International Journal of Applied Mathematics

Volume 33 No. 1 2020, 99-108

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

doi: http://dx.doi.org/10.12732/ijam.v33i1.8

THE TRUNCATED ISHITA DISTRIBUTION: PROPERTIES AND APPLICATION

Issaraporn Thiamsorn¹, Sirinapa Arvuvuen²§

^{1,2}Department of Mathematics and Computer Science Rajamangala University of Technology Thanyaburi Pathum Thani – 12110, THAILAND

In this paper, a new truncated distribution, which is called the Abstract: truncated Ishita (TI) distribution, is proposed. Some statistical properties including moments, survival, and hazard functions, are discussed. Moreover, the maximum likelihood estimation is constructed for estimating the parameters of the TI distribution. Finally, an application based on real data is conducted to illustrate the usefulness of the proposed distribution.

AMS Subject Classification: 60E05, 62H10

Key Words: truncated distribution; truncated Ishita distribution; survival funciton; lifetime data

1. Introduction

Lifetime data modeling is studied extensively by several researchers for different lifetime distributions with unbounded support, such as exponential, gamma, and Lindley distributions with the probability density function (pdf) as follows (see [1, 2]), respectively

$$g_1(x;\theta) = \theta e^{-\theta x}, \tag{1}$$

$$g_1(x;\theta) = \theta e^{-\theta x},$$

$$g_2(x;\alpha,\theta) = \frac{\theta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x},$$
(1)
(2)

$$g_3(x;\theta) = \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x},\tag{3}$$

respectively, where x>0 and $\alpha>0, \theta>0$. Lifetime data modeling is studied extensively by several researchers for different lifetime distributions with

Received: August 23, 2019

© 2020 Academic Publications

[§]Correspondence author

unbounded support. These are mainly based on some modifications and generalizations of distribution. However, in practice, some lifetime data when using these distributions are not suitable due to their shapes, nature of hazard rate functions, and mean residual life, amongst others. However, the finite mixture distribution is available as it is a flexible and powerful probabilistic modeling tool for univariate and multivariate data, which is widely acknowledged in the area which involves the statistical modeling of data, [3].

One mixture distribution, Ishita distribution is proposed by [4], which is the modeling to analysis lifetime data in biomedical science and engineering. The condition under which Ishita distribution is over-dispersed and under-dispersed.

Moreover, data analysis is undertaken in various fields, including engineering, medicine, finance, and demographics, where such types of truncated data arise in practical statistics. It is used in cases where the ability to record exists, or even when occurrences are limited to values that lie above or below a given threshold or within a specified range. Truncated distributions are quite effective in using data analysis, [5, 6]. In 1994, [7] introduced the truncated normal distribution. This distribution has wide applications in Statistics and Econometrics. For example, it is used to model the probabilities of binary outcomes in the probit model and to fit censored data in the Tobit model. The truncated distribution is a probability distribution, which is derived from a normal random variable by bounding it from either below or above (or both) [6]. Many researchers have therefore been attracted to the problems of analyzing such truncated data encountered in various disciplines and have proposed truncated versions of the distributions, i.e., truncated Pareto [8], truncated Birnbaum-Saunders [9], truncated Weibull [10], truncated two-parameter Lindley [11], and truncated Power Lomax [12].

The rest of the paper has been organized into the following sections. In the preliminary method is introduced. Next, a new truncated version of Ishita distribution is proposed. Some statistical properties, the method for the parameters estimating of proposed distribution are illustrated. Moreover, the application of the proposed distribution is discussed. Finally, some conclusions are presented

2. Preliminary Methods

2.1. The Ishita distribution

The Ishita distribution is proposed by [4], which is a mixture distribution between the exponential distribution with scale parameter $\theta > 0$ and the gamma distribution with the shape parameter $\alpha = 3$ and scale parameter θ with the mixing proportion of $p = \frac{\theta^3}{\theta^3 + 2}$. Let X be a random variable which is distributed the Ishita distribution, then the pdf and cumulative density function (cdf) of X are

$$g(x;\theta) = pg_1(x;\theta) + (1-p)g_2(x;\theta) = \frac{\theta^3}{\theta^3 + 2} (\theta + x^2) e^{-\theta x},$$
 (4)

$$G(x;\theta) = 1 - \left(1 + \frac{\theta x(\theta x + 2)}{\theta^3 + 2}\right)e^{-\theta x}, \ x > 0, \theta > 0,$$
 (5)

respectively. The rth moment about origin of the Ishita distribution is given by

$$\mu_r' = \frac{r! \left[\theta^3 + (r+1)(r+2) \right]}{\theta^r (\theta^3 + 2)} , \quad r = 1, 2, 3, \dots$$
 (6)

