Faculty of Natural and Applied Sciences Journal of Applied Biological Sciences
Print ISSN: 3027-0685, e-ISSN: 3043-6494
www.fnasjournals.com
Volume 2; Issue 1; September 2024; Page No. 49-63.

Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis

Afere, B.A.E.
Department of Mathematical Sciences, Prince Abubakar Audu University, Anyigba, Nigeria

Corresponding author email: baafere@gmail.com

Abstract
In this paper, we introduce a new flexible two-parameter unit interval univariate probability distribution called the "hybrid-Epanechnikov transformed Kumaraswamy distribution (HTKD)". The HTKD distribution, a flexible and robust version of the Kumaraswamy distribution, was created with the hybrid Epanechnikov kernel. We performed an in-depth analysis of its statistical characteristics and determined the parameters using the maximum likelihood estimation technique. The HTKD distribution's applicability was demonstrated using three datasets: the COVID-19 survival rate of Spain, the COVID-19 death rates of Canada, and the COVID-19 mortality rates of the UK. The HTKD distribution consistently offered the best fit when compared to other distributions, including the Beta, Kumaraswamy, BurrXII, and Weibull distributions, as seen by the lowest AIC and BIC values. These findings demonstrate the promise of the HTKD distribution as a flexible and useful tool for statistical analysis in epidemiological research. The uniformity of its performance across several datasets highlights its capacity to offer precise and dependable modelling of epidemiological data.

Keywords: Hybrid-Epanechnikov kernel, Kumaraswamy Distribution, HTKD Distribution, Maximum Likelihood Estimation,

Introduction

Numerous fields, including statistics, machine learning, probability theory, and economics, rely heavily on unit interval distributions. The distributions encompass the range [0, 1]. These distributions are essential for modelling ratios, rates, probabilities, and other naturally bounded values inside this interval.

Karl Pearson first introduced the beta distribution in the early 1900s, and it is among the most well-known and ancient unit interval distributions. The two shape parameters α and β constitute the beta distribution, which is versatile and may be used to model a wide range of phenomena (Pearson, 1905). The investigation of continuous probability distributions was made possible by this fundamental study. Kumaraswamy (1980) presented the Kumaraswamy distribution, a different distribution on the unit interval. In some applications, it is more user-friendly than the Beta distribution because of its simpler algebraic feature. Another notable distribution is the triangular distribution (Okoro et al., 2023), which has a minimum, maximum, and mode. Its adaptability to a variety of settings and practicality led to its popularity.

The Dirichlet distribution, a multivariate variant of the Beta distribution, has gained traction in the domains of machine learning and Bayesian statistics, particularly for modelling compositional data and proportions (Aitchison, 1982). Recently, more flexible versions of the beta distribution have been created to better capture complex data patterns on the unit interval (Ferrari & Cribari-Neto, 2004). Generalised versions such as the generalised Beta distribution and the Beta-Kumaraswamy distribution are proposed (Jones, 2009; Carrasco et al., 2010) to improve fit and flexibility for real-world data.

In probabilistic programming, neural networks, and Bayesian inference in machine learning, unit interval distributions are frequently utilised (Kucukelbir et al., 2017). New models address data with excess zeros or ones, which are common in ecology and economics (Ospina & Ferrari, 2010). Software packages and improved computational approaches have made working with unit interval distributions in complicated modelling settings easier (Stan Development Team, 2022).

Since its debut, there have been substantial changes in the research and application of unit interval distributions. From Pearson's seminal Beta distribution to the most recent advances in flexible distributions

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

and computational approaches, these distributions remain crucial for statistical modelling and data analysis. Because they are effective at defining bounding quantities and proportions, they are frequently used in a variety of industries. Research and development on unit interval distributions has increased dramatically in the last two decades. Scholars have developed several adaptations to the Beta distribution to address more complex data structures, including skewness and kurtosis (Jones, 2009; Carrasco et al., 2010). Generalised variants of unit interval distributions provide more flexibility and a better match for real-world data (Gupta & Kundu, 1999). Copulas are increasingly used to model dependency between random variables in joint distributions of proportions (Nelsen, 2006). Particularly in hierarchical models, the application of Bayesian inference has risen dramatically (Gelman et al., 2013).

Unit interval distributions are widely used in finance and economics to characterise probabilities, rates, and limited outcomes (Romano, 2002). Their integration into machine learning frameworks has enabled more sophisticated models and inference techniques (Kucukelbir et al., 2017). Specialised models that handle excess zeros or ones have a better fit for such data (Ospina & Ferrari, 2010). Improved algorithms and computational tools have made it easier to use unit interval distributions in complex and high-dimensional scenarios (Stan Development Team, 2022; Carpenter et al., 2017). In recent studies, Osatohanmwen et al. (2020) introduced a new class of generalised unit interval distributions with emphasis on the T-Kumaraswamy family of distributions; Demick and Liu (2022) explored copula modelling to analyse financial data; and Congdon (2019) developed unit interval-based Bayesian hierarchical models with applications in R.

The study of unit interval distributions has expanded from the fundamental Beta distribution to encompass a wide variety of adaptive, flexible, and computationally tractable models. Numerous fields, including machine learning, environmental statistics, finance, and economics, can benefit from these advancements. Their motivation stems from the need for accurate modelling of intricate data structures. To advance our understanding and proficiency in modelling data in the unit interval, this research is also being conducted. We introduce a novel unit interval distribution in this study, named the Hybrid-Epanechnikov Transformed Kumaraswamy Distribution (HTKD), analyse its statistical properties, and apply it to COVID-19 mortality rate data from the UK and Canada and COVID-19 survival rate data from Spain. Combining the popular Kumaraswamy distribution with the hybrid-Epanechnikov kernel (Afere, 2021; 2024) gives an opportunity to create the HTKD distribution. As developed in Section 2, this distribution leverages the benefits of both components to offer improved approximation, computational effectiveness, and adaptability to diverse unit interval data structures.

This work's subsequent sections are structured as follows: We examine some statistical characteristics of the HTKD distribution in Section 3, laying the groundwork for additional investigation. The technique for figuring out the parameters of the HTKD distribution is covered in full in Section 4, which dives into the maximum likelihood estimate approach. The HTKD distribution is then applied to real-world data, such as the COVID-19 survival rate data from Spain and the COVID-19 mortality rate data from the UK and Canada, in Section 5. Section 6's discussion of the results of these applications offers insights and interpretations of the findings. The final conclusion is given in Section 7.

Formulation of the New Distribution

The distribution function (CDF) and the density function (PDF) of the Kumaraswamydistribution are given as:

$$F_Y(y;a,b) = 1-(1-y^a)^b, \quad y \in [0,1]. \tag{1}$$

and

$$f_Y(y;a,b) = ab\cdot y^{a-1}\cdot(1-y^a)^{b-1}, \quad y \in [0,1], \tag{2}$$

where $a > 0$ and $b > 0$ are the shape parameters. The PDF of the hybrid Epanechnikov kernel (Afere, 2021) is given as:

$$f_X(x) = \frac{1}{8}(5-3x^2), \quad x \in [-1,1]. \tag{3}$$

The CDF of (3) is achieved by evaluating the antiderivative of (3) from -1 to a random variable u. Hence,we have:

$$F_U(u) = \frac{1}{8}(5u-u^3+4), \quad u \in [-1,1] \tag{4}$$

Equation (4) can be substituted into Equation (1) to obtain the CDF of the newly developed two-parameter HTKD distribution as follows

:

$$F_Y(u;a,b) = 1-\left(1-\left(\frac{1}{8}(5u-u^3+4)\right)^a\right)^b, u \in [0,1]. \tag{5}$$

