

Research Article

Bounded Odd Inverse Pareto Exponential Distribution: Properties, Estimation, and Regression

Suleman Nasiru , Abdul Ghaniyyu Abubakari , and Irene Dekomwine Angbing 

Department of Statistics, Faculty of Mathematical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

Correspondence should be addressed to Suleman Nasiru; sulemanstat@gmail.com

Received 25 March 2021; Accepted 3 May 2021; Published 26 May 2021

Academic Editor: Sergejs Solovjovs

Copyright © 2021 Suleman Nasiru et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we introduce a new three-parameter distribution defined on the unit interval. The density function of the distribution exhibits different kinds of shapes such as decreasing, increasing, left skewed, right skewed, and approximately symmetric. The failure rate function shows increasing, bathtub, and modified upside-down bathtub shapes. Six different frequentist estimation procedures were proposed for estimating the parameters of the distribution and their performance assessed via Monte Carlo simulations. Applications of the distribution were illustrated by analyzing two datasets and its fit compared to that of other distributions defined on the unit interval. Finally, we developed a regression model for a response variable that follows the new distribution.

1. Introduction

The development of distributions defined on the unit interval is increasingly gaining grounds in literature due to their usefulness in the areas of psychology, economics, biology, and engineering among others. These distributions are useful for modeling data that are defined on the unit interval such as proportions, percentages, or rates. In psychology for instance, proportions and percentages play a critical role in assessing the probability of judgments, the proportion of the brain's volume occupied by a specific part of the brain, and the proportion of a period of time spent on an activity [1]. In economics, there are many instances where data are bounded on the unit interval, for example, proportion of income spent on nondurable consumption, pension plan participation rates, market shares, fractional repayment on debts, and capital structures [1–3].

Distributions defined on the unit interval are known to have desirable failure (hazard) rate characteristics such as increasing, decreasing, and bathtub shapes. These failure rate characteristics are vital when modeling datasets. For instance, Rajarshi and Rajarshi [4] and Lawless [5] indicated

different scenarios where distributions with bathtub hazard rates are needed to model lifetime of electronic, electrochemical, and mechanical products. Lai [6] reported that the optimum number of minimal repairs for systems have increasing failure rates. Also, Woosley and Cossman [7] revealed that drugs have increasing failure rate during clinical development.

Although the two-parameter beta distribution [8] is one of the oldest distributions for modeling dataset on the unit interval, its cumulative distribution and quantile functions are not tractable. This makes generation of random observations for simulation from the beta distribution a bit complex. Hence, many researchers aim at developing bounded distributions with tractable cumulative distribution and quantile functions. Some of the existing bounded distributions in the literature include bounded M-O extended exponential distribution [2], unit Gompertz distribution [9], unit gamma distribution [10], Kumaraswamy distribution [11], Topp–Leone distribution [12], unit Burr III distribution [13], unit Weibull distribution [14], unit Lindley distribution [15], log-extended exponential-geometric distribution [16], logit slash distribution [17], unit Burr XII

distribution [18], Arcsecant hyperbolic normal distribution [19], unit Johnson S_U distribution [20], and unit inverse Gaussian distribution [21].

Despite the existence of some bounded distributions in the literature, no single distribution can be considered as the best for modeling all kinds of datasets. We are therefore motivated to develop a new bounded distribution with tractable cumulative distribution and quantile functions called bounded odd inverse Pareto exponential (BOIPE) distribution for modeling datasets on unit intervals. The BOIPE distribution is developed using the transformation $Y = e^{-X}$, where X follows the odd inverse Pareto exponential (OIPE) distribution [22]. The new distribution hubs other existing distributions such as the Kumaraswamy, bounded M-O extended exponential, and power function distributions as submodels.

The remainder of the article is organized as follows: Sections 2 and 3 present the BOIPE distribution and its statistical properties, respectively. In Section 4, different frequentist estimation techniques are discussed. In Section 5, Monte Carlo simulations are carried out to examine their performance of the estimators. In Section 6, the applications of the BOIPE distribution are demonstrated. In Section 7, a regression model is proposed. Finally, the conclusion of the study is presented in Section 8.

2. BOIPE Distribution

Let the random variable X follow the OIPE distribution with probability density function (PDF) given by

$$f(x) = \frac{\alpha\beta\lambda e^{-\lambda x} (1 - e^{-\lambda x})^{\alpha-1}}{(1 - (1 - \beta)e^{-\lambda x})^{\alpha+1}}, \quad \alpha > 0, \beta > 0, \lambda > 0, x > 0. \quad (1)$$

Then, the distribution of the random variable $Y = e^{-X}$ is the BOIPE distribution. The PDF, cumulative distribution function (CDF), and hazard rate of the BOIPE distribution are respectively given by

$$f(y) = \frac{\alpha\beta\lambda y^{\lambda-1} (1 - y^\lambda)^{\alpha-1}}{(1 - (1 - \beta)y^\lambda)^{\alpha+1}}, \quad \alpha > 0, \beta > 0, \lambda > 0, 0 < y < 1, \quad (2)$$

$$F(y) = 1 - \frac{(1 - y^\lambda)^\alpha}{(1 - (1 - \beta)y^\lambda)^\alpha}, \quad 0 < y < 1, \quad (3)$$

$$\tau(y) = \frac{\alpha\beta\lambda y^{\lambda-1}}{(1 - y^\lambda)(1 - (1 - \beta)y^\lambda)}, \quad 0 < y < 1. \quad (4)$$

The BOIPE distribution generalizes some existing distributions defined on the $(0, 1)$ support. These are the Kumaraswamy distribution for $\beta = 1$, the bounded M-O extended exponential (BMOEE) distribution for $\alpha = 1$, and the power function (PF) distribution for $\alpha = \beta = 1$. Figure 1 shows the relationship between the BOIPE distribution and its submodels.

The PDF and hazard rate function of the BOIPE distribution exhibit different kinds of shapes as shown in Figure 2. The PDF exhibits left skewed, right skewed, symmetric, J shape, and reversed J shape for the given parameter values. The hazard rate function displays increasing, bathtub, and modified upside-down bathtub shapes. The R codes for PDF and hazard rate function can be found in the Appendix section.

The limiting behavior of the density and hazard rate functions as $y \rightarrow 0$ and $y \rightarrow 1$ are respectively given by

$$\lim_{y \rightarrow 0} f(y) = \lim_{y \rightarrow 0} \tau(y) = \begin{cases} 0, & \lambda > 1, \\ \alpha\beta, & \lambda = 1, \\ \infty, & \lambda < 1, \end{cases} \quad (5)$$

$$\lim_{y \rightarrow 1} f(y) = \lim_{y \rightarrow 1} \tau(y) = 0.$$

Sometimes, to derive the statistical properties of a developed distribution, the expansion of the density function is required. Using the generalized binomial expansion,

$$(1 - z)^{-a} = \sum_{k=0}^{\infty} \binom{a+k-1}{k} z^k, \quad |z| \leq 1, \quad (6)$$

the density function can be written as

$$f(y) = \begin{cases} \alpha\beta\lambda \sum_{i=0}^{\infty} \binom{\alpha+i}{i} (1-\beta)^i y^{\lambda(i+1)-1} (1-y^\lambda)^{\alpha-1}, & \beta \in (0, 1), \\ \alpha\lambda \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \beta^{-(\alpha+i)} (\beta-1)^i y^{\lambda-1} (1-y^\lambda)^{\alpha+i-1}, & \beta > 1. \end{cases} \quad (7)$$

Proposition 1. Let Y_1 be the BOIPE random variable with CDF $F(y; \alpha_1, \beta, \lambda)$ and Y_2 be the BOIPE random variable

with CDF $F(y; \alpha_2, \beta, \lambda)$. Then, the BOIPE distribution is identifiable if $\alpha_1 = \alpha_2$.

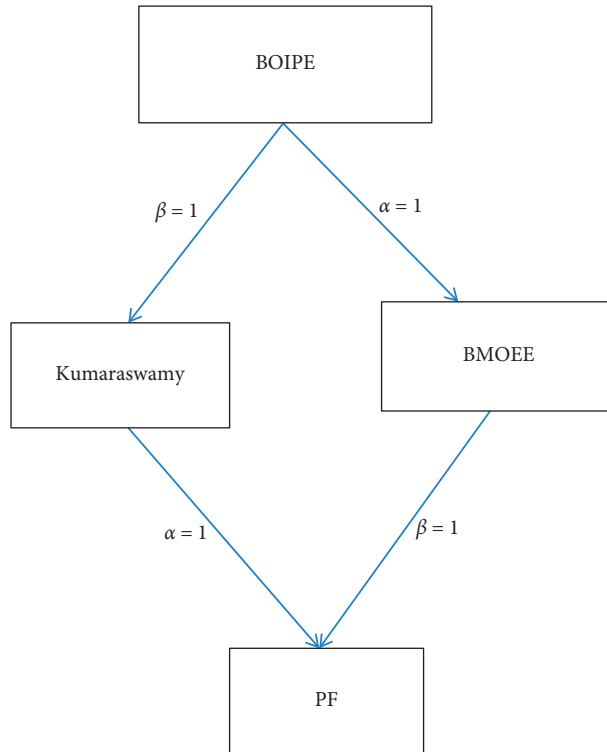


FIGURE 1: Submodels of BOIPE distribution.

