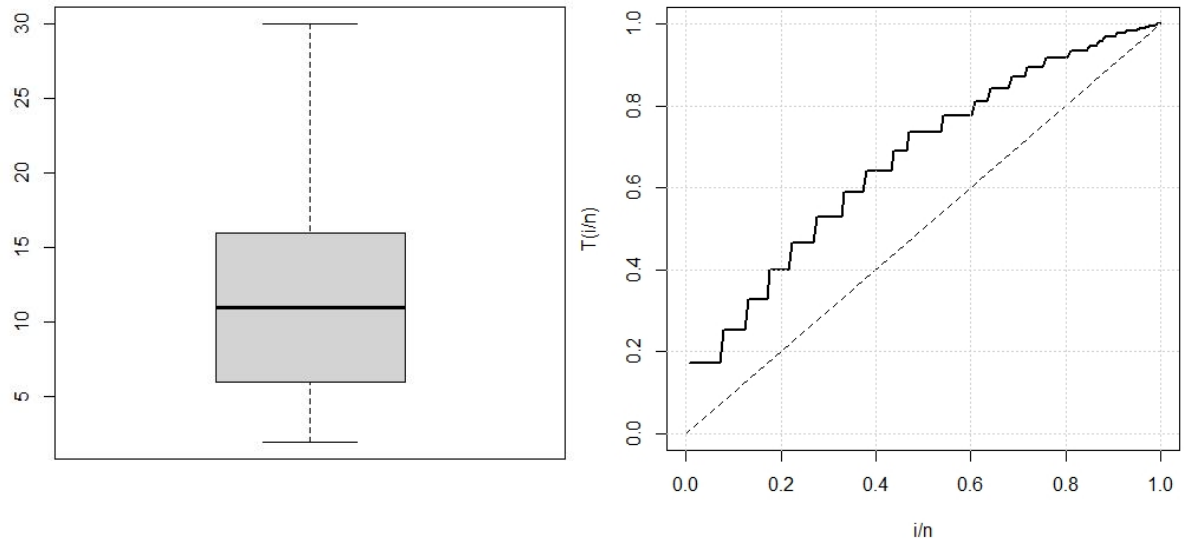

Fig 2. Boxplot (on the left) and TTT plot (on the right)

Table 1. Summary Statistics

Minimum	Q_1	Median	Mean	Q_3	S.D.	Skewness	Kurtosis	Maximum
2	6	11	11.616	16	6.7591	0.508327	2.547717	30

The above findings indicate that the data exhibits positive skewness and deviates from a normal distribution in terms of its shape. Understanding these characteristics is crucial for appropriate data analysis and interpretation.

4.1 Estimated Parameters:

Table 2 presents model parameters estimated through the MLE, LSE and CVM methods. Parameters are estimated using optim () function of R language programming. Table also contains the standard errors.

Table 2: Parameters' Estimated Values and Their Standard Errors

Parameters	MLE	LSE	CVM
Alpha	0.0319(0.0118)	0.0353 (0.1554)	0.0356(0.1555)
Lambda	0.0681(0.0148)	0.0595(0.1403)	0.0599(0.1412)
Beta	0.7986(0.1388)	0.6799(1.1367)	0.6891(1.1569)

We've assessed the model's validity by creating P-P and Q-Q plots to examine its performance and accuracy. The plots demonstrate that the model accurately captures the data.