2.2. The truncated version of distribution

Suppose X be a random variable that is distributed according to some pdf $g(x;\Omega)$, and cdf, $G(x;\Omega)$, which Ω is a parameter vector of X. Let X lies within the interval [a,b] where $-\infty < a \le x \le b < \infty$, then the conditional on $a \le x \le b$ is distributed the truncated distribution. We have the pdf of truncated distribution [6] as

$$f(x;\Omega) = g(x|a \le x \le b;\Omega) = \frac{g(x;\Omega)}{G(b;\Omega) - G(a;\Omega)},$$
(7)

where $f(x;\Omega) = g(x;\Omega)$ for all $a \le x \le b$ and $f(x;\Omega) = 0$ everywhere else. Notice that $f(x;\Omega)$, in fact, $f(x;\Omega)$ is a pdf of X on interval [a,b]:

$$f(x;\Omega) = \int_{a}^{b} f(x;\Omega)dx = \frac{1}{G(b;\Omega) - G(a;\Omega)} \int_{a}^{b} g(x;\Omega)dx$$
$$= \frac{1}{G(b;\Omega) - G(a;\Omega)} G(b;\Omega) - G(a;\Omega) = 1, \tag{8}$$

and $f(x;\Omega) = 0$ every where else. The cdf of truncated distribution is given by

$$F(x;\Omega) = \int_{a}^{x} f(x;\Omega)dx = \frac{G(x;\Omega) - G(a;\Omega)}{G(b;\Omega) - G(a;\Omega)}.$$
 (9)

3. Results

3.1. A new truncated distribution

This section, a new truncated distribution, namely, the truncated Ishita (TI) distribution, is proposed. From the pdf and cdf of Ishita distribution in (4) and (5), respectively, and the pdf and cdf of truncated version in (7) and (9), we have the definition of proposed distribution as follows.

Definition 1. Let X be a random variable which is distributed the TI distribution with scale parameters of a, b and θ , will be denoted $X \sim \text{TI}(\theta, a, b)$. The pdf and cdf of X are respectively:

$$f(x;\theta,a,b) = \frac{\theta^3 (\theta + x^2) (\theta^3 + 2)^{-1} e^{-\theta x}}{\left(1 + \frac{\theta a(\theta a + 2)}{\theta^3 + 2}\right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^3 + 2}\right) e^{-\theta b}},$$
 (10)

$$F(x;\theta,a,b) = \frac{\left(1 + \frac{\theta a(\theta a + 2)}{\theta^3 + 2}\right) e^{-\theta a} - \left(1 + \frac{\theta x(\theta x + 2)}{\theta^3 + 2}\right) e^{-\theta x}}{\left(1 + \frac{\theta a(\theta a + 2)}{\theta^3 + 2}\right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^3 + 2}\right) e^{-\theta b}},$$
(11)

where $-\infty < a \le x \le b < \infty$ and $\theta > 0$.

Some plots of the function in (10) and (11) are illustrated in Figure 1.

3.2. Characteristic properties

3.2.1. Moments

The rth moment about origin of TI distribution is defined as in Definition 2.

Definition 2. Let $X \sim \text{TI}(\theta, a, b)$, then the rth moment about origin of TI distribution is

$$\mu'_{[r]} = \frac{r! \left[\theta^3 + (r+1)(r+2) \right]}{\theta^r(\theta^3 + 2) \left[\left(1 + \frac{\theta a(\theta a + 2)}{\theta^3 + 2} \right) - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^3 + 2} \right) \right]},\tag{12}$$

where $r = 1, 2, 3, ..., \text{ and } \theta > 0.$

From (12), $\mu'_{[r]} = E(X^r)$, we obtained the mean and variance of TI distri-

bution, i.e., respectively,

$$E(X) = \frac{\theta^{3} + 6}{\theta(\theta^{3} + 2) \left[\left(1 + \frac{\theta a(\theta a + 2)}{\theta^{3} + 2} \right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^{3} + 2} \right) e^{-\theta b} \right]},$$

$$V(X) = \frac{2(\theta^{3} + 12)}{\theta^{2}(\theta^{3} + 2)^{2} \left[\left(1 + \frac{\theta a(\theta a + 2)}{\theta^{3} + 2} \right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^{3} + 2} \right) e^{-\theta b} \right]^{2}} \times \left\{ 2(\theta^{3} + 12)(\theta^{3} + 2) \left[\left(1 + \frac{\theta a(\theta a + 2)}{\theta^{3} + 2} \right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^{3} + 2} \right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^{3} + 2} \right) e^{-\theta b} \right] - (\theta^{3} + 6)^{2} \right\}.$$

$$(13)$$

3.2.2. Survival and hazard functions

The survival function, S(x), is the probability that a subject survives longer than time x. Suppose X is a lifetime random variable representing the time until a specified event of interest is occurred, then the survival function of X is defined as S(x) = 1 - F(x). From Eq.11, the survival function of TI distribution is

$$S(x;\theta,a,b) = 1 - \frac{\left(1 + \frac{\theta a(\theta a + 2)}{\theta^3 + 2}\right) e^{-\theta a} - \left(1 + \frac{\theta x(\theta x + 2)}{\theta^3 + 2}\right) e^{-\theta x}}{\left(1 + \frac{\theta a(\theta a + 2)}{\theta^3 + 2}\right) e^{-\theta a} - \left(1 + \frac{\theta b(\theta b + 2)}{\theta^3 + 2}\right) e^{-\theta b}}.$$
 (14)