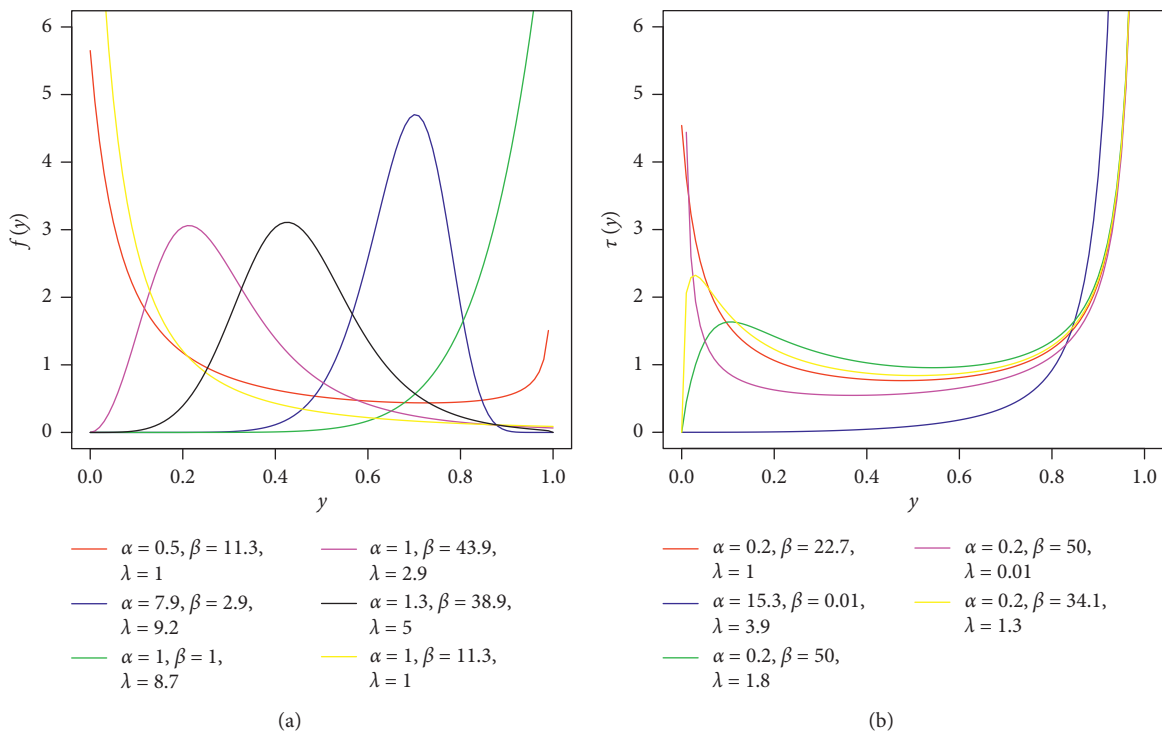


FIGURE 2: Plots of PDF (a) and hazard function (b) of BOIPE distribution.

Proof. For the BOIPE distribution to be identifiable, $F(y; \alpha_1, \beta, \lambda) = F(y; \alpha_2, \beta, \lambda)$, when $\alpha_1 = \alpha_2$. Hence,

$$1 - \frac{(1 - y^\lambda)^{\alpha_1}}{(1 - (1 - \beta)y^\lambda)^{\alpha_1}} = 1 - \frac{(1 - y^\lambda)^{\alpha_2}}{(1 - (1 - \beta)y^\lambda)^{\alpha_2}}. \tag{8}$$

This implies that

$$\frac{(1 - y^\lambda)^{\alpha_1}}{(1 - (1 - \beta)y^\lambda)^{\alpha_1}} = \frac{(1 - y^\lambda)^{\alpha_2}}{(1 - (1 - \beta)y^\lambda)^{\alpha_2}}. \tag{9}$$

When $\alpha_1 = \alpha_2$, $F(y; \alpha_1, \beta, \lambda) = F(y; \alpha_2, \beta, \lambda)$. This completes the proof. \square

3. Statistical Properties

The statistical properties of the BOIPE distribution are presented in this section.

3.1. Quantile. The quantile function is useful when generating random observations from a distribution. It can also be utilized in estimating measures of shapes (skewness and kurtosis) when the moments of the random variable do not exist. The quantile function of the BOIPE distribution is

$$Q_Y(u) = \frac{(1 - (1 - u)^{1/\lambda})^{1/\lambda}}{(1 - (1 - \beta)(1 - u)^{1/\lambda})^{1/\lambda}}, \quad u \in (0, 1). \tag{10}$$

The first quartile, the median, and the upper quartile are obtained by substituting $u = 0.25, 0.5$ and 0.75 , respectively, into equation (10). The quantile function can be used to generate random observations from the BOIPE distribution.

The algorithm for generating random observation from the BOIPE distribution is as follows:

- (i) Generate $U_i \sim \text{uniform}(0, 1), i = 1, \dots, n$;
- (ii) Set $y_i = ((1 - (1 - u)^{1/\lambda})^{1/\lambda}) / ((1 - (1 - \beta)(1 - u)^{1/\lambda})^{1/\lambda})$.

The following R codes can be used to generate random observations from the distribution.

```

quantile <- function(n, alpha, beta, lambda){
  u <- if(n, 0, 1)
  P <- -(1-(1-u)(1/alpha))(1/lambda)
  Z <- -(1-((1-beta)*((1-u)(1/alpha))))(1/lambda)
  y <- P/Z
  return(y).
}
    
```

The histograms of 1000 random observations generated from the BOIPE distribution using different parameter values are shown in Figure 3. The histograms of the random observations show that the distribution can exhibit different degrees of skewness.

3.2. Moments and Incomplete Moments. The moments of a random variable, if they exist, are useful for estimating measures of central tendency, dispersion, and shapes. For the BOIPE random variable, the n^{th} noncentral moment is given by

$$\mu'_n = \int_0^1 y^n f(y) dy. \tag{11}$$

Thus, using the expanded form of the density function yields

$$\mu'_n = \begin{cases} \alpha\beta\lambda \sum_{i=0}^{\infty} \binom{\alpha+i}{i} (1-\beta)^i \int_0^1 y^{n+\lambda(i+1)-1} (1-y^\lambda)^{\alpha-1} dy, & \beta \in (0, 1), \\ \alpha\lambda \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \beta^{-(\alpha+i)} (\beta-1)^i \int_0^1 y^{n+\lambda-1} (1-y^\lambda)^{\alpha+i-1} dy, & \beta > 1, \end{cases} \tag{12}$$

If we let $u = y^\lambda$, then as $y \rightarrow 0, u \rightarrow 0$ and as $y \rightarrow 1, u \rightarrow 1$. Further, $dy = du/\lambda y^{\lambda-1}$. Hence, after some algebraic manipulations, we have

$$\mu'_n = \begin{cases} \alpha\beta \sum_{i=0}^{\infty} \binom{\alpha+i}{i} (1-\beta)^i B\left(\frac{n+\lambda(i+1)}{\lambda}, \alpha\right), & \beta \in (0, 1), \\ \alpha \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \beta^{-(\alpha+i)} (\beta-1)^i B\left(\frac{n+\lambda}{\lambda}, \alpha+i\right), & \beta > 1, \end{cases} \tag{13}$$

where $B(a, b) = \int_0^1 z^{a-1} (1-z)^{b-1} dz$ is the beta function and $n = 1, 2, \dots$. The central moments (μ_s) and the cumulants (κ_s) can be obtained from the noncentral moments as $\mu_s = \sum_{j=0}^s (-1)^j \binom{s}{j} \mu'_1{}^s \mu'_{s-j}$ and $\kappa_s = \mu'_s - \sum_{j=1}^{s-1} \binom{s-1}{j-1} \kappa_j \mu'_{s-j}$, respectively, where $\kappa_1 = \mu'_1$. The skewness and kurtosis are respectively calculated from the third and fourth standardized cumulants as $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and $\gamma_2 = \kappa_4/\kappa_2^2$. Table 1 displays the first six moments, standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), and coefficient of kurtosis (CK). The values for SD, CV, CS, and CK are computed respectively using

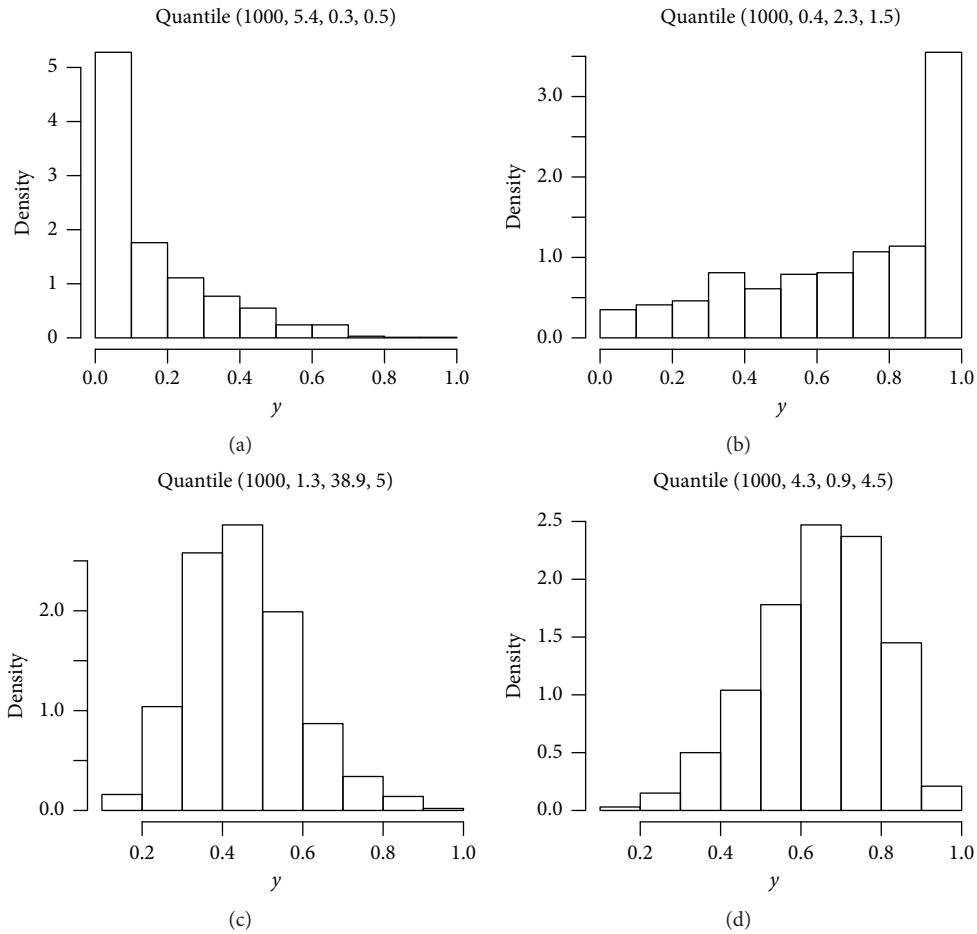


FIGURE 3: Histograms of simulated data from BOIPE distribution.

TABLE 1: First six moments, SD, CV, CS, and CK.

