Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110203

New Mixture of Two Beta Exponential Distributions and Income Distribution

Mahmoud Riad Mahmoud and Nesma Mamdouh Amer *

Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt

Received: 28 Feb. 2021, Revised: 12 May 2021, Accepted: 5 Jun. 2021
Published online: 1 May 2022

Abstract: Due to its nice mathematical properties and its usefulness for modeling failure time data, the beta exponentiated distribution is utilized to construct a mixture distribution that would be useful in a variety of applications. Describing income distributions is one of these applications. For this purpose a mixture of the beta exponential and inverted beta exponential distributions is constructed and studied. Approximate formulas for the expected value of the mixture are obtained. The model parameters are estimated using maximum likelihood and applied to real data as well as simulated data.

Keywords: Income distribution, Maximum likelihood, Mixture distribution

1 Introduction

For many years, the beta distribution of the first kind has been considered to be one of the most important distributions due to its widespread applicability. There are many applications of the beta distribution such as being the distribution of probabilities or proportions of occurrence of some discrete event. It arises naturally when sampling randomly from the uniform distribution as the distribution of a single order statistics. Balding and Nichols [1] used beta distribution as a statistical description of allele frequencies in population genetics. Also, Betkowski and Pownuk [2] modeled time allocation in project management or control systems using beta distribution. Sulaiman et. al. [3] presented the fitting of sunshine data covering a period of time to the beta distribution. de Rooij and Stagnitti [4] applied the beta model on the field of variability of soil properties, proportions of the minerals in rocks in stratigraphy; and heterogeneity in the probability of HIV transmission. Also, the beta distribution is used extensively in Bayesian analysis as an appropriate prior distribution.

Due to the huge growth of data in recent years, statisticians need new distributions to fit the new varieties of data. For some data sets the interval $[0,1]$ may be too compact to show the characteristics of the data in different parts of its range appropriately. In such cases, it may be reasonable to use a transformed version of the beta distribution.The main idea here is shifting the finite interval $[0,1]$ for beta distribution to an infinite interval. One of the goals of this paper is to give a new income distribution in the form of mixture beta-exponential distributions defined on the unbounded interval $(0,\infty)$.

The distribution of income in any society is most relevant in economic studies. This is why the distribution of income have received much attention in the past as well as in recent years. Several distributions have been suggested to describe the distribution of income such as Pareto, Weibull, beta and gamma distributions, see [5]. Esteban [6] shows that log-normal, Weibull, exponential, gamma, Pareto distributions are special cases of the generalized gamma distribution. However, most distributions used in this regard describe the income very well on one side of the spectrum of the income but not on the other. That means these income distributions describe only the large or small values of incomes. The probability density function (pdf) of beta distribution is a power function of the variable Y and its reflection $1-Y$, then beta distribution has a pdf of the form

$$g(y) = \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}, \quad 0 \le y \le 1 \tag{1}$$

* Corresponding author e-mail: gentlewind_85@hotmail.com

Nadarajah and Kotz [7] intoduced a new class of transformed beta distribution by using the transformation method to convert the pdf range of beta distribution from the finite interval $[0,1]$ to the infinite interval $(0,\infty)$. Using the transformation function: $X = \frac{-1}{m_1} lnY$ and substituting it in equation (1) gives the new pdf:

$$f_1(x) = \frac{m_1}{B(a,b)}(e^{m_1 x})^a (1-e^{m_1 x})^{b-1}, \quad m_1, a, b > 0, x > 0 \tag{2}$$

[1]

The cdf is given by:

$$F_1(x) = 1 - I_{e^{-m_1 x}}(a,b), \tag{3}$$

where $I_z(a,b)$ is the regularized beta function defined by:

$$I_z(a,b) = \frac{B(z,a,b)}{B(a,b)},$$

and $B(z,a,b)$ is the incomplete beta function:

$$B(z,a,b) = \int_0^z t^{a-1}(1-t)^{b-1} dt$$

and the next figure shows the the cdf at parameters $a = 3, b = 4, m_1 = 1.8$

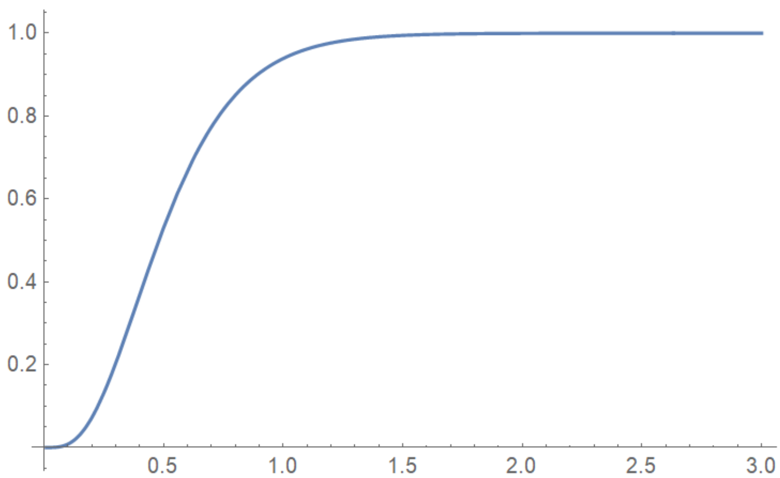

Fig. 1: The cdf of Beta Exponential Distribution

The expectation of X takes the form:

$$E(X) = \frac{\Psi(a+b) - \Psi(a)}{m_1}, \tag{4}$$

and the variance takes the form:

$$Var(X) = \frac{\Psi'(a) - \Psi'(a+b)}{m_1^2}, \tag{5}$$

where Ψ and Ψ' are the digamma function and its first derivative, respectively, see [8]. The characteristic function of X is given by:

$$\phi_x(t) = \frac{B(a - \frac{it}{m_1}, b)}{B(a,b)}, \tag{6}$$

where $i = \sqrt{-1}$.

[1] There are some special cases, In case of $b = 1$, the beta exponential distribution will be exponential distribution

2 Inverted Beta Exponential Distribution

The main purpose in this section is to find the distribution for the large income values that may not be fitted in beta exponential distribution. The pdf and cdf of beta inverted exponential will be presented, also some characteristics of this distribution will be drived. Bakoban and Abu-Zinadah [9] presented $f_2(x)$ be the pdf of the inverted beta distribution which is given by the following:

$$f_2(x) = \frac{m_2}{x^2 B(a,b)}(e^{-m_2/x})^a(1-e^{-m_2/x})^{b-1}; m_2, a, b > 0, x > 0 \tag{7}$$

The cdf, $F_2(x)$, of the inverted beta exponential distribution is given by:

$$F_2(x) = I_{e^{-m_2/x}}(a,b) \tag{8}$$

and the next figure shows the cdf at parameters $a = 3, b = 4, m_2 = 1.7$

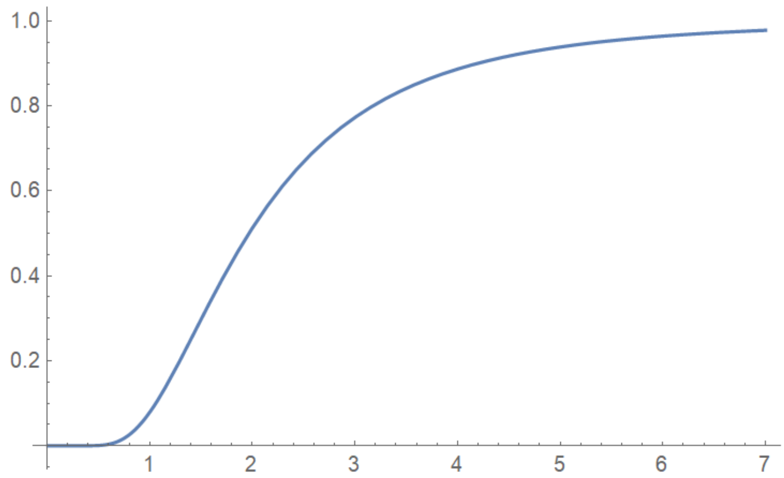

Fig. 2: The cdf of Inverted Beta Exponential Distribution

If b is an integer, the expectation will be in the form

$$E(X) = \frac{m_2}{B(a,b)} \sum_{i=0}^{b-1} \frac{(-1)^{b-i}}{bB(b-i,i+1)} ln[a+b-1-i] \tag{9}$$

with a similar formula if a is an integer. If a and b are both non-intgers, we inerpolate using the above formula between $\lfloor a \rfloor$ and $\lceil a \rceil$, where $\lfloor a \rfloor$ and $\lceil a \rceil$ are floor and ceilng functions.