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Consequently, after replacing Equation (4) with Equation (2), the new HTKD distribution's PDF takes on the following form:

$$f_Y(u;a,b) = ab \cdot \left(\frac{1}{8}(5u - u^3 + 4)\right)^{a-1} \cdot \left(1 - \left(\frac{1}{8}(5u - u^3 + 4)\right)^{a}\right)^{b-1}, u \in [0,1], \quad (6)$$

where $a > 0$ and $b > 0$ are the shape parameters

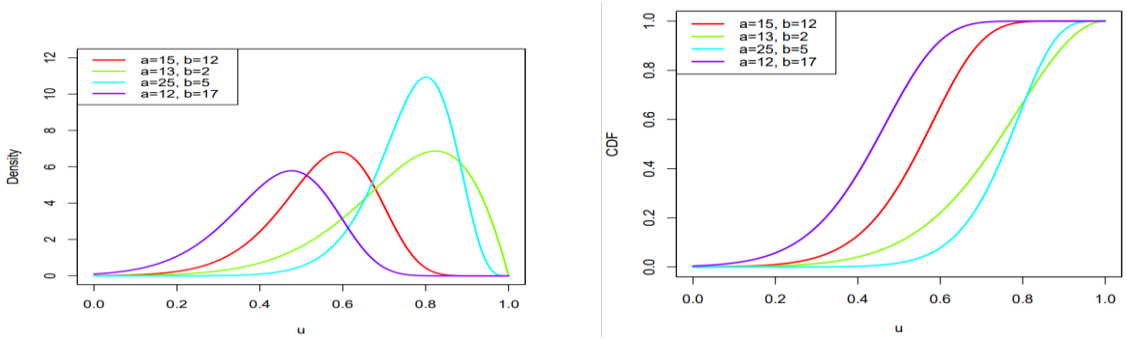

Figure 1: Plots of the PDFs (left) and CDFs (right) of HTKD distribution

Statistical Properties

We will discuss the statistical characteristics of the HTKD distribution in this section. Moments, entropy, mode, quantiles, survival function, hazard function, cumulative hazard function, reversed hazard function, and other attributes will all be covered.

Survival Function

In survival analysis and reliability theory, the survival function, often represented as $S(t)$, is an essential notion. It provides the likelihood that a subject or system will endure after a specific amount of time (t). The survival function can be expressed mathematically as the complement of the time-to-event variable T's cumulative distribution function:

$$S(t) = P(T > t) = 1 - F(t), \quad (7)$$

where $F(t)$ is the cumulative distribution function of T, which represents the probability that the event has occurred at time t. T is the random variable that represents the time of the event of interest. Accordingly, the survival function of the HTKD distribution is given by:

$$S(t;a,b) = \left(1 - \left(\frac{1}{8}(5t - t^3 + 4)\right)^{a}\right)^{b}. \quad (8)$$

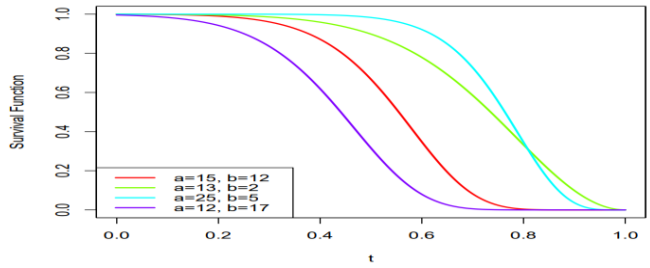

Figure 2: Survival function of HTKD distribute

Hazard Function

The failure rate, or force of mortality, is another name for the hazard function, which is an essential idea in reliability theory and survival analysis. It measures the immediate failure rate at a specific time t, assuming survival through that point. The hazard function $h(t)$
has the following mathematical definition:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t}. \quad (9)$$

As an alternative, Equation (9) can be stated as follows in terms of the survival function $S(t)$ and the probability density function $f(t)$:

$$h(t) = \frac{f(t)}{S(t)}. \quad (10)$$

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Where $f(t)$ is the probability density function of T, which describes the likelihood of the event occurring at an exact time t, $S(t)$ is the survival function, which represents the probability that the subject or system survives beyond time t, and T is as defined in Equation (8). The hazard function $h(t)$ for the HTKD distribution is given as follows:

$$h(t) = \frac{ab\left(\frac{1}{8}(5t - t^3 + 4)\right)^{a-1}\left(1 - \left(\frac{1}{8}(5t - t^3 + 4)\right)^{a}\right)^{b-1}}{\left(1 - \left(\frac{1}{8}(5t - t^3 + 4)\right)^{a}\right)^{b}}$$
$$= ab\left(\tfrac{1}{8}(5t - t^3 + 4)\right)^{a-1}\left(1 - \left(\tfrac{1}{8}(5t - t^3 + 4)\right)^{a}\right)^{-1} \quad (11)$$

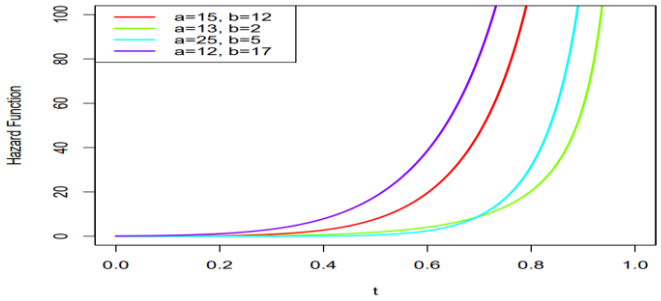

Figure 3: Hazard function of HTKD distribution

Reversed Hazard Function

Reliability theory and survival analysis both make use of the reverse hazard function, sometimes referred to as the reversed hazard rate. If the event hasn't happened in the present, it gives the instantaneous rate of occurrence of an event in the past. In essence, it measures the probability that, in the absence of an event, it would have happened shortly before a certain period.

The reverse hazard function $\tilde{h}(t)$ has the following mathematical definition:

$$\tilde{h}(t) = \lim_{\Delta t \to 0} \frac{P(t - \Delta t \le T < t | T \ge t)}{\Delta t}. \quad (12)$$

As an alternative, the probability density function $f(t)$ and the cumulative distribution function $F(t)$ can be used to represent Equation (12) as follows:

$$\tilde{h}(t) = \frac{f(t)}{1 - F(t)}. \quad (13)$$

where T is defined by Equation (7). $F(t)$ is the cumulative distribution function of T, providing the probability that t he event has occurred by time t. $f(t)$ is the probability density function of T, describing the likelihood of the event occurring at an exact time t. The HTKD distribution's $\tilde{h}(t)$ can be determined as follows:

$$\tilde{h}(t) = \frac{ab\left(\frac{1}{8}(5t - t^3 + 4)\right)^{a-1}\left(1 - \left(\frac{1}{8}(5t - t^3 + 4)\right)^{a}\right)^{b-1}}{1 - \left(1 - \left(\frac{1}{8}(5t - t^3 + 4)\right)^{a}\right)^{b}}. \quad (14)$$

.