μ_r'	$\alpha = 0.4, \beta = 2.5, \lambda = 0.1$	$\alpha = 4.5, \beta = 3.1, \lambda = 1.5$	$\alpha = 0.3, \beta = 0.4, \lambda = 0.8$
μ_1'	0.2459	0.1610	0.8281
μ_2'	0.1860	0.0386	0.7452
μ_3'	0.1580	0.0119	0.6911
μ_4'	0.1408	0.0044	0.6513
μ_5'	0.1288	0.0019	0.6202
μ_6'	0.1198	0.0009	0.5948
SD	0.3543	0.1125	0.2439
CV	1.4406	0.6987	0.2945
SK	1.1365	1.1369	-1.6888
CK	2.6565	4.5168	4.9744

$$\begin{aligned}
 \text{SD} &= \sqrt{\mu_2' - \mu^2}, \\
 \text{CV} &= \frac{\sigma}{\mu} = \sqrt{\frac{\mu_2'}{\mu^2} - 1}, \\
 \text{CS} &= \frac{\mu_3' - 3\mu\mu_2' + 2\mu^3}{(\mu_2' - \mu^2)^{3/2}}, \\
 \text{CK} &= \frac{\mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4}{(\mu_2' - \mu^2)^2}.
 \end{aligned}
 \tag{14}$$

The incomplete moments are important when estimating measures of inequalities such as the Lorenz and Bonferroni curves and measures of deviations such as the mean and median deviations. The incomplete n^{th} moment for the BOIPE distribution is defined as

$$\varphi_n(y) = \int_0^y x^n f(x) dx.
 \tag{15}$$

Substituting the expanded form of the density function into the definition of the incomplete moments and simplifying yields

$$\varphi_n = \begin{cases} \alpha\beta \sum_{i=0}^{\infty} \binom{\alpha+i}{i} (1-\beta)^i B\left(y^\lambda; \frac{n+\lambda(i+1)}{\lambda}, \alpha\right), & \beta \in (0, 1) \\ \alpha \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \beta^{-(\alpha+i)} (\beta-1)^i B\left(y^\lambda; \frac{n+\lambda}{\lambda}, \alpha+i\right), & \beta > 1, \end{cases} \quad (16)$$

where $B(q; a, b) = \int_0^q z^{a-1} (1-z)^{b-1} dz$ is the incomplete beta function.

3.3. *Generating Functions.* The moment generating, characteristic, and cumulant generating functions are derived in

this section. The moment generating function, if it exists, is given by $M_Y(t) = E(e^{tY})$. Hence, employing Taylor series expansion, the moment generating function of the BOIPE random variable Y is given by

$$M_Y(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mu'_n = \begin{cases} \alpha\beta \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \frac{t^n (1-\beta)^i}{n!} B\left(\frac{n+\lambda(i+1)}{\lambda}, \alpha\right), & \beta \in (0, 1), \\ \alpha \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \frac{t^n \beta^{-(\alpha+i)} (\beta-1)^i}{n!} B\left(\frac{n+\lambda}{\lambda}, \alpha+i\right), & \beta > 1. \end{cases} \quad (17)$$

The characteristic function is defined as $\phi_Y(t) = E(e^{itY}) = \sum_{n=0}^{\infty} ((it)^n/n!) \mu'_n, i = \sqrt{-1}$. Thus, the characteristic function is given by

$$\phi_Y(t) = \begin{cases} \alpha\beta \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \frac{(it)^n (1-\beta)^i}{n!} B\left(\frac{n+\lambda(i+1)}{\lambda}, \alpha\right), & \beta \in (0, 1), \\ \alpha \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \frac{(it)^n \beta^{-(\alpha+i)} (\beta-1)^i}{n!} B\left(\frac{n+\lambda}{\lambda}, \alpha+i\right), & \beta > 1. \end{cases} \quad (18)$$

The cumulant generating function of Y , $\kappa_Y(t) = \log \phi_Y(t)$. Hence,

$$\kappa_Y(t) = \begin{cases} \log(\alpha\beta) + \log \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \frac{(it)^n (1-\beta)^i}{n!} B\left(\frac{n+\lambda(i+1)}{\lambda}, \alpha\right), & \beta \in (0, 1), \\ \log(\alpha) + \log \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \binom{\alpha+i}{i} \frac{(it)^n \beta^{-(\alpha+i)} (\beta-1)^i}{n!} B\left(\frac{n+\lambda}{\lambda}, \alpha+i\right), & \beta > 1. \end{cases} \quad (19)$$

3.4. *Entropies.* Entropies are useful measures of variation of a random variable. They have been extensively used in the areas of physics, molecular imaging of tumours, and sparse kernel density estimation. In this subsection, the Rényi [23] and δ entropies are discussed. The Rényi entropy of a random variable Y having the BOIPE distribution is defined as

$$H_\delta(y) = \frac{1}{1-\delta} \log \left[\int_0^1 f^\delta(y) dy \right], \quad \delta > 0, \delta \neq 1. \quad (20)$$

Using the generalized binomial expansion, we have

$$H_\delta(y) = \begin{cases} \frac{1}{1-\delta} \log \left[(\alpha\beta\lambda)^\delta \sum_{i=0}^{\infty} \binom{\delta(\alpha+1)+i-1}{i} (1-\beta)^i \int_0^1 y^{\lambda(\delta+i)-\delta} (1-y^\lambda)^{\delta(\alpha-1)} dy \right], & \beta \in (0, 1), \\ \frac{1}{1-\delta} \log \left[(\alpha\lambda)^\delta \sum_{i=0}^{\infty} \binom{\delta(\alpha+1)+i-1}{i} \beta^{-(\delta\alpha+i)} (\beta-1)^i \int_0^1 y^{\delta(\lambda-1)} (1-y^\lambda)^{\delta(\alpha-1)+i} dy \right], & \beta > 1. \end{cases} \quad (21)$$

Letting $u = y^\lambda$, as $y \rightarrow 0$, $u \rightarrow 0$ and as $y \rightarrow 1$, $u \rightarrow 1$. Also, $dy = du/\lambda y^{\lambda-1}$. Thus, after some algebraic manipulations, the Rényi entropy is obtained as

$$H_\delta(y) = \begin{cases} \frac{1}{1-\delta} \log \left[(\alpha\beta)^\delta \lambda^{\delta-1} \sum_{i=0}^{\infty} \binom{\delta(\alpha+1)+i-1}{i} (1-\beta)^i B \left(\frac{\lambda(\delta-i)-\delta-1}{\lambda}, \delta(\alpha-1)+1 \right) \right], & \beta \in (0, 1) \\ \frac{1}{1-\delta} \log \left[\alpha^\delta \lambda^{\delta-1} \sum_{i=0}^{\infty} \binom{\delta(\alpha+1)+i-1}{i} \beta^{-(\delta\alpha+i)} (\beta-1)^i B \left(\frac{\delta(\lambda-1)+1}{\lambda}, \delta(\alpha-1)+i+1 \right) \right], & \beta > 1. \end{cases} \quad (22)$$

The δ -entropy is defined as

$$I_\delta(y) = \frac{1}{\delta-1} \log \left[1 - \int_0^1 f^\delta(y) dy \right], \quad \delta > 0, \delta \neq 1, \quad (23)$$

and then it follows from equation (22).

3.5. *Stochastic Ordering.* Stochastic ordering is used to examine comparative behavior in reliability theory and other fields. Suppose Y_1 and Y_2 are two continuous random variables with PDFs $f(y)$ and $g(y)$, respectively. If $f(y)/g(y)$ is nondecreasing, then the random variable Y_2 is smaller than Y_1 in likelihood ratio order denoted as $Y_2 \leq_{lr} Y_1$. The likelihood ratio order is stronger than the hazard rate order and the usual stochastic order, which are defined as follows:

- (i) Y_2 is said to be stochastically smaller than Y_1 , denoted by $Y_2 \leq_{st} Y_1$ if $F(y) \leq G(y)$ for all y . $F(y)$ and $G(y)$ are the CDFs of Y_1 and Y_2 , respectively.
- (ii) Y_2 is said to be smaller than Y_1 in hazard rate order denoted by $Y_2 \leq_{hr} Y_1$ if $\tau_2(y) \leq \tau_1(y)$ for all y . $\tau_1(y)$ and $\tau_2(y)$ are the CDFs of Y_1 and Y_2 , respectively.

Proposition 2. Let Y_1 and Y_2 be two random variables having the BOIPE distribution with parameters $(\alpha_1, \beta, \lambda)$ and $(\alpha_2, \beta, \lambda)$, respectively. If $\alpha_2 > \alpha_1$, then $Y_2 \leq_{lr} Y_1$.

Proof. The ratio of the densities of the random variables is

$$\frac{f(y)}{g(y)} = \left(\frac{\alpha_1}{\alpha_2} \right) (1-y^\lambda)^{\alpha_1-\alpha_2} (1-(1-\beta)y^\lambda)^{\alpha_2-\alpha_1}. \quad (24)$$

Next,

$$\frac{d \log \left(\frac{f(y)}{g(y)} \right)}{dy} = \frac{\lambda(\alpha_2 - \alpha_1)y^{\lambda-1}}{1-y^\lambda} + \frac{\lambda(\alpha_2 - \alpha_1)(\beta-1)y^{\lambda-1}}{1+(\beta-1)y^\lambda}. \quad (25)$$

Hence, if $\alpha_2 > \alpha_1$, then $(d \log / dy)(f(y)/g(y)) > 0$ for all y . This implies that $f(y)/g(y)$ is nondecreasing in y and thus $Y_2 \leq_{lr} Y_1$. It is worth noting that $Y_2 \leq_{lr} Y_1 \Rightarrow Y_2 \leq_{hr} Y_1 \Rightarrow Y_2 \leq_{st} Y_1$. \square

3.6. *Order Statistics.* Order statistics are important for estimating summary statistics such as the minimum, maximum, and range of a dataset. They are also used in quality control testing and reliability to forecast failure of future

items based on the times of few early failures. Given that $Y_{(1:n)} \leq \dots \leq Y_{(n:n)}$ are order statistics of a random sample Y_1, \dots, Y_n from the BOIPE distribution, then the PDF of the k^{th} order statistic, $Y_{(k:n)}$, is given by

$$f_{k:n}(y) = \frac{n!f(y)}{(k-1)!(n-k)!} [F(y)]^{k-1} [1-F(y)]^{n-k}. \quad (26)$$

Using the binomial expansion, the PDF of the k^{th} order statistic can be written as

$$f_{k:n}(y) = \frac{n!f(y)}{(k-1)!(n-k)!} \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} [1-F(y)]^{n+i-k}. \quad (27)$$

Substituting the PDF and CDF of the BOIPE distribution defined in equations (2) and (3), respectively, we have

$$f_{k:n}(y) = \frac{n!}{(k-1)!(n-k)!} \sum_{i=0}^{k-1} v_i f(y; \alpha(n+i-k), \lambda, \beta), \quad (28)$$

where $v_i = ((-1)^i / (n+i-k)) \binom{k-1}{i}$ and $f(y; \alpha(n+i-k), \lambda, \beta)$ is the PDF of the BOIPE distribution with parameters $\alpha(n+i-k), \lambda$, and β .