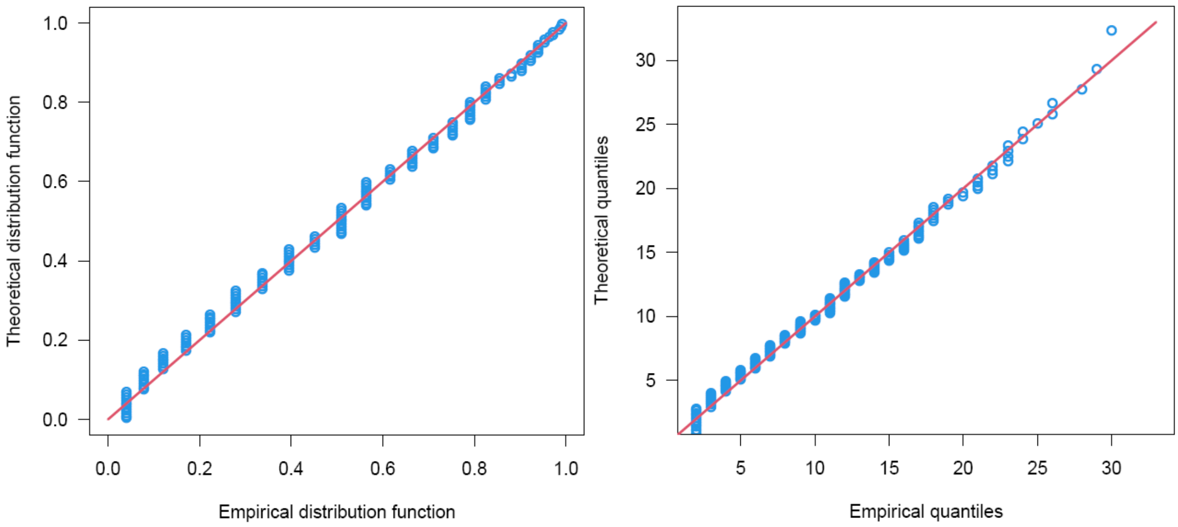

Figure 2: P-P plot (Left) and Q-Q plot (Right) of the model

We created histograms and display fitted density curves on the left panel of Figure 3 to visually assess the goodness of fit. The Figure 3's right panel showcases both the empirical CDF and the fitted CDF of the model.

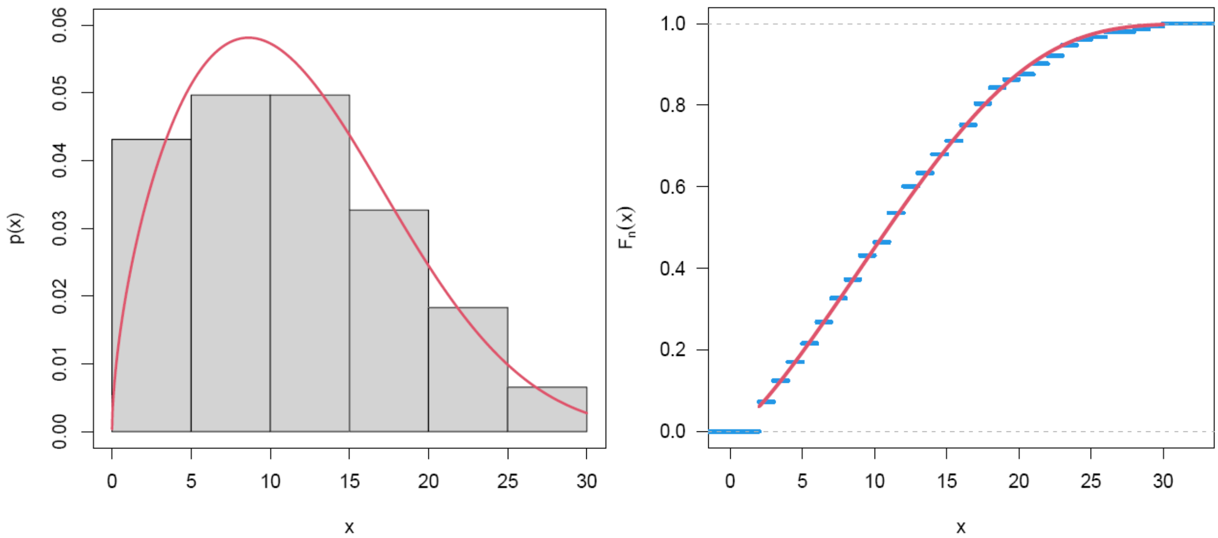

Figure 3: Histogram versus Pdf (Left panel) and ECDF versus Fitted CDF (Right panel)

Table 3 compares the information criteria values corresponding to different methods of estimation used in study. In essence, information criteria serve as tools for selecting models by comparing their fit to the same data. It's important to note that the models under comparison don't have to be nested. These criteria are measures of model fit based on likelihood, incorporating a penalty for complexity, specifically in terms of the number of parameters.