Consequently, the ratio of pdf and survival function, is given by h(x) = f(x)/S(x), which is called the hazard function. From the pdf in (10) and the survival function in (14), we have

$$h(x;\theta,a,b) = \frac{\theta^3(\theta+x^2)(\theta^3+2)^{-1}e^{-\theta x}}{\left(1 + \frac{\theta x(\theta x+2)}{\theta^3+2}\right)e^{-\theta x} - \left(1 + \frac{\theta b(\theta b+2)}{\theta^3+2}\right)e^{-\theta b}}.$$
 (15)

3.3. Maximum likelihood estimation

Let X_1, X_2, \ldots, X_n be a random sample sized n from $TI(\theta, a, b)$. Then the log likelihood function is given by

$$\log L = \prod_{i=1}^{n} \log f(x_i; \theta, a, b)$$

$$= 3n \log(\theta) + \sum_{i=1}^{n} \log(\theta + x_i^2) - \theta \sum_{i=1}^{n} x_i - n \log(\theta^3 + 2)$$

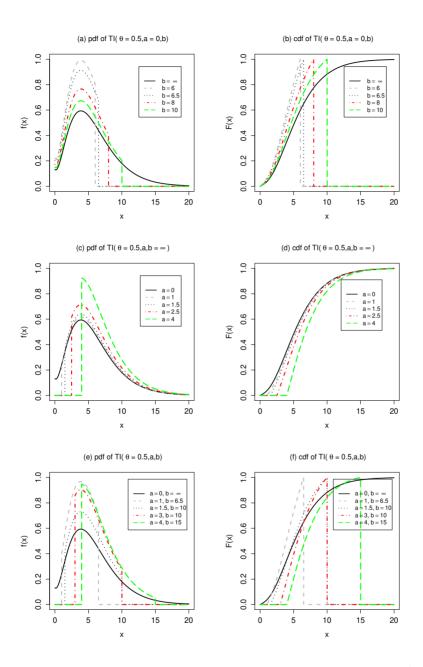


Figure 1: Some plots of pdf and cdf of X for $X \sim \text{TI}(\theta, \mathbf{a}, \mathbf{b})$ and specified the values of θ, a and b.

$$-\log\Big((1+\frac{\theta a(\theta a+2)}{\theta^{3}+2})e^{-\theta a}-(1+\frac{\theta b(\theta b+2)}{\theta^{3}+2})e^{-\theta b}\Big).$$

For $\hat{a} = \min(x)$ and $\hat{b} = \max(x)$, the maximum likelihood estimated value of parameter θ is obtained by $\hat{\theta} = \frac{\partial \log L}{\theta} = 0$, which the expression is not in closed form. This work is introduced the numerical optimization with **nlm** function of **stat** package [13, 14] in R [15] that to estimate the parameters of TI distribution.

3.3.1. R code for MLE of TI parameters

```
library(stats)
logTI<-function(x,t){
theta<-t[1]; a<-min(x); b<-max(x)
prob<-((theta^3)/(theta^3+2)*(theta+x^2)*exp(-theta*x))
/((1+(theta*a*(theta*a+2))/(theta^3+2))*exp(-theta*a)
-(1+(theta*b*(theta*b+2))/(theta^3+2))*exp(-theta*b))
lpdf1<-log(prob)
loglike<--sum(lpdf1)
return(loglike)
}
out<-nlm(logTI,x=x,c(0.01))</pre>
```

3.4. Application study

The goodness of fit of TI distribution has been done on several lifetime data sets. In this section, we present the goodness of fit of proposed distribution using maximum likelihood estimate of the parameter on two data sets and the fit has been compared with Ishita, exponential, and Lindley distributions. The real data set, the strength data of glass of the aircraft window [16] is considered as follows: 18.83, 20.80, 21.657, 23.03, 23.23, 24.05, 24.321, 25.50, 25.52, 25.80, 26.69, 26.77, 26.78, 27.05, 27.67, 29.90, 31.11, 33.20, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381.