3 A New Mixture of Beta Exponential Distribution and Its Inverse (MBEI)

Finding a distribution that fits income data, specially both small and large values, is one of the goals of this paper. So, let X distributed as beta exponential random variable and $\frac{1}{X}$ distributed as inverted beta exponential random variable, we shall construct a distribution that is a mixture of the distributions of X and $\frac{1}{X}$. The pdf of the new mixture distribution (MBEI) will be the combination of there corresponding pdf's and it will be as follow:

$$f(x) = p\frac{m_1}{B(a,b)}(e^{-m_1 x})^a(1-e^{-m_1 x})^{b-1} + q\frac{m_2}{x^2 B(a,b)}(e^{-m_2/x})^a(1-e^{-m_2/x})^{b-1}; \tag{10}$$

where $x > 0, am_1 > 0, am_2 > 0, a, b > 0$ and $p + q = 1$. And, the cdf will be as follow:

$$F(x) = p(1 - I_{e^{-m_1 x}}(a,b)) + qI_{e^{-m_2/x}}(a,b), \tag{11}$$

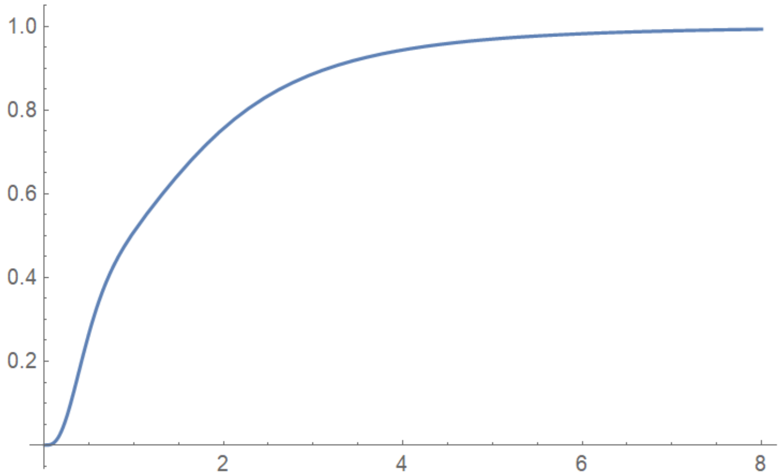

Fig. 3: The cdf of MBEI

where $x > 0$ also the next figure will present the cdf of the (MBEI) distribution at parameters $a = 3, b = 4, m_1 = 1.8, m_2 = 1.7$
If b is an integer, The expectation of the new mixture of beta exponential distribution is given by:

$$E(X) = p\frac{\Psi(a+b) - \Psi(a)}{m_1} + q\frac{m_2}{B(a,b)}\sum_{i=0}^{b-1}\frac{(-1)^{b-i}}{bB(b-i,i+1)}ln[a+b-1-i] \tag{12}$$

with a similar formula if a is an integer.
The next sections will present the estimation of parameters for (MBEI) distribution and the application using simulation and real data will present.

4 Parameter Estimation of MBEI Distribution

The estimation of the parameters for MBEI distribution can be obtained using maximum likelihood estimation (MLE) method [10]. Given a random sample $X_1, X_2, ..., X_n$ from the MBEI distribution with pdf as shown in the equation (10), the likelihood function will be in the form

$$L(\theta) = \prod_{i=1}^{n}[p\frac{m_1}{B(a,b)}(e^{-m_1x_i})^a(1-e^{-m_1x_i})^{b-1} + q\frac{m_2}{x_i^2B(a,b)}(e^{-m_2/x_i})^a(1-e^{-m_2/x_i})^{b-1}] \tag{13}$$

The MLEs can be obtained by taking logarithm of the last equation and differentiate it with respect to a, b, m_1, m_2 which give the following equations respectively:

$$\frac{\partial ln(L(\theta))}{\partial a} = \sum_{i=1}^{n}\frac{m_1p\left(e^{-m_1x_i}\right)^a\left(1-e^{-m_1x_i}\right)^{b-1}\ln\left(e^{-m_1x_i}\right)}{(B_i+A_i)B(a,b)} - \sum_{i=1}^{n}\frac{m_1p(\psi(a)-\psi(a+b))\left(e^{-m_1x_i}\right)^a\left(1-e^{-m_1x_i}\right)^{b-1}}{(B_i+A_i)B(a,b)} + \sum_{i=1}^{n}\frac{m_2q\left(e^{-\frac{m_2}{x_i}}\right)^a\left(1-e^{-\frac{m_2}{x_i}}\right)^{b-1}\ln\left(e^{-\frac{m_2}{x_i}}\right)}{x_i^2(B_i+A_i)B(a,b)} - \sum_{i=1}^{n}\frac{m_2q(\psi(a)-\psi(a+b))\left(e^{-\frac{m_2}{x_i}}\right)^a\left(1-e^{-\frac{m_2}{x_i}}\right)^{b-1}}{x_i^2(B_i+A_i)B(a,b)}, \tag{14}$$