Different information criteria vary in the form of the penalty they apply and may favor different models.

Irrespective of the specific information criterion used, when comparing values across multiple models, smaller criterion values indicate a better and more parsimonious fit. In this context, we considered Akaike's, Bayesian, Hannan-Quinn, and Corrected Akaike's Information Criteria. Among these, the Maximum Likelihood Estimation (MLE) method yielded the lowest information criterion values, suggesting a superior fit compared to other methods. Therefore, based on the information criteria analysis, the MLE method emerges as a robust choice for modeling the given data.

Table 3: Information criteria corresponding to methods of estimation

Methods	LL	AIC	BIC	CAIC	HQIC
MLE	-496.5925	999.185	1008.276	999.340	1002.878
LSE	-497.2673	1000.535	1009.626	1000.696	1004.228
CVM	-497.1079	1000.216	1009.307	1000.377	1003.909

The minimum log-likelihood (LL) and information criteria values linked to Maximum Likelihood Estimation (MLE) suggest that MLE offers a superior fit to the data in comparison to the Cramér-von Mises (CVM) and Least Squares Error (LSE) estimation approaches. We used the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests to evaluate how well the data fits. All the data analysis was conducted using the R programming language. The test statistics and corresponding p-values for various estimation techniques are presented in Table 4, revealing that the CVM method exhibits lower test statistics values with higher p-values for A^2 and W. Additionally, the MLE method demonstrates lower test statistics values with higher p-values for the Kolmogorov-Smirnov test. These observations enhance our comprehension of how various estimation methods perform in accurately capturing the inherent characteristics of the dataset.

Table 4: Test statistics and p values corresponding to methods of estimation

Methods	KS(p-values)	A^2(p- values)	W(p-values)
MLE	0.0488(0.8586)	0.0597(0.8167)	0.5208(0.7254)
LSE	0.0555(0.7344)	0.0413(0.9267)	0.4407(0.8074)
CVM	0.0539(0.7654)	0.04087(0.9292)	0.4268(0.8215)

4.2 Comparative study of the model

To demonstrate the suitability of our proposed model, we conducted a thorough comparison with eight other prominent probability models found in the existing literature. These selected models are as follows:

1. Exponentiated Generalized Odd Lomax Exponential (EGOLE) distribution (Telee et al., 2023)
2. Exponentiated Odd Lomax Exponential (EOLE) distribution (Dhungana & Kumar, 2022)
3. Lindley Inverse Weibull (LIW) distribution (Joshi & Kumar, 2020b)
4. Half Logistic Nadarajah Haghighi (HLNHE) distribution (Joshi & Kumar, 2020a)
5. Lomax Exponentiated Weibull (LEW) distribution (Ansari and Nofal, 2020)
6. Odd Lomax Exponential (OLE) distribution (Ogunsanya et al., 2019)
7. Exponentiated Generalized Inverted Exponential (EGIE) distribution (Oguntunde et al., 2014)
8. Generalized Inverted Generalized Exponential (GIGE) distribution (Oguntunde et al., 2015)

For our comparative analysis, we estimated the parameters and standard errors of these models using the same dataset. The values of the estimated parameters for the recommended model are demonstrated in Table 5 for reference.