4. Parameter Estimation

The methods for estimating the parameters of the BOIPE distribution are presented in this section. These include maximum likelihood, ordinary least-squares (OLS), maximum product spacing (MPS), Cramér-von Mises (CVM), Anderson-Darling (AD), and percentile (PC) methods.

4.1. Maximum Likelihood Method. If y_1, \dots, y_n are n independent and identically distributed observations from the BOIPE distribution and $\theta = (\alpha, \beta, \lambda)^T$, then the total log-likelihood function, $\ell = \ell(\theta)$, is given by

$$\begin{aligned} \ell &= n \log(\alpha\beta\lambda) + (\lambda-1) \sum_{i=1}^n \log(y_i) \\ &+ (\alpha-1) \sum_{i=1}^n \log(1-y_i^\lambda) - (\alpha+1) \sum_{i=1}^n \log(1-(1-\beta)y_i^\lambda). \end{aligned} \quad (29)$$

The maximum likelihood estimates (MLE) of parameters can be obtained by directly maximizing equation (29) using the *R* software or equating the following system of equations to zero and solving them simultaneously using numerical methods:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log(1-y_i^\lambda) - \sum_{i=1}^n \log(1-(1-\beta)y_i^\lambda), \\ \frac{\partial \ell}{\partial \beta} &= \frac{n}{\beta} - (\alpha+1) \sum_{i=1}^n \frac{y_i^\lambda}{1-(1-\beta)y_i^\lambda}, \\ \frac{\partial \ell}{\partial \lambda} &= \frac{n}{\lambda} + \sum_{i=1}^n \log(y_i) - (\alpha-1) \sum_{i=1}^n \frac{\log(y_i)y_i^\lambda}{1-y_i^\lambda} + (\alpha+1) \sum_{i=1}^n \frac{(1-\beta)\log(y_i)y_i^\lambda}{1-(1-\beta)y_i^\lambda}. \end{aligned} \quad (30)$$

When the regularity conditions are satisfied, the multivariate normal $N_3(0, J(\theta)^{-1})$ distribution, where $J(\theta)^{-1}$ is the observed information estimated at $\hat{\theta}$, can be utilized to estimate the approximate confidence intervals for the BOIPE distribution parameters.

4.2. Ordinary Least Squares. The OLS technique is an estimation procedure introduced by Swain et al. [24] for estimating the parameters of a model. Suppose $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ are ordered observations from the BOIPE distribution with CDF $F(y|\alpha, \beta, \lambda)$. The OLS estimates are obtained by minimizing

$$L(\alpha, \beta, \lambda) = \sum_{i=1}^n \left[F(y_{(i)} | \alpha, \beta, \lambda) - \frac{i}{n+1} \right]^2, \quad (31)$$

with respect to the parameters α, β , and λ .

4.3. Maximum Product Spacing. MPS technique was developed as an alternative method to the maximum likelihood approach using the Kullback-Leibler information measure [25–29]. Suppose the uniform spacing

$$D_i = F(y_{(i)}|\alpha, \beta, \lambda) - F(y_{(i-1)}|\alpha, \beta, \lambda), \quad (32)$$

where $F(y_{(0)}|\alpha, \beta, \lambda) = 0, F(y_{(n+1)}|\alpha, \beta, \lambda) = 1$ and $D_0(\alpha, \beta, \lambda) + D_1(\alpha, \beta, \lambda) + \dots + D_{n+1}(\alpha, \beta, \lambda) = 1$. The MPS estimates are obtained by maximizing the logarithm of the geometric mean of the spacing

$$M(\alpha, \beta, \lambda) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i(\alpha, \beta, \lambda), \quad (33)$$

with respect to α, β , and λ .

4.4. Percentile Method. The PC method is also another approach for estimating the parameters of a model [30, 31]. Suppose $u_i = 1/(n + 1)$ is an unbiased estimator of $F(y_{(i)}|\alpha, \beta, \lambda)$. The PC estimates of the BOIPE distribution parameters are obtained by minimizing

$$P(\alpha, \beta, \lambda) = \sum_{i=1}^n [y_{(i)} - Q_Y(u_i)]^2, \tag{34}$$

with respect to the parameters and $Q_Y(u_i)$ is given by equation (10).

4.5. Cramér–von Mises Method. The CVM estimation method is considered to have less bias than other minimum distance estimators [32]. The CVM estimates for the BOIPE distribution are obtained by minimizing

$$C(\alpha, \beta, \lambda) = \frac{1}{12n} + \sum_{i=1}^n \left[F(y_{(i)}|\alpha, \beta, \lambda) - \frac{2i - 1}{2n} \right]^2, \tag{35}$$

with respect to the parameters α, β , and λ .

4.6. Anderson–Darling Method. The AD estimator is another type of minimum distance estimators. The AD estimates of the BOIPE distribution are obtained by minimizing

$$A(\alpha, \beta, \lambda) = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\log F(y_{(i)}|\alpha, \beta, \lambda) + \log(1 - F(y_{(i)}|\alpha, \beta, \lambda))], \tag{36}$$

with respect to the parameters α, β , and λ .

5. Monte Carlo Simulation

In this section, the performance of the estimators for the parameters of the BOIPE distribution is examined via Monte Carlo simulations. The simulation exercise was carried out using two sets of parameter values, that is, $(\alpha, \beta, \lambda) = (0.8, 0.5, 0.2)$ and $(0.9, 3.8, 0.8)$. The sample sizes $n = 25, 75, 150, 200, 300$, and 500 were used to generate random observations from the BOIPE distribution using its quantile function. For each sample size, the experiment was replicated for $N = 5,000$ times and the average estimate (AE), absolute bias (AB), and mean square error (MSE) were estimated. The results, as shown in Tables 2 and 3, revealed that all the estimators are consistent. For the first case (Table 2), the maximum likelihood estimators tend to have the least MSEs compared to the other estimators. For the second case (Table 3), when the sample size was 25, all the estimators tend to over estimate the parameter β . However, as the sample size increases, the estimates tend to converge to the actual parameter value. Again, the maximum likelihood estimators had the smallest of the MSEs, as the sample size increases.

6. Applications

The empirical applications of the BOIPE distribution are illustrated in this section using two real datasets. The first dataset (data I) can be found in the study by Yousof et al. [33] and consists of transformed total milk production in the first birth of 107 cows from the SINDI race. The data are 0.4365, 0.4260, 0.5140, 0.6907, 0.7471, 0.2605, 0.6196, 0.8781, 0.4990, 0.6058, 0.6891, 0.5770, 0.5394, 0.1479, 0.2356, 0.6012, 0.1525, 0.5483, 0.6927, 0.7261, 0.3323, 0.0671, 0.2361, 0.4800, 0.5707, 0.7131, 0.5853, 0.6768, 0.5350, 0.4151, 0.6789, 0.4576, 0.3259, 0.2303, 0.7687, 0.4371, 0.3383, 0.6114, 0.3480, 0.4564, 0.7804, 0.3406, 0.4823, 0.5912, 0.5744, 0.5481, 0.1131, 0.7290, 0.0168,

0.5529, 0.4530, 0.3891, 0.4752, 0.3134, 0.3175, 0.1167, 0.6750, 0.5113, 0.5447, 0.4143, 0.5627, 0.5150, 0.0776, 0.3945, 0.4553, 0.4470, 0.5285, 0.5232, 0.6465, 0.0650, 0.8492, 0.8147, 0.3627, 0.3906, 0.4438, 0.4612, 0.3188, 0.2160, 0.6707, 0.6220, 0.5629, 0.4675, 0.6844, 0.3413, 0.4332, 0.0854, 0.3821, 0.4694, 0.3635, 0.4111, 0.5349, 0.3751, 0.1546, 0.4517, 0.2681, 0.4049, 0.5553, 0.5878, 0.4741, 0.3598, 0.7629, 0.5941, 0.6174, 0.6860, 0.0609, 0.6488, and 0.2747.

The second dataset (data II) was first reported by Dumonceaux and Antle [34] and comprises the maximum flood level (in millions of cubic feet per second) for the Susquehanna River at Harrisburg, Pennsylvania. Each data point is the maximum flood level for a four-year period. The first being, 0.654, for the period 1890–1893 and the last being, 0.265, for the period 1966–1969. The data are 0.654, 0.613, 0.315, 0.449, 0.297, 0.402, 0.379, 0.423, 0.379, 0.324, 0.269, 0.740, 0.418, 0.412, 0.494, 0.416, 0.338, 0.392, 0.484, and 0.265. The performance of the BOIPE distribution was compared to that of the beta (B), Kumaraswamy (K), bounded Marshall–Olkin extended exponential (BMOEE), and exponentiated Topp–Leone (ETL) [35] distributions using goodness-of-fit statistics such as the Akaike information criterion (AIC), corrected Akaike information criterion (AICc), -2ℓ , Anderson–Darling method (AD), and Cramér–von Mises (CVM) method. The p values of the AD and CVM statistics are given in the parentheses. The distribution with the smallest values of the goodness-of-fit statistics is considered the best for a given dataset. The R codes for the empirical illustration can be found in the appendix section. Table 4 presents the maximum likelihood estimates of the parameters of the fitted distributions with their corresponding standard errors in parentheses for data I.

The goodness-of-fit statistics for the fitted distribution for the first dataset are shown in Table 5. It can be seen that the BOIPE distribution provides the best fit to the dataset since it has the least values for all the goodness-of-fit statistics.