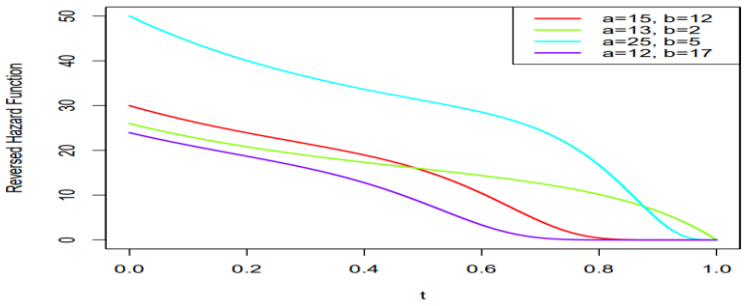

Figure 4. Reversed hazard function of HTKD distribution

Cumulative Hazard Function

A fundamental idea in reliability theory and survival analysis is the cumulative hazard function. It shows the overall amount of risk that has built up at that point t. This feature is helpful in determining the likelihood that an event (like failure or death) will occur over time. The integral of the hazard function $h(t)$ from time 0 to time t is the mathematical definition of the cumulative hazard function $H(t)$:

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

$$H(t) = \int_0^t h(u)\, du. \tag{15}$$

The hazard function, denoted as $h(t)$ in Equation (15), provides the instantaneous rate of occurrence of the event at time t, provided that it has not happened before time t. As an alternative, the survival function $S(t)$ can be used to express the cumulative hazard function as follows:

$$H(t) = -\ln S(t). \tag{16}$$

The cumulative hazard function for the HTKD distribution function is provided by:

$$H(t) = \int_0^t \frac{ab\left(\frac{1}{8}(5u-u^3+4)\right)^{a-1}\left(1-\left(\frac{1}{8}(5u-u^3+4)\right)^{a}\right)^{b-1}}{1-\left(1-\left(\frac{1}{8}(5u-u^3+4)\right)^{a}\right)^{b}} du$$
$$= -\ln\left[1-\left(1-\left(\frac{1}{8}(5t-t^3+4)\right)^{a}\right)^{b}\right] \tag{17}$$

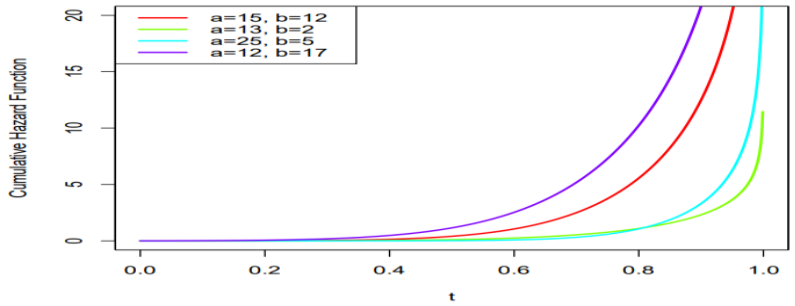

Figure 5: Cumulative hazard function of HTKD distribution

Quantile Function

The following is the derivation of the quantile function u_q of the HTKD distribution: Let U be a random variable and let $F_U(u)$ be its CDF. $Q(q)$, the q-th quantile, is defined as follows:

$$Q(q) = \inf\{u \in \mathbb{R}: F_U(u) \geq q\}. \tag{18}$$

For the HTKD distribution, we set $F_Y(u) = q$ and solve for u in terms of q to obtain the quantile function $Q(q)$. Therefore,

$$q = 1-\left(1-\left[\frac{1}{8}(5u-u^3+4)\right]^{a}\right)^{b}, \tag{19}$$

But $F_U(u) = \frac{1}{8}(5u-u^3+4)$. Hence, the CDF $F_Y(u)$ becomes:

$$q = 1-(1-[F_U(u)]^{a})^{b}. \tag{20}$$

On solving for $F_U(u)$, we have:

$$F_U(u) = \left(1-(1-q)^{1/b}\right)^{1/a}. \tag{21}$$

This implies that:

$$\frac{1}{8}(5u-u^3+4) = \left(1-(1-q)^{1/b}\right)^{1/a}.$$

Solving for u, we have:

$$u^3 - 5u + \left(4 - 8\left(1-(1-q)^{1/b}\right)^{1/a}\right) = 0. \tag{22}$$

Thus, using Cardano's formula (Cardano, 1993), which provides solutions for the roots of a cubic equation , Equation (22) becomes

$$u_q = \inf\left|\sqrt[3]{-\frac{4-8\left(1-(1-q)^{1/b}\right)^{1/a}}{2}+\sqrt{\left(\frac{4-8\left(1-(1-q)^{1/b}\right)^{1/a}}{2}\right)^2-\left(\frac{5}{3}\right)^3}}\right.$$
$$\left.+\sqrt[3]{-\frac{4-8\left(1-(1-q)^{1/b}\right)^{1/a}}{2}-\sqrt{\left(\frac{4-8\left(1-(1-q)^{1/b}\right)^{1/a}}{2}\right)^2-\left(\frac{5}{3}\right)^3}}\right|. \tag{23}$$

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Quantile Approach for Skewness and Kurtosis for HTKD Distribution

Alternative measurements of a distribution's form based on quantiles (percentiles) as opposed to moments are offered by the quantile approach to skewness and kurtosis computation. When the distribution is given in closed form or within a straightforward analytical framework, this method is especially helpful. However, programs such as R can handle the distribution if it is not in a closed form. A quantile-based method for skewness was presented by Kenney and Keeping (1962), and a similar approach for kurtosis was offered by Moore (1988).

For the HTKD distribution, we can use the quantile function in Equation (23) to get Moor's kurtosis and Bowley's skewness (Bowley, 1920) using the following formula:

$$\text{Bowley's Skewness} = \frac{Q_{\frac{1}{4}} + Q_{\frac{3}{4}} - 2\times Q_{\frac{1}{2}}}{Q_{\frac{3}{4}} - Q_{\frac{1}{4}}}. \quad (24)$$

: and

$$\text{Moor's Kurtosis} = \frac{(Q_{\frac{7}{8}} - Q_{\frac{5}{8}}) + (Q_{\frac{3}{4}} - Q_{\frac{1}{4}})}{Q_{\frac{3}{4}} - Q_{\frac{1}{4}}}. \quad (25)$$

where, $Q_{\frac{1}{8}}, Q_{\frac{1}{4}}, Q_{\frac{3}{8}}, Q_{\frac{5}{8}}, Q_{\frac{3}{4}}, Q_{\frac{7}{8}}$ are specific quantiles at 12.5%, 25%, 37.5%, 62.5%, 75%, and 87.5% respectively.

The first quartile $Q_{\frac{1}{4}}$ and the third quartile $Q_{\frac{3}{4}}$ in a symmetric distribution are equally spaced from the median $Q_{\frac{1}{2}}$, leading to skewness that is almost zero. A longer right tail is indicated by positive skewness when $Q_{\frac{3}{4}}$ is farther from $Q_{\frac{1}{2}}$; a longer left tail is indicated by negative skewness when $Q_{\frac{1}{4}}$ is farther from $Q_{\frac{1}{2}}$. On the other hand, kurtosis quantifies how heavy or light the tails are in relation to a normal distribution. Negative kurtosis (platykurtic) indicates lighter tails and fewer extreme values, whereas positive kurtosis (leptokurtic) indicates heavier tails and more extreme values.