Table 5: The estimated parameter values and their corresponding standard errors using MLE for both the proposed and competing models

Models	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\delta}$	$\hat{\theta}$	$\hat{\gamma}$
MEKwE	0.0319(0.0118)	0.7986(0.1388)	0.0681(0.0148)			
EGOLE	54.800(156.0923)	38.711(106.9083)	0.053(0.0192)		1.682(0.3362)	
EOLE	0.0266(0.0096)		28.6034(4.9965)	7.5254(5.2967)	1.5457(0.1237)	
LIW	8.6377(0.8695)	0.7461(0.1045)			4.1896(1.4514)	
HLNHE	0,04697(-)	1.6400(0.2991)	1.33342(-)			
LEW	6.8995(0.00049)	1.05636(-)			0.0860(0.0070)	
GIGE	2.3108(0.3062)		4.8503(-)			2.4017(-)
EGIE	7.6227(-)	0.2307(0.0186)	44.14876(-)			
OLE	3.6466(0.9066)	14.6636(4.1751)			0.1264(0.0187)	

We evaluated the model's goodness of fit to the data using multiple information criteria, as outlined in Table 6. These statistical measures, namely Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), Hannan Quinn Information Criterion (HQIC), and Corrected Akaike Information Criterion (CAIC), are employed for model selection and evaluation across diverse fields. The proposed model exhibits the lowest log likelihood value, surpassing all others except for EGOLE. In terms of information criteria values (AIC, BIC, CAIC, and HQIC), the proposed model MEKwE consistently records the lowest values. These lower values across various metrics imply a superior goodness of fit for the proposed model. This reinforces its appropriateness for accurately representing the data when compared to alternative models, including EGOLE. In summary, the proposed model demonstrates a superior fit to the dataset compared to competing models.

Table 6: Log likelihood and values of information criteria

Models	LL	AIC	BIC	CAIC	HQIC
MEKwE	-496.5925	999.185	1008.276	999.340	1002.878
EGOLE	-496.4853	1000.971	1013.092	1001.241	1003.433
EOLE	-496.8049	1001.609	1013.732	1001.879	1006.534
LIW	-514.9643	1035.929	1045.020	1036.090	1039.622
HLNHE	-506.3778	1018.756	1027.847	1018.917	1022.449
LEW	-512.8346	1031.669	1040.761	1031.830	1035.362
GIGE	-520.2300	1046.460	1055.551	1046.621	1050.153
EGIE	-507.7668	1021.534	1030.625	1021.695	1025.227

We assessed how well the recommended model fits compared to the competing models. The histogram and fitted density curve corresponding to MLE methods are plotted and displayed in left panel of the figure5. The empirical cumulative distribution function (ecdf) curve and the fitted cumulative distribution function (cdf) curve for each competing model, as well as the proposed model, are shown in the right panel of Figure 5. Analyzing these curves in the right panel allows for a thorough assessment of the goodness-of-fit for both the existing models and the proposed one. This examination assists in evaluating their respective adequacy in capturing the underlying patterns in the data.