$$\frac{\partial ln(L(\theta))}{\partial b} = \sum_{i=1}^{n} \frac{m_1 p\left(e^{-m_1 x_i}\right)^a \left(1-e^{-m_1 x_i}\right)^{b-1} \ln\left(1-e^{-m_1 x_i}\right)}{(B_i+A_i)B(a,b)} - \sum_{i=1}^{n} \frac{m_1 p(\psi(b)-\psi(a+b))\left(e^{-m_1 x_i}\right)^a \left(1-e^{-m_1 x_i}\right)^{b-1}}{(B_i+A_i)B(a,b)} + \sum_{i=1}^{n} \frac{m_2 q\left(e^{-\frac{m_2}{x_i}}\right)\left(1-e^{-\frac{m_2}{x_i}}\right)^{b-1} \ln\left(1-e^{-\frac{m_2}{x_i}}\right)}{x_i^2(B_i+A_i)B(a,b)} - \sum_{i=1}^{n} \frac{m_2 q(\psi(b)-\psi(a+b))\left(e^{-\frac{m_2}{x_i}}\right)^a \left(1-e^{-\frac{m_2}{x_i}}\right)^{b-1}}{x_i^2(B_i+A_i)B(a,b)}, \quad (15)$$

$$\frac{\partial ln(L(\theta))}{\partial m_1} = \sum_{i=1}^{n} \frac{p\left(e^{-m_1 x_i}\right)^a \left(1-e^{-m_1 x_i}\right)^{b-1}}{(B_i+A_i)B(a,b)} - \sum_{i=1}^{n} \frac{a m_1 p x_i \left(e^{-m_1 x_i}\right)^a \left(1-e^{-m_1 x_i}\right)^{b-1}}{(B_i+A_i)B(a,b)} + \sum_{i=1}^{n} \frac{(b-1) m_1 p x_i \left(e^{-m_1 x_i}\right)^{a+1} \left(1-e^{-m_1 x_i}\right)^{b-2}}{(B_i+A_i)B(a,b)}, \quad (16)$$

and

$$\frac{\partial ln(L(\theta))}{\partial m_2} = \sum_{i=1}^{n} \frac{q\left(e^{-\frac{m_2}{x_i}}\right)^a \left(1-e^{-\frac{m_2}{x_i}}\right)^{b-1}}{x_i^2(B_i+A_i)B(a,b)} - \sum_{i=1}^{n} \frac{a m_2 q\left(e^{-\frac{m_2}{x_i}}\right)^a \left(1-e^{-\frac{m_2}{x_i}}\right)^{b-1}}{x_i^3(B_i+A_i)B(a,b)} + \sum_{i=1}^{n} \frac{(b-1) m_2 q\left(e^{-\frac{m_2}{x_i}}\right)^{a+1} \left(1-e^{-\frac{m_2}{x_i}}\right)^{b-2}}{x_i^3(B_i+A_i)B(a,b)} \quad (17)$$

where $A_i = \frac{(1-p)m_2 e^{\frac{-m_2}{x_i}}(1-e^{-\frac{m_2}{x_i}})^{b-1}}{(x_i)^2 B(a,b)}$, and $B_i = \frac{p m_1 e^{-m_1 x_i}(1-e^{-m_1 x_i})^{b-1}}{B(a,b)}$. Equating the equations from (14) to (17) by zero, the estimated values for a, b, m_1, m_2 can be found, since there is no exact form for the four parameters a, b, m_1, m_2. In the next section, a numerical example will be presented using the above equations to find the estimation of parameters.

5 Numerical Illustration

In this section, a numerical examples are applied to illustrate the computational methods detailed in the preceding sections, but it will be presented some of these examples. Using Mathematica (12.0) package, the next steps were followed:

Step 1Let the weights be $p = q = 0.5$.
Step 2Using equation (11), generate a 1000 random samples of sizes $n = 50, 100, 150$ at different groups of parameters.
Step 3Using equations (14) to (17), calculate the estimators for these groups of parameters.