TABLE 2: Simulation results for $\alpha = 0.8, \beta = 0.5$, and $\lambda = 0.2$.

n	Estimate	Parameter	MLE	MPS	OLS	CVM	AD	PC
25	AE	α	1.0646	1.4949	0.9073	1.0274	0.9796	0.7282
		β	1.7111	1.7185	3.1877	3.2045	1.8115	113.9482
		λ	0.2654	0.3224	0.2515	0.2806	0.2385	0.5679
	AB	α	0.2646	0.6949	0.1073	0.2274	0.1796	0.0718
		β	1.2111	1.2185	2.6877	2.7045	1.3115	113.4482
		λ	0.0654	0.1224	0.0515	0.0806	0.0385	0.3679
	MSE	α	1.3050	3.7320	0.9636	1.0335	1.2921	0.5436
		β	7.9240	7.9016	21.4493	23.0297	10.3399	1543.1780
		λ	0.1748	0.2183	0.2294	0.2477	0.1741	0.5791
75	AE	α	0.8540	0.9189	0.8376	0.8684	0.8401	0.7710
		β	0.6806	0.6922	0.8020	0.8120	0.6755	2.4090
		λ	0.2215	0.2464	0.2162	0.2277	0.2142	0.3342
	AB	α	0.0540	0.1189	0.0376	0.0684	0.0401	0.0290
		β	0.1806	0.1922	0.3020	0.3120	0.1755	1.9090
		λ	0.0215	0.0464	0.0162	0.0277	0.0142	0.1342
	MSE	α	0.2571	0.3060	0.3348	0.3705	0.2449	0.2644
		β	0.6085	0.5965	0.9434	0.9430	0.6282	5.6751
		λ	0.0798	0.0911	0.1102	0.1138	0.0843	0.2673
150	AE	α	0.8253	0.8582	0.8237	0.8395	0.8233	0.7884
		β	0.5745	0.5915	0.5997	0.6141	0.5682	1.1278
		λ	0.21205	0.2273	0.2053	0.2111	0.2071	0.2673
	AB	α	0.0253	0.0582	0.0237	0.0395	0.0233	0.0116
		β	0.0745	0.0915	0.0997	0.1141	0.0682	0.6278
		λ	0.0120	0.0273	0.0053	0.0111	0.0071	0.0673
	MSE	α	0.1413	0.1638	0.1764	0.1932	0.1452	0.1740
		β	0.3009	0.3077	0.4330	0.4792	0.3239	1.7357
		λ	0.0540	0.0599	0.0744	0.0767	0.0590	0.1799
200	AE	α	0.8141	0.8375	0.8068	0.8169	0.8115	0.7773
		β	0.5507	0.5679	0.6021	0.6128	0.5573	1.0201
		λ	0.2071	0.2194	0.2076	0.2126	0.2055	0.2625
	AB	α	0.0141	0.0375	0.0068	0.0169	0.0115	0.0227
		β	0.0507	0.0679	0.1021	0.1128	0.0573	0.5201
		λ	0.00712	0.0194	0.0076	0.0126	0.0055	0.0625
	MSE	α	0.1121	0.1263	0.1447	0.1539	0.1206	0.1507
		β	0.2499	0.2495	0.3862	0.4133	0.2822	1.2590
		λ	0.0426	0.0482	0.0660	0.0700	0.0491	0.1639
300	AE	α	0.8094	0.8263	0.8070	0.8144	0.8071	0.7890
		β	0.5342	0.5481	0.5590	0.5619	0.5388	0.7746
		λ	0.2055	0.2146	0.2050	0.2074	0.2047	0.2331
	AB	α	0.0094	0.0263	0.0070	0.0144	0.0071	0.0110
		β	0.0342	0.0481	0.0590	0.0619	0.0388	0.2746
		λ	0.0055	0.0146	0.0050	0.0074	0.0047	0.0331
	MSE	α	0.0926	0.1017	0.1166	0.1212	0.0973	0.1227
		β	0.1929	0.2012	0.2839	0.2898	0.2133	0.7608
		λ	0.0357	0.0394	0.0536	0.0541	0.0417	0.1271
500	AE	α	0.8091	0.8196	0.8050	0.8109	0.8081	0.8007
		β	0.5135	0.5226	0.5285	0.5315	0.5174	0.6352
		λ	0.2026	0.2085	0.2021	0.2038	0.2019	0.2160
	AB	α	0.0091	0.0196	0.0058	0.0109	0.0081	0.0007
		β	0.0135	0.0226	0.0285	0.0315	0.0174	0.1352
		λ	0.0026	0.0085	0.0021	0.0038	0.0019	0.0160
	MSE	α	0.0682	0.0744	0.0863	0.0912	0.0764	0.0959
		β	0.1306	0.1377	0.1925	0.2043	0.1581	0.4831
		λ	0.0265	0.0279	0.0393	0.0417	0.0320	0.1022

Figure 4 displays the PDF and CDF plots of the fitted distributions for data I. The graph clearly shows that the BOIPE distribution provides a good fit to the dataset.

The probability-probability (P-P) plots of the fitted distributions for data I are shown in Figure 5. The plots again reveal that the BOIPE distribution fits the data well.

TABLE 3: Simulation results for $\alpha = 0.9, \beta = 3.8$, and $\lambda = 0.8$.

n	Estimate	Parameter	MLE	MPS	OLS	CVM	AD	PC
25	AE	α	1.3479	1.7895	1.2252	1.2601	1.2060	1.5544
		β	21.2341	128.6030	264.7218	137.1928	24.5256	11.7387
		λ	0.8930	0.9900	0.9103	0.9187	0.8812	0.8575
	AB	α	0.4479	0.8895	0.3252	0.3601	0.3060	0.6544
		β	17.4341	124.8030	260.9218	20.7255	1.3115	7.9387
		λ	0.0930	0.1900	0.1103	0.1187	0.0812	0.0575
	MSE	α	1.7616	2.7191	1.6121	1.4790	1.4195	3.9597
		β	117.5867	2766.3810	3432.5360	1500.9980	120.9186	14.5361
		λ	0.3314	0.4250	0.5048	0.4511	0.3397	0.3941
75	AE	α	0.9983	1.0709	1.0246	0.8684	0.9723	0.9287
		β	5.0327	5.6193	10.0174	0.8120	5.4761	8.4316
		λ	0.8198	0.8564	0.8211	0.2277	0.8218	0.8515
	AB	α	0.0983	0.1709	0.1246	0.1539	0.0723	0.0287
		β	1.2327	1.8193	6.2174	4.2008	1.6761	4.6316
		λ	0.0198	0.0564	0.0211	0.0330	0.0218	0.0515
	MSE	α	0.3547	0.5474	0.5977	0.7598	0.3726	0.3827
		β	5.0275	7.2441	83.5301	25.3746	6.8930	10.1803
		λ	0.1621	0.1832	0.2360	0.2205	0.1715	0.3147
150	AE	α	0.9370	0.9743	0.9393	0.9616	0.9328	0.9050
		β	4.4916	4.3804	5.1789	5.0559	4.5906	6.2644
		λ	0.8176	0.8297	0.8123	0.8149	0.8153	0.8359
	AB	α	0.0370	0.0743	0.0393	0.0616	0.0328	0.0050
		β	0.6916	0.5804	1.3789	1.2559	0.7906	2.4644
		λ	0.0176	0.0297	0.0123	0.0149	0.0153	0.0359
	MSE	α	0.2041	0.2180	0.2952	0.3085	0.2219	0.2440
		β	2.8083	2.7220	4.7300	5.0739	3.0367	6.5710
		λ	0.1158	0.1191	0.1478	0.1505	0.1229	0.2284
200	AE	α	0.9229	0.9602	0.9303	0.9446	0.9139	0.8958
		β	4.2587	4.1614	4.7214	4.7278	4.4483	5.9325
		λ	0.8106	0.8215	0.8041	0.8147	0.8134	0.8424
	AB	α	0.0229	0.0602	0.0303	0.0446	0.0139	0.0042
		β	0.4587	0.3614	0.9214	0.9278	0.6483	2.1325
		λ	0.0106	0.0215	0.0041	0.0147	0.0134	0.0424
	MSE	α	0.1560	0.1791	0.2324	0.2617	0.1635	0.2107
		β	0.1642	2.1522	3.9443	3.5360	2.5156	5.7578
		λ	0.0974	0.1023	0.1261	0.1295	0.1062	0.2088
300	AE	α	0.9128	0.9434	0.9178	0.9300	0.9100	0.9076
		β	4.1064	3.990	4.3935	4.2727	4.2022	4.7893
		λ	0.8091	0.8132	0.8035	0.8082	0.8096	0.8122
	AB	α	0.0128	0.0434	0.0178	0.0300	0.0100	0.0076
		β	0.3064	0.1990	0.5935	0.4727	0.4022	0.9893
		λ	0.0091	0.0132	0.0035	0.0082	0.0096	0.0122
	MSE	α	0.1215	0.1422	0.1874	0.1845	0.1358	0.1736
		β	1.6734	1.6217	2.5497	2.3692	1.8528	3.4530
		λ	0.0803	0.0818	0.1017	0.1013	0.0873	0.1630
500	AE	α	0.9141	0.9260	0.9140	0.9128	0.9102	0.9020
		β	3.9092	3.9358	4.0276	4.1266	3.9692	4.4265
		λ	0.8029	0.8110	0.8000	0.8087	0.8033	0.8128
	AB	α	0.0141	0.0260	0.0140	0.0128	0.0102	0.0020
		β	0.1092	0.1358	0.2276	0.3266	0.1692	0.6265
		λ	0.0029	0.0110	0.0002	0.0087	0.0033	0.0128
	MSE	α	0.0938	0.1001	0.1281	0.1288	0.0996	0.1205
		β	1.1206	1.1582	1.5660	1.6586	1.2290	2.4066
		λ	0.0605	0.0627	0.0765	0.0803	0.0629	0.1243

The maximum likelihood estimates for the parameters of the fitted distributions for data II are given in Table 6.

The goodness-of-fit statistics for the fitted distributions for the second dataset are given in Table 7. The results revealed that

the BOIPE distribution again provides the best fit to the second dataset as compared to the other competing distributions.

The PDF and CDF plots of the fitted distributions, shown in Figure 6, give a pictorial representation of how

TABLE 4: Maximum likelihood estimates for data I.

Model	Estimates		
BOIPE	$\hat{\alpha} = 46.8165$ (1.9334×10^{-4})	$\hat{\beta} = 0.0379$ (9.4107×10^{-3})	$\hat{\lambda} = 1.7085$ (2.5741×10^{-1})
B	$\hat{\alpha} = 2.4125$ (0.3145)	$\hat{\beta} = 2.8297$ (0.3744)	—
K	$\hat{\alpha} = 3.4363$ (0.5820)	$\hat{\lambda} = 2.1949$ (0.2224)	—
BMOEE	$\hat{\beta} = 6.3399$ (2.0703)	$\hat{\lambda} = 2.5683$ (0.3213)	—
ETL	$\hat{\alpha} = 2.4684$ (0.3016)	$\hat{\beta} = 1.3050$ (0.1792)	—

TABLE 5: Goodness-of-fit statistics for data I.