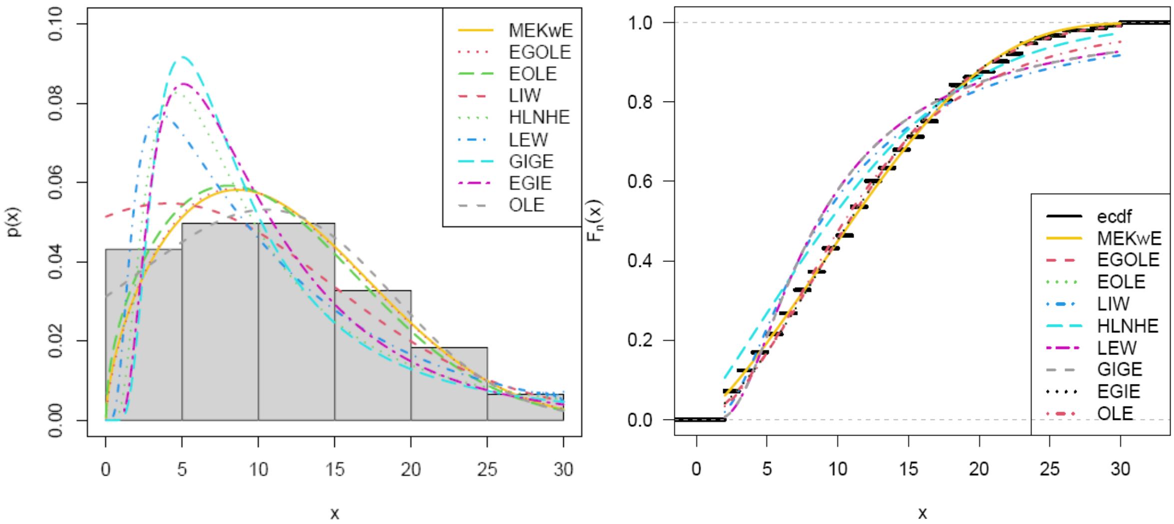

Figure 5: Histogram versus fitted pdf (left panel) and ecdf versus fitted cdf (Right panel)

5 Conclusion

We've developed a novel probability model named the Modified Extended Kumaraswamy Exponential Distribution in this study. Some properties like hazard rate, survival rate, quantile function etc are studied. Three methods of parameter estimation are used and the suitability of methods is compared using three methods of information criteria as well as by three methods goodness of fit statistics. Probability density curves and Hazard rate curves displayed. The hazard rate curves exhibit a monotonic increase, followed by a decrease, a period of constancy, constancy followed by an increase, and a J-shaped pattern across various parameter sets. We have utilized a real dataset on daily COVID-19 death counts in Nepal during the initial wave from January 23, 2020, to December 24, 2020 to assess the model's suitability. In this study, we have examined eight additional probability models from the existing literature to assess how well our proposed model compares to these competing models. We have also evaluated various information criteria values using four different methods. The results of our study reveal that the suggested model outperforms the other models we examined in terms of its capacity to accurately represent the data. We have plotted histograms and fitted density curves corresponding to Maximum Likelihood Estimation (MLE) methods. Additionally, we present cumulative distribution function (CDF) curves for both empirical and fitted distributions for all competing models, as well as for the proposed model. This graphical representation aids in assessing how well the proposed model aligns with the observed data and its ability to capture the underlying distribution. We conducted all the calculations and generated the graphs using the R programming language.

Acknowledgements

The authors would like to acknowledge the valuable contribution, comments and helpful suggestions made by the referees and the editors of the journal who have helped finalize this manuscript.

References

Aljarrah, M. A., Lee, C., and Famoye, F. (2014). On generating TX family of distributions using quantile functions. *Journal of Statistical Distributions and Applications*, *1*, 1-17. https://doi.org/10.1186/2195-5832-1-2

Al-saiary, Z. A., Bakoban, R. A., and Al-zahrani, A. A. (2019). Characterizations of the Beta Kumaraswamy Exponential Distribution. *Mathematics*, 8(1), 23. https://doi.org/10.3390/math8010023

Ansari, S. I., and Nofal, Z. M. (2021). The Lomax exponentiated Weibull model. *Japanese Journal of Statistics and Data Science*, *4*, 21-39. https://doi.org/10.1007/s42081-020-00073-0

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE transactions on reliability*, 36(1), 106-108. https://doi.org/10.1109/TR.1987.5222310

Cakmakyapan, S., and Gamze, O. Z. E. L. (2016). The Lindley family of distributions: properties and applications. *Hacettepe Journal of Mathematics and Statistics*, *46*(6), 1113-1137. http://dx.doi.org/10.15672/HJMS.201611615850

Chaudhary, A. K., Telee, L. B. S., and Kumar, V. (2022). Inverse Exponentiated Odd Lomax Exponential Distribution: Properties and Applications. *Nepalese Journal of Statistics*, 6(01), 29-50. http://dx.doi.org/10.3126/njs.v6i01.50801