Table 1: Descriptive statistics for HTKD distribution

A	*B*	Median	LQ	UQ	Bowley's Skewness	Moor's Kurtosis	IQR
4	4.5	0.1847	6.6107e-05	0.3572	-0.0341	0.8350	0.3572
5	3.0	0.3776	1.9307e-01	0.5436	-0.0528	1.2150	0.3506
6	2.0	0.5344	3.5336e-01	0.6916	-0.0705	1.2067	0.3382
7	1.5	0.6408	4.6726e-01	0.7854	-0.0908	1.1967	0.3182
4	2.5	0.3295	1.1903e-01	0.5206	-0.0484	1.1167	0.4016
2	1.5	0.1744	6.6107e-05	0.4624	0.2460	0.7625	0.4623

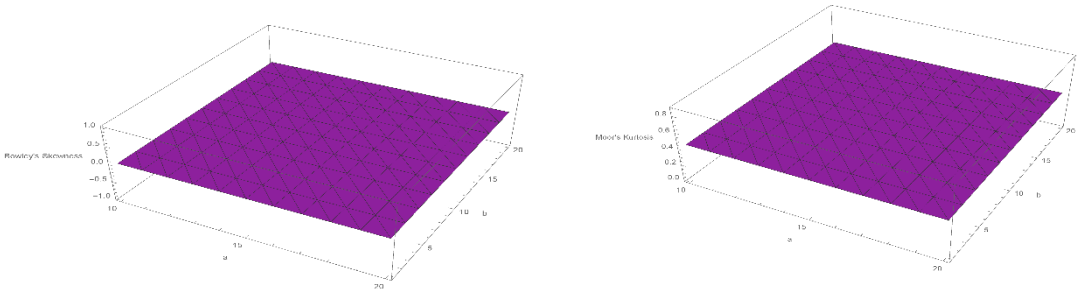

Figure 6: 3D Plots of Bowley's Skewness (Left), Moor's Kurtosis (Right) of HTKD distribution

Table 1 and Figure 6 provide a thorough grasp of the properties of the HTKD distribution for different combinations of parameters. Several important insights about the behaviour of the distribution and its sensitivity to variations in parameters a and b are revealed by evaluating these data.

First off, there is a general symmetry or mild left skewness in the distributions. The Bowley's skewness values, which are either slightly negative or hang around zero for the majority of parameter combinations, demonstrate this. The parameter pair (2, 1.5) stands out as an example, exhibiting a positive skewness that suggests a little right skew.

Second, larger median and quartile values are the result of greater values of a and lower values of b. In Table 1, the median, upper quartile (UQ), and lower quartile (LQ) numbers all clearly show this pattern. An increase in a indicates a higher concentration of data around the median since it causes the distribution's central tendency to shift upward and the gap between the quartiles to widen.

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Thirdly, for all parameter combinations, the distributions show prominent tails, a feature called leptokurtosis. According to Moor, the distributions have larger tails than a normal distribution because their kurtosis values are constantly greater than 0. Significantly, smaller b values are associated with higher kurtosis values, indicating that lower b may be a contributing factor to more severe values in the tails.

Lastly, when increases, the interquartile range (IQR), which represents the spread of the centre 50% of the data, tends to get smaller. With increasing value, this suggests that the data become more concentrated around the median, indicating a decrease in variability within the central section of the distribution.

These observations are helpful in comprehending how the HTKD distribution behaves and how sensitive it is to variations in parameters a and b. Through a thorough examination of the descriptive statistics and graphical displays, the subtle impacts of these parameters on the form, central tendency, and variability of the distribution can be identified.

Mode

To find the mode of the PDF of the HTKD distribution, we need to determine the value of u that maximizes $f_Y(u)$. Thus, differentiating $f_Y(u)$ and solving for u, we obtain the expression:

$$\left(\frac{a-1}{a}\right)\left(1-\left(\frac{1}{8}(5u-u^3+4)\right)^a\right)=\left(\frac{1}{8}(5u-u^3+4)\right)^{a-1}. \tag{26}$$

Due to the complexity of (26), numerical solutions are typically needed. Various root-finding techniques, such as Newton-Raphson, are employed to determine the value of u that maximizes $f_Y(u)$. R can be used to accomplish these techniques. The value of u that fulfills Equation (26) is the mode of the HTKD distribution.

Series Expansion of the PDF and CDF of HTKD Distribution

The compact version of the density function of the HTKD distribution can be expressed by power series expansion, as detailed as follows:

$$\begin{aligned} f_Y(u) &= ab\left(\frac{1}{8}(5u-u^3+4)\right)^{a-1}\left(1-\left(\frac{1}{8}(5u-u^3+4)\right)^a\right)^{b-1} \\ &= ab\,\tfrac{1}{2^{ai_1+a-1}}\textstyle\sum_{i_1=0}^{\infty}(-1)^{i_1}\binom{b-1}{i_1}\left(1+\left(-\frac{1}{4}u^3+\frac{5}{4}u\right)\right)^{ai_1+a-1} \end{aligned} \tag{27}$$

Let $m=ai_1+a-1$, then we have:

$$f_Y(u)=ab\sum_{i_1=0}^{\infty}\sum_{j_1=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{i_1+j_1-k}\binom{b-1}{i_1}\binom{m}{j_1}\binom{j_1}{k}\frac{5^k}{2^{m+2j_1}}u^{3j_1-2k}$$

$$=\textstyle\sum_{i_1=0}^{\infty}C_{i_1,j_1,k}u^{3j_1-2k} \tag{28}$$

where $C_{i_1,j_1,k}=ab\sum_{j_1=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{i_1+j_1-k}\binom{b-1}{i_1}\binom{m}{j_1}\binom{j_1}{k}\frac{5^k}{2^{m+2j_1}}$. Similarly, the power series expansion can be used to lower the distribution function of the HTKD distribution as follows:

$$\begin{aligned} F_Y(u) &= 1-\sum_{i_2=0}^{\infty}\sum_{j_2=0}^{\infty}(-1)^{i_2+ai_2-j_2}\binom{b}{i_2}\binom{ai_2}{j_2}\frac{5^{j_2}}{2^{3ai_2}}u^{3ai_2-2j_2} \\ &= 1-\textstyle\sum_{i_2=0}^{b}D_{i_2,j_2}u^{3ai_2-2j_2} \end{aligned} \tag{29}$$

where, $D_{i_2,j_2}=\sum_{j_2=0}^{\infty}(-1)^{i_2+ai_2-j_2}\binom{b}{i_2}\binom{ai_2}{j_2}\frac{5^{j_2}}{2^{3ai_2}}$.

Moments and Moments Generating Function

A probability distribution's moments are numerical measurements that characterise the distribution's form, centre, spread, and other features. Particular distributional information is given each time. The expected value of a random variable U's n-th power is its n-th moment. The n-th moment has the following mathematical definition:

$$\mathrm{E}(U^n)=\textstyle\int_{-\infty}^{\infty}u^n\cdot f_Y(u)\,du. \tag{30}$$

where n is a non-negative integer, and $f_Y(u)$ is the PDF of the HTKD distribution. Equation (28) can be substituted into Equation (30) to yield the following:

$$E(u^n)=\int_0^1 u^n\sum_{i_1=0}^{\infty}C_{i_1,j_1,k}u^{3j_1-2k}$$

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

$$= \sum_{i_1=0}^{\infty} C_{i_1,j_1,k} \int_0^1 u^n u^{3j_1-2k} du$$
$$= \sum_{i_1=0}^{\infty} C_{i_1,j_1,k} \int_0^1 u^{n+3j_1-2k} du$$
$$= \frac{\sum_{i_1=0}^{\infty} C_{i_1,j_1,k}}{n+3j_1-2k+1}$$

As a result, the HTKD distribution's n-th moments are provided by:

$$E(u^n) = \frac{\sum_{i_1=0}^{\infty} C_{i_1,j_1,k}}{n+3j_1-2k+1}. \quad (31)$$

Equally, the MGF $m(t)$ is defined by:

$$m(t) = E(e^{tU}) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(u^n). \quad (32)$$

Therefore, the MGF of the HTKD distribution can be obtained by substituting Equation (31) into Equation (32) and having:

$$m(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \left(\frac{\sum_{i_1=0}^{\infty} C_{i_1,j_1,k}}{n+3j_1-2k+1} \right)$$
$$= \sum_{n=0}^{\infty} \sum_{i_1=0}^{\infty} C_{i_1,j_1,k} \frac{t^n}{n!\,(n+3j_1-2k+1)} \quad (33)$$