The distribution selection is carried out that is using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which are given respectively as follows: AIC = $-2 \log L + 2k$ and BIC = $-2 \log L + k \log(n)$, where $\log L$ denotes the log-likelihood function evaluated at the maximum likelihood estimates of the parameters, k is the parameter number of any distributions, and n is the sample size. In addition, we used KolmogorovSmirnov (K-S) test, K-S = $\sup_{x} |F_n(x) - F_0(x)|$ where $F_n(x)$ and $F_0(x)$ are the empirical distributions

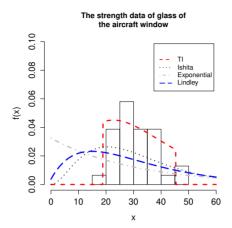


Figure 2: Plots among histogram of the strength data of the glass of the aircraft window and the theoretical pdf of distributions.

tion function and the theoretical cumulative distribution of distributions being tested, respectively. Results of the parameter estimation and goodness of fit test are shown in Table 1. For testing the goodness of fit test (minimum of K-S values) in this data set, the observations about strength data of glass of the aircraft window is distributed as close to the TI distribution, which it has $\hat{\theta} = 0.0883$, $\hat{a} = 18.83$, and $\hat{b} = 45.381$, greater than the Ishita, Lindley, and exponential distributions, respectively. Moreover, the proposed distribution has the minimum value of AIC and BIC when it compares the other distributions. Plots for the histogram of data and the theoretical pdf of distributions are shown in Figure 2.

Table I: Maximum likelihood estimates of the parameter in each distributions for glass of the aircraft window

	Distributions			
	TI	Ishita	Exponential	Lindley
$\hat{ heta}$	0.0883	0.0974	0.0325	0.0630
$-2\log L$	115.9686	154.4796	274.5288	253.9884
AIC	117.9686	156.4796	276.5288	255.9884
BIC	119.4025	157.9136	277.9628	257.4224
K-S	0.1115	0.2984	0.4586	0.3655

4. Conclusion

This paper is proposed a truncated version of Ishita distribution, which is called the truncated Ishita (TI) distribution. The probability functions, i.e., the probability density function, cumulative density function, moments, survival function, and hazard function, are introduced. The unknown parameter of TI distribution is estimated by using the maximum likelihood estimation, and R code to estimate the parameter is presented. Finally, the proposed distribution has been fitted with the real data to illustrate the usefulness of distribution, and it compares among the other distribution, i.e., Ishita, exponential, Lindley distributions.

References

- [1] R. Shanker, F. Hagos and S. Sujatha, On modeling of Lifetimes data using exponential and Lindley distributions, *Biometrics & Biostatistics International Journal*, **2** (2015), 1-9.
- [2] H.C Thom, A note on the gamma distribution, Monthly Weather Review, 86 (1958), 117-122.
- [3] M.A.T. Figueiredo and A.K. Jain, Unsupervised learning of finite mixture models, *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24** (2002), 381-396.
- [4] R. Shanker and K.K. Shukla, Ishita distribution and its applications, *Biometrics & Biostatistics International Journal*, **5** (2017), 1-9.
- [5] Y. Dodge, *The Oxford Dictionary of Statistical Terms*, Oxford, The Oxford University Press (2003).
- [6] S.K. Singh, U. Singh and V.K. Sharma, The truncated Lindley distribution: Inference and application, *Journal of Statistics Applications & Probability*, 3 (2014), 219-228.
- [7] N.L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Wiley, New York (1994).
- [8] L. Zaninetti and M. Ferraro, On the truncated Pareto distribution with applications, *Central European Journal of Physics*, **6** (2008), 1-6.

- [9] S.E. Ahmed, C. Castro-Kuriss, E. Flores, V. Leiva, A truncated version of the Birnbaum-Saunders distribution with an application in financial risk, *Pakistan Journal of Statistics*, **26** (2010), 293-311.
- [10] T. Zange and M. Xie, On the upper truncated Weibull distribution and its reliability implications, *Reliability Engineering & System Safety*, **96** (2011), 194-200.
- [11] S. Aryuyuen, Truncated two-parameter Lindley distribution and its application, *Journal of Applied Sciences*, **17** (2018), 19-32.
- [12] S. Aryuyuen and W. Bodhisuwan, The truncated power Lomax distribution: Properties and applications, *Walailak Journal of Science and Technology*, **16** (2018), 655-668.
- [13] J.E. Dennis and R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, N.J. (1983).
- [14] R.B. Schnabel, J.E. Koontz and B.E. Weiss, A modular system of algorithms for unconstrained minimization, ACM Transactions on Mathematical Software, 11 (1985), 419-440.
- [15] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org (2018).
- [16] E.J. Fuller, S. Frieman, J. Quinn, G. Quinn and W. Carter, Fracture mechanics approach to the design of glass aircraft windows: A case study, SPIE Proc., 2286 (1994), 419-430.