Model	-2ℓ	AIC	AICc	AD (<i>P</i> value)	CVM (<i>P</i> value)
BOIPE	-54.3796	-48.3796	-48.1466	0.6471 (0.6043)	0.1057 (0.5583)
B	-47.5545	-43.5545	-43.4391	1.3853 (0.2064)	0.2281 (0.2190)
K	-50.7894	-46.7894	-46.6740	1.0030 (0.3555)	0.1522 (0.3836)
BMOEE	-30.0969	-26.0969	-25.9815	2.5737 (0.0454)	0.3316 (0.1103)
ETL	-46.6856	-42.6856	-42.5702	1.5421 (0.1668)	0.2632 (0.1723)

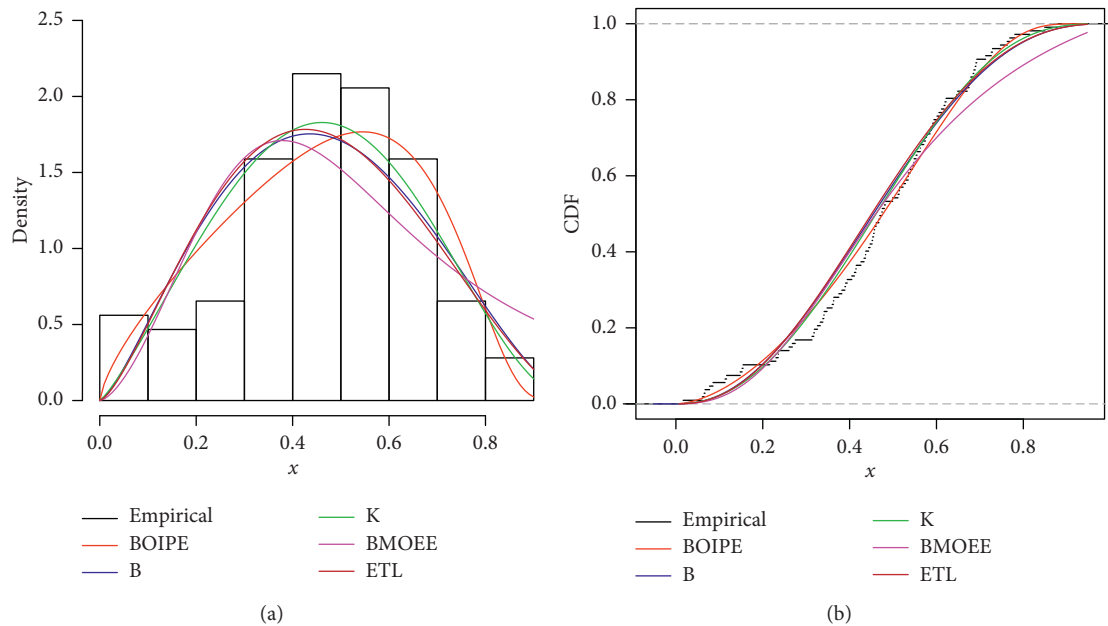


FIGURE 4: Plots of PDF (a) and CDF (b) of fitted distributions for data I.

well the distributions fit data II. It can be seen that the BOIPE distribution mimics the empirical distribution of the dataset.

The P-P plots shown in Figure 7 also revealed that the BOIPE distribution provides a good fit to data II compared to the other fitted distributions.

7. BOIPE Regression Model

Sometimes, one may be interested in investigating the effects of some exogenous variables on an endogenous variable and

a regression model may be required to accomplish this task. Thus, we proposed a new parametric regression model with assumption that the underlying distribution of the response variable follows the BOIPE distribution. In order to establish the regression model, we relate the parameters α and λ to exogenous variables by the logarithm link functions $\alpha_i = \exp(\mathbf{x}_i^T \alpha_1)$ and $\lambda_i = \exp(\mathbf{x}_i^T \lambda_1)$, $i = 1, \dots, n$, respectively, where $\alpha_1 = (\alpha_{10}, \dots, \alpha_{1p})^T$ and $\lambda_1 = (\lambda_{10}, \dots, \lambda_{1p})^T$ constitute the vectors of the regression coefficients and $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$. The survival function of $Y|X$ from equation (3) follows as

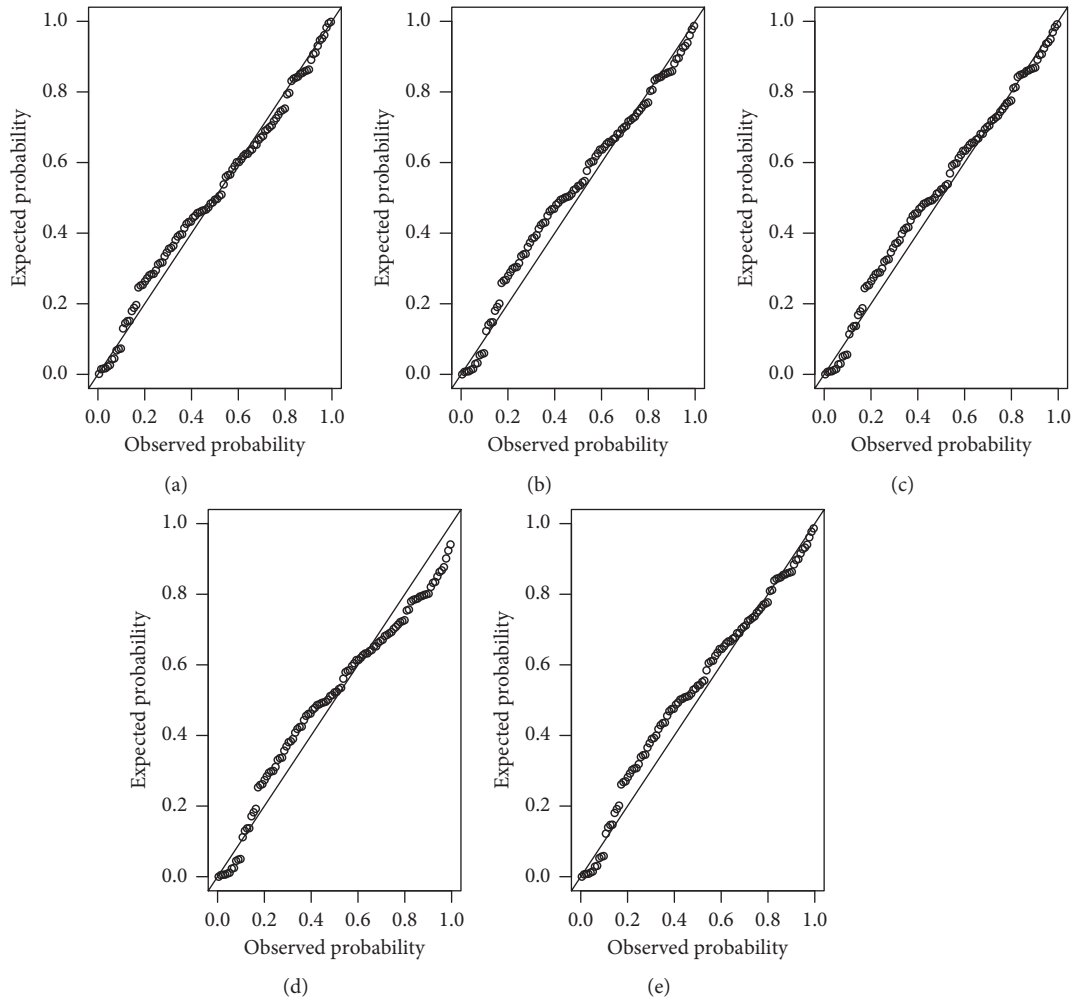


FIGURE 5: P-P plots of fitted distributions for data I: (a) BOIPE distribution. (b) B distribution. (c) K distribution. (d) BMOEE distribution. (e) ETL distribution.

TABLE 6: Maximum likelihood estimates for data II.

Model	Estimates		
BOIPE	$\hat{\alpha} = 0.5764$ (1.6068×10^{-1})	$\hat{\beta} = 5000.4970$ (8.4895×10^{-5})	$\hat{\lambda} = 8.3109$ (5.8128×10^{-1})
B	$\hat{\alpha} = 6.7594$ (2.0954)	$\hat{\beta} = 9.1141$ (2.8525)	—
K	$\hat{\alpha} = 11.7932$ (5.3616)	$\hat{\lambda} = 3.3639$ (0.6034)	—
BMOEE	$\hat{\beta} = 10250.1239$ (4.3889×10^{-5})	$\hat{\lambda} = 9.7920$ (4.4583×10^{-1})	—
ETL	$\hat{\alpha} = 4.6079$ (0.9498)	$\hat{\beta} = 4.0457$ (1.4757)	—

$$S(y|\mathbf{x}) = \frac{\left(1 - y^{\exp(\mathbf{x}^T \lambda_1)}\right)^{\exp(\mathbf{x}^T \alpha_1)}}{\left(1 - (1 - \beta)y^{\exp(\mathbf{x}^T \lambda_1)}\right)^{\exp(\mathbf{x}^T \alpha_1)}}. \tag{37}$$

To estimate the parameters of the regression model, the maximum likelihood technique was employed. The total log-likelihood function that needs to be maximized in order to obtain the estimates of the regression parameters is given by

TABLE 7: Goodness-of-fit statistics for data II.

Model	-2ℓ	AIC	AICc	AD (<i>P</i> value)	CVM (<i>P</i> value)
BOIPE	-32.0306	-26.0306	-24.5306	0.2437 (0.9735)	0.0387 (0.9438)
B	-28.1303	-24.1303	-23.4244	0.7329 (0.5301)	0.1237 (0.4844)
K	-25.7371	-21.7371	-21.0312	0.9326 (0.3934)	0.1637 (0.3525)
BMOEE	-25.7947	-21.7947	-21.0888	1.4356 (0.1927)	0.1271 (0.471)
ETL	-27.1904	-23.1904	-22.4846	0.8147 (0.4689)	0.1434 (0.4132)

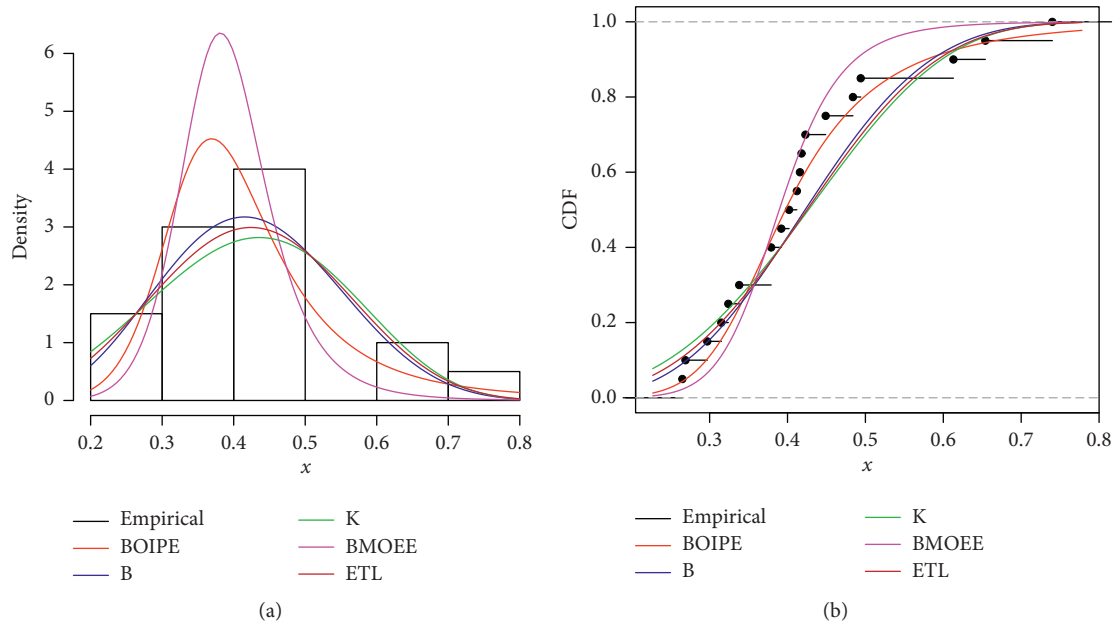


FIGURE 6: Plots of PDF (a) and CDF (b) of fitted distributions for data II.