Rényi Entropy

A term from information theory and statistics called Rényi entropy measures the uncertainty or randomness of a probability distribution or random variable. It quantifies the average level of surprise, unpredictability, or information present in a random variable's results (Rényi, 1961). In the event where U is a random variable with a PDF $f_U(u)$, the following formula provides the Rényi entropy:

$$I_y(u) = \frac{1}{1-\lambda} \log \int_{-\infty}^{\infty} \left(f_Y(u) \right)^{\lambda} du. \quad (34)$$

Substituting Equation (28) into Equation (34), the entropy of the HTKD distribution is given as:

$$I_y(u) = \frac{1}{1-\lambda} \log \int_0^1 \left(\sum_{i_1=0}^{\infty} C_{i_1,j_1,k} u^{3j_1-2k} \right)^{\lambda} du$$
$$= \frac{1}{1-\lambda} \left(\sum_{i_1=0}^{\infty} C_{i_1,j_1,k} u^{3j_1-2k} \right)^{\lambda} \log \int_0^1 u^{\lambda(3j_1-2k)} du \quad (35)$$

Hence, Equation (35) becomes:

$$I_y(u) = \frac{1}{1-\lambda} \left(\sum_{i_1=0}^{\infty} C_{i_1,j_1,k} u^{3j_1-2k} \right)^{\lambda} \log \frac{1}{\lambda(3j_1-2k)}$$

Order Statistics of the PDF

The formula for the order statistics of a random sample is utilized to get the PDF of the k-th order statistics $U_{(k)}$:

$$f_{U_{(k)}}(u) = \frac{n!}{(k-1)!(n-k)!} (F_Y(u))^{k-1} (1 - F_Y(u))^{n-k} f_Y(u). \quad (36)$$

where, $F_Y(u)$ and $f_Y(u)$ are respectively the CDF and PDF of the HTKD distribution. Now, on substituting the series form of $F_Y(u)$ and $f_Y(u)$ given in Equations (28) and (29) respectively into Equation (34), we have:

$$f_{U_{(k)}}(u) = \frac{n!}{(k-1)!(n-k)!} \sum_{r=0}^{\infty} \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} (-1)^r \binom{k-1}{r} C_{i_1,j_1,k} \left(D_{i_2,j_2} \right)^{r+n-k} u^{\tau}. \quad (37)$$

where, $\tau = 2r(ai_2 - j_2) + 2(n-k)(ai_2 - j_2) + 3j_1 - 2k$, $C_{i_1,j_1,k}$ and D_{i_2,j_2} are as previously defined.

Maximum Likelihood Estimation of the Parameters

For an entire random independent sample $u_1, u_2, \ldots, u_n$, the HTKD distribution's log-likelihood function can be found as follows:

$$\ell(a,b) = \sum_{i=1}^{n} \left[\ln b + \ln a + (a-1) \ln \left(\frac{1}{8}(5u_i - u_i^3 + 4) \right) + (b-1) \ln \left(1 - \left(\frac{1}{8}(5u_i - u_i^3 + 4) \right)^a \right) \right]. \quad (38$$

The score vector $\mathbf{S}(a,b)$ of the parameters a and b of the log-likelihood function $\ell(a,b)$ is given by:

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

$$\mathbf{S}(a,b) = \begin{pmatrix} \frac{\partial \ell(a,b)}{\partial a} \\ \frac{\partial \ell(a,b)}{\partial b}. \end{pmatrix} \tag{39}$$

where:

$$\frac{\partial \ell(a,b)}{\partial a} = \sum_{i=1}^{n} \left[\frac{1}{a} + \ln\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right) - (b-1) \cdot \frac{\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a \ln\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)}{1 - \left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a} \right], \tag{40}$$

$$\frac{\partial \ell(a,b)}{\partial b} = \sum_{i=1}^{n} \left[\frac{1}{b} + \ln\left(1 - \left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a\right)\right]. \tag{41}$$

Setting the components of $\mathbf{S}(a,b)$ to zero results in a system of linear equations. These equations can be solved numerically to find the maximum likelihood estimates $\hat{a}$ and $\hat{b}$. Methods like Newton-Raphson or other iterative algorithms can be employed for this purpose, using software such as R.

The Fisher Information Matrix (FIM) $\mathbf{I}(a,b)$ is computed to evaluate the accuracy of the parameter estimations [(Lauritzen, 1996)]. The variance-covariance matrix, which explains the standard errors and correlations between the parameter estimations, is obtained by taking the inverse of the FIM. The following are the elements of the Fisher Information Matrix $\mathbf{I}(a,b)$:

$$\mathbf{I}(a,b) = \begin{pmatrix} -\theta_{11} & -\theta_{12} \\ -\theta_{21} & -\theta_{22} \end{pmatrix} \tag{42}$$

where,

$$\theta_{11} = \sum_{i=1}^{n} \left(-\frac{1}{a^2} - \frac{(b-1)\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a \left(\ln\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)\right)^2}{\left(1 - \left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a\right)^2} \right), \tag{43}$$

$$\theta_{12} = \theta_{21} = \sum_{i=1}^{n} \left(-\frac{\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a \ln\left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)}{1 - \left(\frac{1}{8}(5u_i - u_i^3 + 4)\right)^a} \right), \tag{44}$$

$$\theta_{22} = \sum_{i=1}^{n} \left(-\frac{1}{b^2} \right). \tag{45}$$

The inverse of the Fisher Information Matrix provides the variance-covariance matrix:

$$\mathbf{I}^{-1}(\hat{a},\hat{b}) = \begin{pmatrix} \mathrm{Var}(\hat{a}) & \mathrm{Cov}(\hat{a},\hat{b}) \\ \mathrm{Cov}(\hat{a},\hat{b}) & \mathrm{Var}(\hat{b}) \end{pmatrix}. \tag{46}$$

We generate a random variable U from a Beta distribution with $n = 50$, specifically $\mathrm{Beta}(2,2)$, and compute the Fisher Information Matrix (FIM) and its inverse in order to evaluate the accuracy of HTKD parameter estimates:

$$\mathbf{I} = \begin{pmatrix} 50.00000 & -44.00334 \\ -44.00334 & 50.00000 \end{pmatrix}$$

and

$$\mathbf{I}^{-1} = \begin{pmatrix} 0.08869873 & 0.07806081 \\ 0.07806081 & 0.08869873 \end{pmatrix}$$

The Fisher Information Matrix ($\mathbf{I}$) offers valuable insights on the correlation and degree of uncertainty associated with the parameter underestimation. In the case of that specific parameter, a bigger diagonal value (50.00000) denotes reduced uncertainty and higher precision in the estimation process. The estimations of the two parameters appear to have an inverse connection, as indicated by the negative off-diagonal value (-44.00334). Conversely, the covariance matrix of the parameter estimates, which shows the correlations and uncertainties following estimation, is given by $\mathbf{I}^{-1}$. The variance of the parameter estimations is represented by the diagonal elements (0.08869873), where smaller numbers denote greater precision. A positive correlation between the estimations of the parameters of the HTKD distribution is indicated by the positive off-diagonal elements (0.07806081).

Monte Carlo Experiments

In order to evaluate the efficiency and efficacy of the maximum likelihood estimators of the parameters of the HTKD distribution, we run Monte Carlo experiments in this section. These tests are carried out using various parameter mixes and sample sizes. The experiments are repeated for $r = 10000$ runs, with mixtures of parameters $a = 5, b = 2$, $a = 2, b = 1.5$, $a = 4, b = 4.5$, and $a = 4, b = 2.5$, and sample sizes $n = 25$, $n = 60$, $n = 200$, $n = 500$, and $n = 3000$.