Tables 1 to 4 will show the results for some of these groups of parameters.
For the first group of parameters which is $\{m_1 = 1.8,\ m_2 = 2,\ a = 3,\ b = 4\}$ and applying the previous three steps, the results will be showed in Table 1

Table 1: Estimation of Parameters with Different Sample Sizes at m_1=1.8, m_2=2, a=3 and b=4

Sample Size	Estimators			
n	m_1	m_2	a	b
50	1.40	2.39	3.48	3.97
100	1.34	2.33	3.61	3.81
150	1.29	2.18	3.89	3.83
Sample Size	Mean Square Error of Parameters			
n	$MSE(m_1)$	$MSE(m_2)$	$MSE(a)$	$MSE(b)$
50	0.32	1.42	3.02	2.12
100	0.32	1.59	3.21	1.18
150	0.33	1.51	3.32	0.63

Table 2: Estimation of Parameters with Different Sample Sizes at m_1=1.8, m_2=2, a=3 and b=4.5

Sample Size	Estimators			
n	m_1	m_2	a	b
50	1.77	2.11	3.29	4.87
100	1.73	1.98	3.32	4.70
150	1.70	1.91	3.37	4.62
Sample Size	Mean Square Error of Parameters			
n	$MSE(m_1)$	$MSE(m_2)$	$MSE(a)$	$MSE(b)$
50	0.20	0.30	0.40	1.44
100	0.14	0.16	0.32	0.70
150	0.13	0.16	0.40	0.45

Applying the previous three steps at the group of parameters $m_1 = 1.8, m_2 = 2, a = 3, b = 4.5$, the results will be shown in Table 2

For the group of parameters $m_1 = 1.8, m_2 = 2, a = 3, b = 5$ and applying the previous three steps, the results will be shown in Table 3

Table 3: Estimation of Parameters with Different Sample Sizes at m_1=1.8, m_2=2, a=3 and b=5

Sample Size	Estimators			
n	m_1	m_2	a	b
50	1.75	2.15	3.32	5.37
100	1.71	2.00	3.34	5.18
150	1.66	1.93	3.41	5.11
Sample Size	Mean Square Error of Parameters			
n	$MSE(m_1)$	$MSE(m_2)$	$MSE(a)$	$MSE(b)$
50	0.24	0.42	0.51	1.72
100	0.16	0.23	0.38	0.94
150	0.14	0.18	0.43	0.76

For the group of parameters $m_1 = 1.8, m_2 = 2, a = 3, b = 3.5$ and applying the previous three steps, the results will be shown in Table 4

It is found that the estimate value obtained from the simulation model closely approximates the estimators at different sample sizes and and also the mean square error for each parameter is relatively small. For further study may be required to study the estimation of parameters using other ways of estimation such as censored samples especially when using a panel data.
In the next section, a numerical example on a real data from a sample of households survey will be present. Also the results of goodness of fit test of this data using different tests will be shown.

Table 4: Estimation of Parameters with Different Sample Sizes at m_1=1.8, m_2=2, a=3 and b=3.5

Sample Size	Estimators			
n	m_1	m_2	a	b
50	1.83	2.07	3.24	3.81
100	1.76	1.95	3.25	3.65
150	1.73	1.92	3.27	3.61
Sample Size	Mean Square Error of Parameters			
n	$MSE(m_1)$	$MSE(m_2)$	$MSE(a)$	$MSE(b)$
50	0.23	0.25	0.44	0.87
100	0.11	0.11	0.28	0.35
150	0.07	0.08	0.20	0.23

6 Example On a Real Data

This section presents the results of fitting the MBEI distribution to a real data set. The data is a sample based on households income survey of Egypt where the survey sample size is 11000 households. All the calculations were carried out using the *Mathematica*(12.0) package after taking logarithm of income data and the resuts are summarized in Figure 4 shows the histogram of logarithm income data:

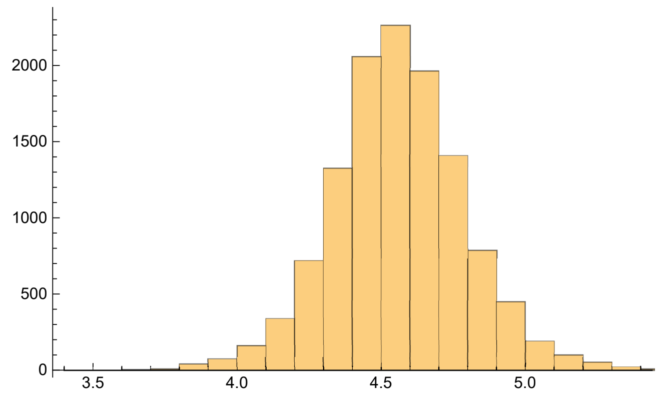

Fig. 4: Histogram of Log Income Data

Using the real data of income and the log likelihood of (MBEI) distribution and after maximizing it, it is found that the parameters a, b, m_1, m_2 that maximize the log likelihood could take the values 375.9, 463.3, 0.18 and 3.67 respectively. The cdf using these parameters and the empirical distribution of log income household data can show how the data fits of the concerning distribution and Figure 5 shows how the two curves of the empirical and the cumulative distributions are almost identical.