Chaudhary, A. K., Sapkota, L. P., and Kumar, V (2023). Inverse Exponential Power distribution: Theory and Applications. *International Journal of Mathematics, Statistics and OperationsResearch,* 3(1),175-185. https://doi.org/10.47509/IJMSOR.2023.v03i01.11

Chaudhary, A. K., Sah Telee, L. B., and Kumar, V. (2023). Extended Kumaraswamy Exponential Distribution with Application to COVID-19 Data set. *Journal of Nepal Mathematical Society*, *6*(1), 1–10. https://doi.org/10.3126/jnms.v6i1.57657

Dhungana, G. P., and Kumar, V. (2022). Exponentiated Odd Lomax Exponential distribution with application to COVID-19 death cases of Nepal. *PloS one*, 17(6), e0269450. https://doi.org/10.1371/journal.pone.0269450

Gupta, R. D., and Kundu, D. (2001). Generalized exponential distribution: different method of estimations. *Journal of Statistical Computation and Simulation*, *69*(4), 315-337. https://doi.org/10.1080/00949650108812098

Hassan, A. S., & Nassr, S. G. (2019). Power Lindley-G family of distributions. *Annals of Data Science*, 6, 189-210. https://doi.org/10.1007/s40745-018-0159-y

Joshi, R. K., and Kumar, V. (2020a). Half Logistic NHE: Properties and Application. *International Journal for Research in Applied Science & Engineering Technology*, *8*(IX), 742-753. https://doi.org/10.22214/ijraset.2020.31557

Joshi, R. K., and Kumar, Vijay. (2020b). Lindley inverse Weibull distribution: Theory and Applications. *Bull. Math. & Stat. Res*, *8*(3), 32-46. http://www.bomsr.com/

Lai, C. D., Xie, M., and Murthy, D. N. P. (2003). A modified Weibull distribution. *IEEE Transactions on reliability*, *52*(1), 33-37. https://doi.org/10.1109/TR.2002.805788

Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37 (1), 25-32. https://doi.org/10.2307/2348376

Mudholkar, G. S., and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, *42*(2), 299-302. https://doi.org/10.1109/24.229504

Ogunsanya, A. S., Sanni, O. O., and Yahya, W. B. (2019). Exploring some properties of odd Lomax-exponential distribution. *Annals of Statistical Theory and Applications (ASTA)*, *1*, 21-30. http://dx.doi.org/10.13140/RG.2.2.12002.27840

Oguntunde, P. E., and Adejumo, A. O. (2015). The generalized inverted generalized exponential distribution with an application to a censored data. *Journal of Statistics Applications & Probability*, *4*(2), 223. http://dx.doi.org/10.12785/jsap/040204

Oguntunde, P. E., Adejumo, A., and Balogun, O. S. (2014). Statistical properties of the exponentiated generalized inverted exponential distribution. *Applied Mathematics*, *4*(2), 47-55. DOI: 10.5923/j.am.20140402.02

Tahir, M. H., Hussain, M. A., Cordeiro, G. M., El-Morshedy, M., and Eliwa, M. S. (2020). A new Kumaraswamy generalized family of distributions with properties, applications, and bivariate extension. *Mathematics,* 8(11), 1989.

https://doi.org/10.3390/math8111989

Telee, L. B. S., and Kumar, V. (2023). Modified Generalized Exponential Distribution. *Nepal Journal of Mathematical Sciences*, *4*(1). https://doi.org/10.3126/njmathsci.v4i1.53154

Telee, L. B. S., Karki, M., and Kumar, V. (2023). Model and Properties of Exponentiated Generalized Odd Lomax Exponential Distribution. *Interdisciplinary Journal of Management and Social Sciences*, 4(2). http://dx.doi.org/10.3126/ijmss.v4i2.57210

Worldometer (2023). Second wave of COVID-19 in Nepal. Retrieved from https://covid.ourworldindata.org/data/owid-covid-data.xlsx