If $\hat{\ell}$ represents the maximum likelihood estimate of ℓ, then the following evaluation metrics are used to compare the efficiency and performance of the parameters of the HTKD distribution:

1. The maximum likelihood estimator's mean estimates (ME) are provided by:

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

$$ME = \frac{1}{N}\sum_{i=1}^{N} \hat{\ell}_i$$

2. The maximum likelihood estimator's average bias (AVB) is provided by:
$$AVB = \frac{1}{N}\sum_{i=1}^{N} (\hat{\ell}_i - \ell_i)$$
3. The maximum likelihood estimator's root mean squared error (RMSE) is given by:
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{\ell}_i - \ell_i)^2}$$
4. Average width (AW) of the parameter ℓ 's 95% confidence intervals.
5. The 95% confidence intervals of the parameters' ℓ coverage probability (CP).

The Monte Carlo experiment results for four parameter combinations $a = 5, b = 2$, $a = 2, b = 1.5$, $a = 4, b = 4.5$, and $a = 4, b = 2.5$ are shown in Tables 2, 3, 4, and 5 respectively.

Table 2: Monte Carlo experiment for parameter $a = 5, b = 2$ for different values of n

Parameter	Sample Size	ME	AVB	RMSE	AW	CP
a	25	5.3893	4.9408	4.9487	4.6149	0.9443
	60	5.1765	4.7289	4.7366	2.8876	0.9543
	200	5.0640	4.6145	4.6222	1.5588	0.9614
	500	5.0196	4.5724	4.5802	0.9781	0.9443
	3000	5.0013	4.5542	4.5619	0.3981	0.9486
b	25	2.2956	1.8470	1.8708	2.8306	0.9514
	60	2.1265	1.6789	1.7016	1.6415	0.9571
	200	2.0221	1.5725	1.5953	0.8415	0.9600
	500	2.0141	1.5669	1.5896	0.5295	0.9657
	3000	2.0009	1.5537	1.5765	0.2142	0.9529

Table 3: Monte Carlo experiment for parameter $a = 2, b = 1.5$ for different values of n

Parameter	Sample Size	ME	AVB	RMSE	AW	CP
a	25	2.1881	2.0344	2.0757	1.9772	0.9500
	60	2.0653	1.9086	1.9516	1.2262	0.9557
	200	2.0278	1.8739	1.9165	0.6617	0.9529
	500	2.0102	1.8541	1.8972	0.4165	0.9386
	3000	2.0002	1.8447	1.8879	0.1693	0.9486
b	25	1.7441	1.5904	1.6465	2.0210	0.9757
	60	1.5737	1.4169	1.4759	1.1409	0.9571
	200	1.5383	1.3845	1.4420	0.6055	0.9486
	500	1.5102	1.3540	1.4127	0.3743	0.9357
	3000	1.5007	1.3452	1.4039	0.1515	0.9386

Table 4: Monte Carlo experiment for parameter $a = 4, b = 4.5$ for different values of n

Parameter	Sample Size	ME	AVB	RMSE	AW	CP
a	25	4.2950	4.1264	4.1348	3.2151	0.9443
	60	4.1427	3.9718	3.9803	2.0233	0.9514
	200	4.0221	3.8512	3.8601	1.0820	0.9514
	500	4.0103	3.8402	3.8491	0.6828	0.9586
	3000	3.9997	3.8291	3.8379	0.2783	0.9429
b	25	5.6601	5.4915	5.4991	8.6174	0.9629
	60	4.9398	4.7688	4.7764	4.6230	0.9714
	200	4.5679	4.3970	4.4049	2.2792	0.9500
	500	4.5410	4.3709	4.3787	1.4297	0.9571
	3000	4.4968	4.3262	4.3340	0.5762	0.9600

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Table 5: Monte Carlo experiment for parameter $a = 4, b = 2.5$ for different values of n

Parameter	Sample Size	ME	AVB	RMSE	AW	CP
a	25	4.2821	3.9738	3.9850	3.5139	0.9429
	60	4.1584	3.8490	3.8597	2.2199	0.9571
	200	4.0353	3.7265	3.7376	1.1888	0.9614
	500	4.0127	3.7037	3.7149	0.7491	0.9529
	3000	4.0018	3.6927	3.7039	0.3052	0.9400
b	25	2.9228	2.6145	2.6336	3.8082	0.9600
	60	2.7016	2.3922	2.4103	2.1976	0.9671
	200	2.5478	2.2389	2.2576	1.1142	0.9500
	500	2.5182	2.2093	2.2279	0.6942	0.9529
	3000	2.5034	2.1943	2.2130	0.2812	0.9386

For the HTKD distribution, which is defined by parameters *a* and *b*, the Monte Carlo experiments shown in Tables 2, 3, 4, and 5 investigate the performance of parameter estimation under various settings. Each table corresponds to particular combinations of these parameters and describes how different sample sizes affect estimation accuracy, bias, and interval precision.

For example, the mean estimate (ME) progressively converges to the real value of 5 as the sample size n grows in Table 2, where $a = 5$ and $b = 2$. This convergence points to a fair approximation. Smaller numbers indicate less bias and better accuracy. The average absolute bias (AVB) and root mean squared error (RMSE) quantify the divergence between estimated and true values. The accuracy of the estimate is gauged by the average width of the 95% confidence interval (AW); smaller intervals correspond to more accurate estimations. Coverage Probability (CP) evaluates how effectively these intervals capture the genuine parameter values; for 95% confidence intervals, it should ideally be around 0.95.

Table 3 examines parameters $a = 2$ and $b = 1.5$, showing a similar trend of unbiased estimation and increased precision with increasing sample numbers. The interpretation of AVB, RMSE, AW, and CP is consistent with Table 2, demonstrating the performance of the estimate approach at various parameter values.

The ME values in Table 4, where $a = 4$ and $b = 4.5$, indicate the degree to which estimations resemble the genuine values of 4 and 4.5, respectively. Important insights into bias reduction, accuracy enhancement, and interval dependability are provided here, as in the other tables, by AVB, RMSE, AW, and CP. Table 5 similarly concentrates on $a = 4$ and $b = 2.5$, showing how ME values improve with larger sample numbers, signifying more precise estimation. Metrics like AVB, RMSE, AW, and CP provide further insight into how well the estimate process performs in various experimental setups. Together, these tables essentially highlight how sample size affects estimate quality. As seen by lower AVB and RMSE, narrower AWs, and higher CPs, larger sample sizes of ten result in more accurate and less biased parameter estimations. This thorough analysis highlights the value of the Monte Carlo simulation technique in statistical modelling and inference by demonstrating how reliable it is in evaluating the HTKD distribution parameters.

Applications

Here, we show that the HTKD distribution is applicable by utilising COVID-19 mortality rate data from the UK and Canada in Tables 6 and 8, respectively, and COVID-19 survival rate data from Spain in Table 10. The datasets are secondary data from Obulezi et al. (2023). We assess how well the novel HTKD distribution performs in comparison to some of the existing two-parameter distributions (Kumaraswamy, Beta, Weibull, and BurrXII).