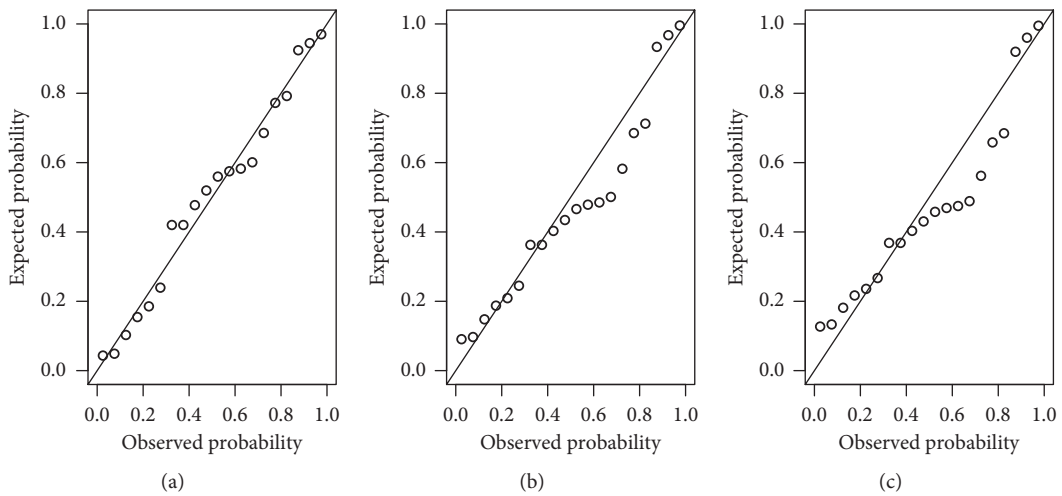


FIGURE 7: Continued.

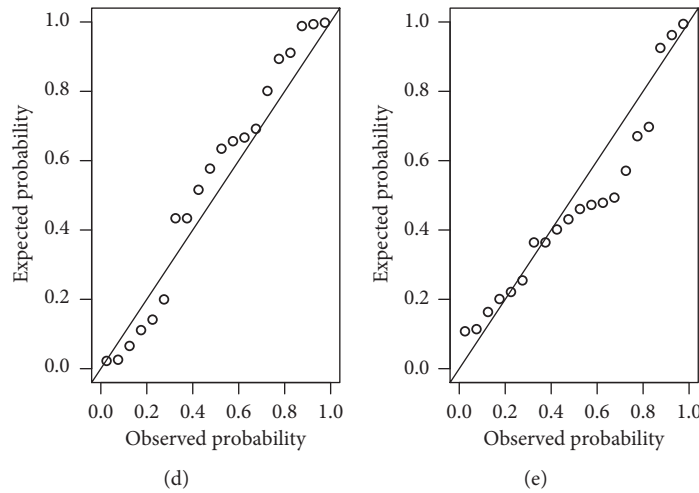


FIGURE 7: P-P plots of fitted distributions for data II: (a) BOIPE distribution. (b) B distribution. (c) K distribution. (d) BMOEE distribution. (e) ETL distribution.

TABLE 8: Estimates of regression parameters and goodness-of-fit statistics.

Model	Parameters	Estimates	P value	Goodness-of-fit
BOIPE	β	100.1200 (0.0019)	$<2.2000 \times 10^{-6}$	$-\ell = -106.5700$
	α_{10}	0.4675 (0.4754)	0.3253	AIC = -203.1476
	α_{11}	2.6183 (0.9544)	0.0061	AICc = -201.0047
	λ_{10}	0.3315 (0.0982)	0.0007	—
	λ_{11}	0.1945 (0.0705)	0.0058	—
Beta	β_0	-3.7591 (0.1686)	<0.0010	$-\ell = -103.8000$
	β_1	-0.3716 (0.1722)	0.0309	AIC = -201.6000
	ψ	71.2000 (18.8100)	<0.0010	AICc = -200.8000
Simplex	β_0	-3.7986 (0.3286)	<0.0010	$-\ell = -94.3895$
	β_1	-0.4421 (0.1213)	<0.0010	AIC = -182.7790
	ψ	2.5375 (0.1988)	0.0262	AICc = -181.9790

$$\begin{aligned}
 \ell = n \log(\beta) + \sum_{i=1}^n \log(\alpha_i \lambda_i) + \sum_{i=1}^n (\lambda_i - 1) \log(y_i) \\
 + \sum_{i=1}^n (\alpha_i - 1) \log(1 - y_i^{\lambda_i}) \\
 - \sum_{i=1}^n (\alpha_i + 1) \log(1 - (1 - \beta) y_i^{\lambda_i}).
 \end{aligned}
 \tag{38}$$

We demonstrated the application of the BOIPE regression by modeling the relationship between long-term interest (LTI) rates of the Organization for Economic Cooperation Development (OECD) countries and foreign direct investment (FDI). The data can be found in the study by Altun and Cordeiro [36] and are as follows:

LTI rate (%): 2.640, 0.596, 0.680, 2.190, 4.560, 2.140, 0.410, 0.530, 0.750, 0.280, 4.390, 3.390, 5.190, 0.800, 2.160, 2.640, 0.060, 2.549, 0.930, 0.310, 0.540, 7.750, 0.470, 2.810, 1.760, 3.170, 1.760, 1.010, 0.990, 1.318, 0.550, 0.040, 1.374, and 2.890.

FDI stocks (outward) (% GDP): 30.78, 57.87, 121.52, 90.17, 45.39, 11.08, 55.92, 51.54, 56.31, 43.34, 11.64, 20.85, 21.99, 276.22, 28.81, 27.56, 30.6, 21.02, 5.93, 7.24, 380.1, 15.76, 305.44, 8.94, 48.05, 5.41, 23.68, 3.56, 14.53, 41.9, 71.7, 162.75, 61.86, and 40.43.

The performance of the BOIPE regression model was compared to that of the beta and simplex regression models. The beta and simplex regression models were fitted using the betareg and simplexreg packages of the R software respectively. The estimated parameters of the BOIPE regression model were obtained using the mle2 function of the bbmle package of the R software. The R codes can be found in the appendix section. Table 8 presents the estimated parameters (standard errors) of the fitted regression model and their goodness of fit statistics. For all the fitted models, the coefficient of the FDI is significant. The coefficient of the FDI in the BOIPE regression model is positive indicating that a change in the FDI increases the LTI rate. However, the coefficient of the FDI in the beta and simplex regression models is negative implying that the FDI decreases the LTI

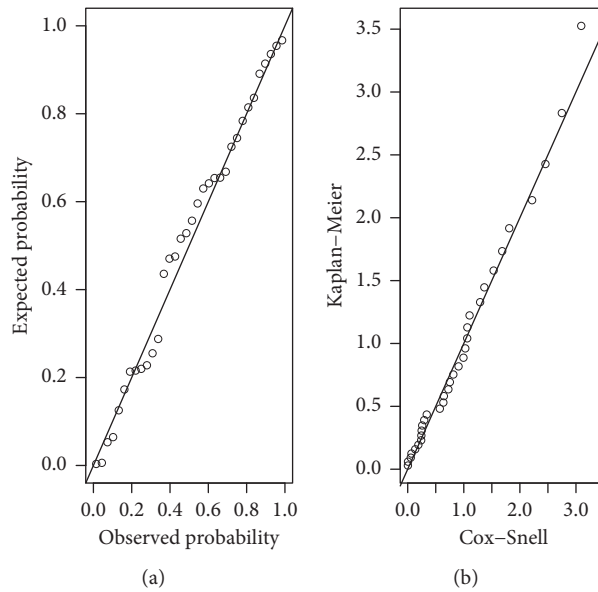


FIGURE 8: P-P plot (a) and Cox-Snell versus $-\log$ of Kaplan-Meier estimates (b).

rate. But, since the BOIPE regression model provides a better fit to the data than the beta and simplex regression models, we conclude that a change in the FDI increases the LTI rate.

From the estimated parameters of the BOIPE regression model, we have

$$\begin{aligned} \log(\alpha_i) &= 0.4675 + 2.6183 \text{ FDI}_i, \\ \log(\lambda_i) &= 0.3315 + 0.1945 \text{ FDI}_i. \end{aligned} \tag{39}$$

In order to examine the adequacy of the BOIPE regression model, we estimated the Cox-Snell residuals [37]. The Cox-Snell residual is defined as

$$\hat{r}_i = -\log \hat{S}(y_i | x_i), \tag{40}$$

where $\hat{S}(\cdot)$ is the estimated survival function. If the model fits the data well, the Cox-Snell residuals are expected to behave like a sample from the standard exponential distribution [5]. Also, the plot of the Cox-Snell residuals versus $-\log \hat{S}(r_i)$, where $\hat{S}(r_i)$ is the Kaplan-Meier estimate of the Cox-Snell residuals, is expected to be a straight line with zero intercept and unit slope. Figure 8 shows the P-P plot of the Cox-Snell residuals and the plot of the Cox-Snell residuals versus the negative logarithm of the Kaplan-Meier estimate of the residuals. It can be seen from the P-P plot that the plotted points are closer to the diagonal line indicating that the model provides an adequate fit to the data. Also, the plot of the Cox-Snell residuals versus the negative logarithm of the Kaplan-Meier estimate of the residuals is a straight line with zero intercept, as shown in Figure 8.