The data is tested if it is fitted the distribution or not where the null hypothesis and its alternative will be:
H_0: The dataset have the same distribution (MBEI) against H_1: the dataset haven't the same distribution (MBEI).
After using different tests such as Kolmogorov Smirnov and Anderson Darling and so on, The results of goodness of fit test for the log income data using diffrent tests give that the null hypothesis the datasets have the same distribution is not rejected at the 5 percent of significance level. the next table shows results of the diffrent kinds of tests and test statistics for each test and also the $p$values:

7 Conclusion

It has been shown that the new MBEI distribution with four parameters is an extension of several known distributions, which means that this new distribution has more flexibility to model many kinds of data specially data generated by

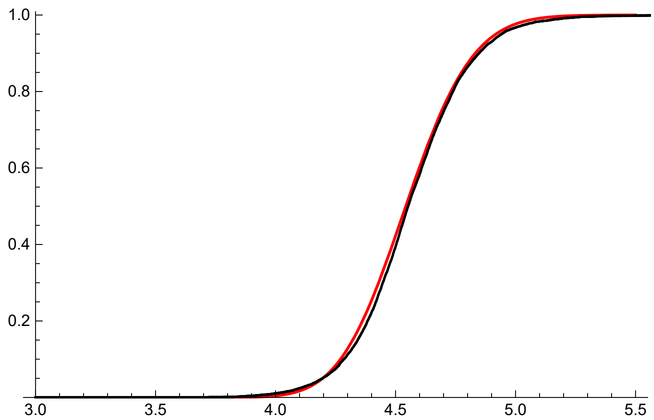

Fig. 5: Empirical and Cumulative Distribution

Table 5: Results Using Different Goodness of Fit Tests

Test Name	Test statistics	pValue	Result
Kolmogorov Smirnov	0.023	0.95	Is not reject H_0
Anderson Darling	0.78	0.49	Is not reject H_0
Cramer Von Mises	0.06	0.84	Is not reject H_0
Kuiper	0.05	0.57	Is not reject H_0
PearsonChiSquare	28.41	0.60	Is not reject H_0

complex processes such as genetic data and income data. Using Kolomogrove-Smirnov test statistics, it has been shown that empirically the new MBEI distribution gives a better fit than some popular distributions used to model income. It is observed that as it is expected, in most cases as the sample size increases the estimates impove with respect to bias and MSE. The mixing proportions were assumed to be known to simplify the presentation, but they could have been estimated were they unkown. An approximate formula for the expected value of the new distribution is given and with little more work the variance could be approximated.

Conflicts of Interests: The authors declare that they have no conflicts of interests

References

[1] D.J. Balding and R.A. Nichols, A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity, *Genetica*, **96(1-2)**, 3-12 (1995).
[2] M.L. Betkowski, and A. Pownuk, Calculating risk of cost using Monte Carlo simulations with fuzzy parameters in civil engineering, *Changes*, **10**, 6 (2004).
[3] M.Y. Sulaiman, W.M.H. Oo, M. Abd Wahab and A. Zakaria, Application of beta distribution model to Malaysian sunshine data, *Renewable Energy* **18(4)**, 573-579 (1999).
[4] G. de Rooij and F. Stagnitti, Applications of the beta distribution in soil science, *Handbook of Beta distribution and Its Application*, 535-550, (2004).
[5] J.B. McDonald, *Some generalized functions for the size distribution of income*, Modeling Income Distributions and Lorenz Curves, Springer, 37-55 (2008).
[6] J. Esteban, Income-share elasticity and the size distribution of income, *International Economic Review*, **27(2)**, 439-444 (1986).
[7] S. Nadarajah and S. Kotz, The beta exponential distribution, *Reliability Engineering & system safety*, **91(6)**, 689-697 (2006).
[8] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, US Government printing office, (1948).
[9] R.A. Bakoban and H.H Abu-Zinadah, The beta generalized inverted exponential distribution with real data applications, *REVSTAT-Statistical Journal* **15(1)**, 65-88 (2017).
[10] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, (2011).