Table 6: First Data Set: UK's COVID-19 Mortality Rate

0.1292	0.3805	0.4049	0.2564	0.3091	0.2413	0.1390	0.1127	0.3547	0.3126
0.2991	0.2428	0.2942	0.0807	0.1285	0.2775	0.3311	0.2825	0.2559	0.2756
0.1652	0.1072	0.3383	0.3575	0.2708	0.2649	0.0961	0.1565	0.1580	0.1981
0.4154	0.3990	0.2483	0.1762	0.1760	0.1543	0.3238	0.3771	0.4132	0.4602
0.3520	0.1882	0.1742	0.4033	0.4999	0.3930	0.3963	0.3960	0.2029	0.1791
0.4768	0.5331	0.3739	0.4015	0.3828	0.1718	0.1657	0.4542	0.4772	0.3402

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Table 7: Model Comparison of Different Distributions for UK's COVID-19 Mortality Rate Dataset

Distribution	Parameter 1	Parameter 2	AIC	BIC
HTKD	*a* : 10.99207	*b* : 44.45731	-146.04328	-141.85459
Beta	*Shape1* : 4.05028	*Shape2* : 10.01369	-86.79892	-82.61023
Kumaraswamy	*Shape1* : 2.68061	*Shape2* : 19.58734	-87.72880	-83.54011
BurrXII	*a* : 2.82378	*b* : 24.49530	-86.80269	-82.61400
Weibull	*Shape*: 2.76070	*Scale*: 0.32540	-87.22986	-83.04117

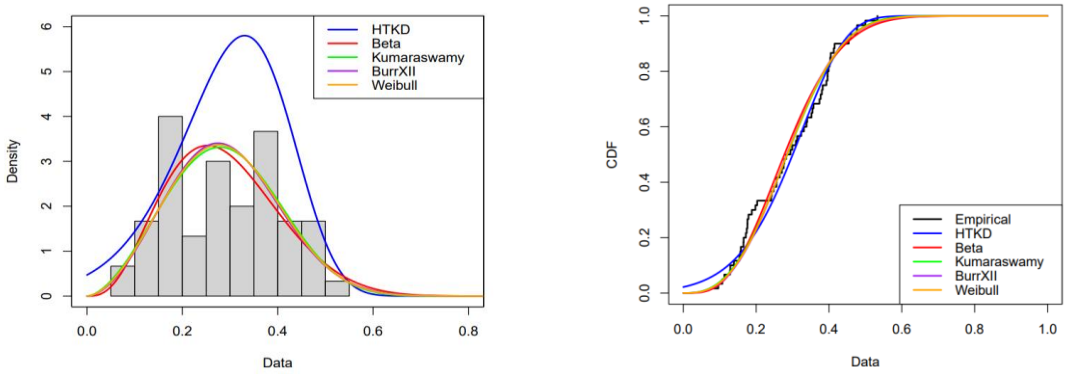

Figure 7: Plots of the histogram and fitted PDFs (left), empirical and fitted CDFs (right) of HTKD distribution for the UK's COVID-19 mortality rate

Table 8. Second Data Set: Canada's COVID-19 Mortality Rate

0.1622	0.1159	0.1897	0.1260	0.3025	0.2190	0.2075	0.2241	0.2163	0.1262
0.1627	0.2591	0.1989	0.3053	0.2170	0.2241	0.2174	0.2541	0.1997	0.3333
0.2594	0.2230	0.2290	0.1536	0.2024	0.2931	0.2739	0.2607	0.2736	0.2323
0.1563	0.2677	0.2181	0.3019	0.2136	0.2281	0.2346	0.1888	0.2729	0.2162
0.2746	0.2936	0.3259	0.2242	0.1810	0.2679	0.2296	0.2992	0.2464	0.2576
0.2338	0.1499	0.2075	0.1834	0.3347	0.2362				

Table 9: Model Comparison of Different Distributions for Canada's COVID-19 Mortality Rate Dataset

Distribution	Parameter 1	Parameter 2	AIC	BIC
HTKD	*a* : 22.42300	*b* : 12315.14098	-222.0884	-218.0377
Beta	*Shape1* : 14.50032	*Shape2* : 48.44678	-167.8800	-163.8293
Kumaraswamy	*Shape1* : 5.03087	*Shape2* : 1049.62460	-169.2000	-165.1493
BurrXII	*a* : 5.03599	*b* : 1058.64400	-169.2027	-165.1520
Weibull	*Shape*: 5.03354	*Scale*: 0.25087	-169.2014	-165.1507

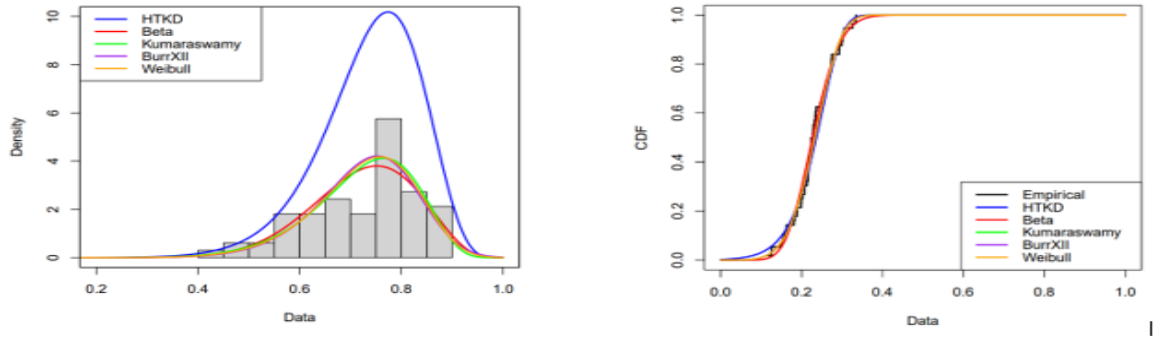

Figure 8: Plots of the histogram and fitted PDFs (left), empirical and fitted CDFs (right) of HTKD distribution for Canada's COVID-19 mortality rate dataset

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

Table 10: Third Data Set: Spain's COVID-19 Survival Rate

0.6670	0.5000	0.5000	0.4286	0.7500	0.6531	0.5161	0.7895	0.7689	0.6873
0.5200	0.7251	0.6375	0.6078	0.6289	0.5712	0.5923	0.6061	0.5924	0.5921
0.5592	0.5954	0.6164	0.6455	0.6725	0.6838	0.6850	0.6947	0.7210	0.7315
0.7412	0.7508	0.7519	0.7547	0.7645	0.7715	0.7759	0.7807	0.7838	0.7847
0.7871	0.7902	0.7934	0.7913	0.7962	0.7971	0.7977	0.8007	0.8038	0.8289
0.8322	0.8354	0.8371	0.8387	0.8456	0.8490	0.8535	0.8547	0.8564	0.8580
0.8604	0.8628	0.6586	0.7070	0.7963	0.8516				

Table 11: Model comparison of different distributions for Spain's COVID-19 survival rate dataset

Distribution	Parameter 1	Parameter 2	AIC	BIC
HTKD	*a* : 23.44384	*b* : 5.88810	-226.9411	-222.5618
Beta	*Shape1* : 12.79436	*Shape2* : 4.89945	-111.1486	-106.7692
Kumaraswamy	*Shape1* : 8.07822	*Shape2* : 7.73824	-113.6686	-109.2893
BurrXII	*a* : 9.05443	*b* : 11.51520	-109.4720	-105.0927
Weibull	*Shape*: 8.67669	*Scale*: 0.76840	-111.1991	-106.8198