8. Conclusion

In this study, a three-parameter distribution called bounded odd inverse Pareto exponential distribution was proposed. The hazard rate function of the proposed distribution exhibits different kinds of shapes making it suitable for

modeling the dataset with both monotonic and nonmonotonic failure rates defined on the (0, 1) interval. Different estimation techniques were proposed for estimating the parameters of the model. However, the Monte Carlo simulation results revealed that the maximum likelihood procedure estimates the parameters better compared to the other estimation procedures. The empirical applications of the model using real datasets indicated that the new distribution provides good fit to the given datasets compared to other existing distributions. Finally, we proposed the BOIPE regression model and compared its performance with the beta and simplex regression models using the real datasets. The goodness-of-fit statistics revealed that the BOIPE regression model fitted the given data better than the beta and simplex regression models.

Appendix

- (i) ##### BOIPE PDF #####
- (ii) BOIPE_PDF < -function(y, alpha, beta, lambda) {A < -alpha*beta*lambda*(y^(lambda-1))*((1-y)^lambda)^(alpha-1); B < -(1-((1-beta)*(y^lambda))^(alpha+1)); PDF < -A/B; return(PDF)}
- (iii) ##### CDF #####
- (iv) BOIPE_CDF < -function(y, alpha, beta, lambda) {C < -((1-y)^lambda)^(alpha); D < -(1-((1-beta)*(y^lambda))^(alpha)); CDF < -1-(C/D); return(CDF)}
- (v) ##### Hazard Function #####
- (vi) BOIPE_H < -function(y, alpha, beta, lambda) {I < -alpha*beta*lambda*(y^(lambda-1))*((1-y)^lambda)^(-1) J < -(1-((1-beta)*(y^lambda))); H < -I/J; return(H)}
- (vii) ##### Negative Loglikelihood Function #####

- (viii) BOIPE_LL < -function(alpha, beta, lambda) {A < -alpha*beta*lambda*(y^(lambda-1))*((1-y)^lambda)^(alpha-1); B < -(1-((1-beta)*y^lambda))^(alpha+1); PDF < -A/B; LL < -sum(log(PDF)); return(LL)}
- (ix) #####Negative Loglikelihood for Regression Model #####
- (x) BOIPE_LL < -function(alpha0, alpha1, beta, lambda0, lambda1) {alpha < -exp(alpha0+alpha1*x1) lambda < -exp(lambda0+lambda1*x1); A < -alpha*beta*lambda*(y^(lambda-1))*((1-y)^lambda)^(alpha-1); B < -(1-((1-beta)*y^lambda))^(alpha+1); PDF < -A/B; LL < -sum(log(PDF)); return(LL)}
- (xi) #####Optimization #####
- (xii) library (bbmle) ##### Calling the Package bbmle #####
- (xiii) ##### Optimizing the Regression Model #####
- (xiv) fit < -mle2(BOIPE_LL,start = list(alpha0 = 0.03455,alpha1 = 0.005,beta = 100.235, lambda0 = 0.345,lambda1 = 0.0008945),data = list(y,x1),method = "BFGS"); summary(fit)
- (xv) ##### Optimizing the BOIPE distribution for Milk Data #####
- (xvi) fit1 < -mle2(BOIPE_LL,start = list(alpha = 38.94589368,beta = 0.04548055, lambda = 0.70421823),data = list(x),method = "BFGS"); summary(fit1).

Data Availability

The study is on methodological improvement, and the data used can be found within the paper with the appropriate source duly cited.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] M. Smithson and Y. Shou, "CDF-quantile distributions for modelling random variables on the unit interval," *British Journal of Mathematical and Statistical Psychology*, vol. 70, no. 3, pp. 412–438, 2017.
- [2] I. Ghosh, S. Dey, and D. Kumar, "Bounded M-O extended exponential distribution with applications," *Stochastic and Quality Control*, vol. 34, 2019.
- [3] L. E. Papke and J. M. Wooldridge, "Econometric methods for fractional response variables with an application to 401(k) plan participation rate," *Journal of Applied Econometrics*, vol. 11, pp. 619–632, 1996.
- [4] S. Rajarshi and M. B. Rajarshi, "Bathtub distributions: a review," *Communications in Statistics-Theory and Methods*, vol. 17, no. 8, pp. 2597–2621, 1998.
- [5] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Hoboken, NJ, USA, 2003.
- [6] M. T. Lai, "Optimum number of minimal repairs for a system under increasing failure rate shock model with cumulative repair-cost limit," *International Journal of Reliability and Safety*, vol. 7, pp. 95–107, 2013.
- [7] K. L. Woosley and J. Cossman, "Drug development and FDA's critical path initiative," *Public Policy*, vol. 81, pp. 129–133, 2007.
- [8] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 2, John Wiley & Sons, Hoboken, NJ, USA, second edition, 1995.
- [9] J. Mazucheli, A. F. Menezes, and S. Dey, "Unit-Gompertz distribution with applications," *Statistica*, vol. 79, no. 1, pp. 26–43, 2019a.
- [10] A. Grassia, "On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma distribution and derived compounded distributions," *Australian Journal of Statistics*, vol. 19, no. 2, pp. 108–114, 1977.
- [11] P. Kumaraswamy, "Generalized probability density function for double-bounded random processes," *Journal of Hydrology*, vol. 46, pp. 79–88, 1980.
- [12] C. W. Topp and F. C. Leone, "A family of J-shaped frequency functions," *Journal of the American Statistical Association*, vol. 50, pp. 209–219, 1955.
- [13] K. Modi and V. Gill, "Unit Burr III distribution with application," *Journal of Statistics and Management System*, vol. 23, 2019.
- [14] J. Mazucheli, A. F. Menezes, and M. E. Ghitany, "The unit-Weibull distribution and associated inference," *Journal of Applied Probability and Statistics*, vol. 13, pp. 1–22, 2018.
- [15] J. Mazucheli, A. F. B. Menezes, and S. Chakraborty, "On the one parameter unit-Lindley distribution and its associated regression model for proportion data," *Journal of Applied Statistics*, vol. 46, no. 4, pp. 700–714, 2019b.
- [16] P. Jodrá and M. Jiménez-Gamero, "A quantile regression model for bounded responses based on the exponential geometric distribution," *Revstat*, vol. 18, no. 4, pp. 415–436, 2020.
- [17] M. Korkmaz, "A new heavy-tailed distribution defined on the bounded interval: the logit slash distribution and its applications," *Journal of Applied Statistics*, vol. 47, no. 12, pp. 2097–2119, 2019.
- [18] M. Korkmaz and C. Chesneau, "On the unit Burr-XII distribution with the quantile regression modeling and applications," *Computational and Applied Mathematics*, vol. 40, no. 1, pp. 1–26, 2021.
- [19] M. Korkmaz, C. Chesneau, and Z. S. Korkmaz, "On the arcsecant hyperbolic normal distribution. Properties, quantile regression modeling and applications," *Symmetry*, vol. 13, no. 117, pp. 2–24, 2021.
- [20] S. Gündüz and M. Korkmaz, "A new unit distribution based on the unbounded Johnson distribution rule: the unit Johnson SU distribution," *Pakistan Journal of Statistics and Operations Research*, vol. 16, no. 3, pp. 471–490, 2020.
- [21] M. Ghitany, J. Mazucheli, A. F. B. Menezes, and F. Alqallaf, "The unit-inverse Gaussian distribution: a new alternative to two-parameter distributions on the unit interval," *Communications in Statistics-Theory and Methods*, vol. 48, 2018.
- [22] M. A. Aldahlan and A. Z. Afify, "A new three-parameter exponential distribution with applications in reliability and engineering," *Journal of Nonlinear Sciences and Applications*, vol. 13, pp. 258–269, 2020.
- [23] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561, University of California Press, Berkeley, CA, USA, 1961.

- [24] J. J. Swain, S. Venkatraman, and J. R. Wilson, "Least-Squares estimation of distribution functions in Johnson's translation system," *Journal of Statistical Computation and Simulations*, vol. 29, pp. 271–297, 1988.
- [25] R. Cheng and N. Amin, "Maximum product of spacings estimation with application to the lognormal distribution," *Mathematical report*, University of Wales IST, Cardiff, UK, 1979.
- [26] R. Cheng and N. Amin, "Estimating parameters in continuous univariate distributions with a shifted origin," *Journal of the Royal Statistical Society, Series B*, vol. 45, pp. 394–403, 1983.
- [27] B. Ranneby, "The maximum spacing method: an estimation method related to the maximum likelihood method," *Scandinavian Journal of Statistics*, vol. 11, pp. 93–112, 1984.
- [28] S. Ali, S. Dey, M. H. Tahir, and M. Mansoor, "The comparison of different estimation methods for the parameters of flexible Weibull distribution," *Communications Faculty of Sciences University of Ankara Series A-Mathematics and Statistics*, vol. 69, no. 1, pp. 794–814, 2020.
- [29] M. S. Eliwa, El-Morshedy, and S. Ali, "Exponentiated odd Chen-G family of distributions: statistical properties, Bayesian and non-Bayesian estimation with applications," *Journal of Applied Statistics*, vol. 13, 2020.
- [30] J. H. K. Kao, "Computer methods for estimating Weibull parameters in reliability studies," *IRE Transactions Reliability and Quality Control*, vol. 13, pp. 15–22, 1958.
- [31] J. H. K. Kao, "A graphical estimation of mixed Weibull parameters in life test electron tube," *Technometrics*, vol. 1, pp. 389–407, 1959.
- [32] P. Macdonald, "An estimation procedure for mixture of distribution," *Journal of the Royal Statistical Society, Series B*, vol. 33, pp. 326–329, 1971.
- [33] H. M. Yousof, M. Alizadeh, S. M. A. Jahanshahi, T. G. Ramires, I. Ghosh, and G. G. Hamedani, "The transmuted Topp-Leone G family of distributions: theory, characterizations and applications," *Journal of Data Science*, vol. 15, pp. 723–740, 2017.
- [34] R. Dumonceaux and C. E. Antle, "Discrimination between the log-normal and Weibull distributions," *Technometrics*, vol. 15, no. 4, pp. 923–926, 1973.
- [35] A. Pourdarvish, S. M. T. K. Mirmostafae, and K. Naderi, "The exponentiated Topp-Leone distribution: properties and application," *Journal of Applied Environmental and Biological Sciences*, vol. 5, pp. 251–256, 2015.
- [36] E. Altun and G. M. Cordeiro, "The unit-improved second-degree Lindley distribution: inference and regression modeling," *Computational Statistics*, vol. 35, 2019.
- [37] D. R. Cox and E. J. Snell, "A general definition of residuals," *Journal of the Royal Statistical Society, Series B*, vol. 30, pp. 248–275, 1968.