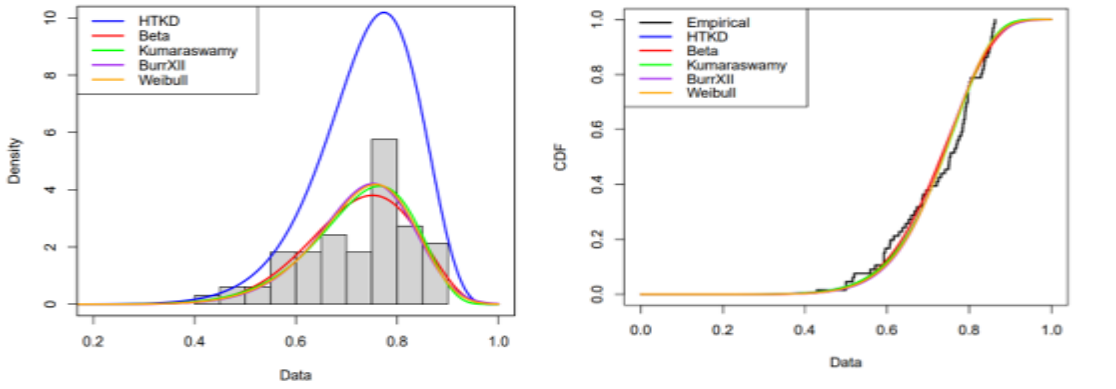

Figure 9: Plots of the histogram and fitted PDFs (left), empirical and fitted CDFs (right) of HTKD distribution for Spain's COVID-19 survival rate dataset

Discussion

An extensive examination of several distributions was carried out in order to ascertain which statistical distribution would be most suitable for the UK data on the COVID-19 mortality rate. Among all the distributed values evaluated, the HTKD distributionhas the lowest AIC and BIC values, which suggests that it is the most appropriate. With an AIC of -146.04328 and a BIC of -141.85459 (see Table 7), the HTKD distribution performs better than the Beta, Kumaraswamy, BurrXII, and Weibull distributions. These results imply that the HTKD distribution provides the most accurate model for the UK's COVID-19 mortality rate dataset.

Graphical representations in Figure 7 further support this conclusion. The HTKD distribution successfully fits the dataset, as seen by the fitted PDF, which closely reflects the shape of the histogram, and the fitted CDF, which agrees well with the empirical CDF. This close agreement demonstrates the accuracy and dependability of the HTKD distribution in simulating the COVID-19 death rate in the UK.

Comparing the HTKD distribution with other distributions using the COVID-19 mortality rate dataset from Canada revealed similar results. With the lowest AIC and BIC values of -222.0884 and -218.0377 (see Table 9), respectively, the HTKD distribution scored better than the other models, including the Beta, Kumaraswamy, BurrXII, and Weibull distributions. It may be inferred from this that the HTKD distribution fits the data from Canada the best. The better match of the HTKD distribution for Canada's COVID-19 mortality rate dataset is further supported by further graphical evidence. The empirical and fitted curves near agreement demonstrate the model's precision and potency in capturing the data.

The HTKD distribution has the lowest AIC (-226.9411) and BIC (-222.5618) values among the distribution fits tested for the COVID-19 survival rate dataset for Spain. The flexibility and durability of the HTKD distribution are shown by its consistent performance across a variety of datasets, which makes it a dependable

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63

option for modelling epidemiological data. The graphical displays of the Spain dataset's histograms, fitted PDFs, empirical data, and fitted CDFs provide more evidence of the HTKD distribution's efficacy in predicting survival rates. The ability of the HTKD distribution to appropriately depict Spain's survival rates is validated by the excellent agreement between the fitted curves and the actual data.

When compared to other studied distributions, including Beta, Kumaraswamy, BurrXII, and Weibull, the HTKD distribution consistently offers the greatest match, according to an examination of COVID-19 mortality and survival rate statistics for the UK, Canada, and Spain. The lowest AIC and BIC values found for the HTKD distribution across all datasets, as well as the tight correspondence between the fitted HTKD curves and the real data in graphical representations, confirm this result. The HTKD distribution's outstanding modelling performance for these datasets demonstrates its promise as a reliable and adaptable tool for statistical analysis in epidemiological research. To further confirm the efficacy and generalisability of the HTKD distribution, future studies might investigate its application to different epidemiological datasets.

Conclusion

This work presented and investigated the HTKD distribution, a new flexible unit interval distribution. Numerous statistical characteristics of it were examined. The robustness and correctness of the parameter estimates were validated by Monte Carlo simulations. The HTKD distribution consistently offered the best fit when applied to COVID-19 mortality rates in the UK, Canada, and Spain, as evidenced by the lowest AIC and BIC values and close alignment of fitted curves with empirical data, in contrast to the Beta, Kumaraswamy, BurrXII, and Weibull distributions. The HTKD distribution's flexibility and strong performance across datasets highlighted its potential as a versatile tool for epidemiological analysis. Future research could extend its application to diverse epidemiological data and more complex data structures to enhance its utility in public health research.

References

Afere, B. A. (2021). A new family of hybrid classical polynomial kernels in density estimation. *International Journal of Research and Innovation in Applied Science (IJRIAS)*, *6*, 145-151.

Afere, B. A. E. (2024). On the fourth-order hybrid beta polynomial kernels in kernel density estimation. *Journal of the Nigerian Society of Physical Sciences*, *6*(1), 1631.

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 139-160.

Bowley, A. L. (1920). *Elements of statistics* (4th ed.). Charles Scribner's Sons.

Cardano, G. (1993). *Ars magna or The rules of algebra* (T. R. Witmer, Trans.). Dover Publications.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1-32. https://doi.org/10.18637/jss.v076.i01

Carrasco, J. M. F., Ferrari, S. L. P. & Cordeiro, G. M. (2010). A New Generalized Kumaraswamy Distribution, 1004.0911 (arxiv.org)

Congdon, P. D. (2019). *Bayesian hierarchical models: With applications in R.* taylorfrancis.com

Dewick, P. R. & Liu, S. (2022). Copula modelling to analyse financial data. *Journal of Financial Management, 15*(3), 104. https://doi.org/10.3390/jrfm15030104

Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799-815.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics, 41*(2), 173-188.

Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, *6*(1), 70-81.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, *46*(1–2), 19-88.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *Journal of Machine Learning Research*, *18*, 1-45.

Moore, D. S. (1985). *Statistics: Concepts and controversies* (3rd ed.). W. H. Freeman, New York.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Obulezi, O. J., Nwankwo, B. C., Nwankwo, M. P., Ifeanyi, A. C., Igbokwe, C. P., & Igwebudu, C. N. (2023). Modelling mortality rate of COVID-19 patients in the United Kingdom, Canada, and the survival rate of COVID-19 patients in Spain. *Journal of Xidian University*, *17*(11), 520-538. https://doi.org/10.37896/jxu17.11/048

Okoro, I., Uka, C. O., & Ogbara, C. O. (2023). Application of three probability distributions to justify central limit theorem. *African Journal of Mathematics and Statistics Studies*, *6*(4), 77-80.

Osatohanmwen, P., Oyegue, F. O., Ewere, F. & Ajibade, B. (2020). A New Family of Generalized Distributions on the Unit Interval: The T-Kumaraswamy Family of Distributions. *Journal of Data Science, 18*(4), 218.

Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers, 51*, 111-126.

Pearson, K. (1905). The problem of the random walk. *Nature, 72*, 294-296.

Romano, C. (2002). Applying copula function to risk management. *Capitalia, Italy*. http://www.icer.it/workshop/Romano

Stan Development Team. (2022). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1).

Cite this article as:
Afere, B. A. E. (2024). Hybrid-Epanechnikov Transformed Kumaraswamy Distribution: Applications to COVID-19 Mortality and Survival Analysis. *FNAS Journal of Applied Biological Sciences. 2*(1), 49-63