The Journal of Applied Science **Vol. 17 No. 1: 19-32 [2018]**
วารสารวิทยาศาสตร์ประยุกต์ doi: 10.14416/j.appsci.2018.05.005
ISSN 1513-7805 Printed in Thailand ISSN 2586-9663 (Online)

Research Article

Truncated two-parameter Lindley distribution and its application

Sirinapa Aryuyuen[1]*

[1]Department of Mathematics and Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi, Pathum Thani, 10110, Thailand
***E-mail**: sirinapa_a@rmutt.ac.th
Received: 05/04/2018; Accepted: 30/05/2018

Abstract

In this paper, the truncated two-parameter Lindley (T-TPL) distribution is proposed, which is classified as doubly, left and right truncated distributions. Some statistical properties, i.e., the moments, and functions of survival, hazard and quantile are also discussed. The maximum likelihood estimators are constructed for estimating the unknown parameters of the T-TPL distribution. Moreover, the distributions have been fitted with some real data to illustrate the efficiency of the T-TPL distribution when it compared to other distributions (i.e., exponential, Lindley, two-parameter Lindley, and truncated Lindley distributions). The results have shown that the T-TPL distribution gives a reasonable better fit to the real data about bladder cancer patients and the flood discharge rates than other distributions.

Keywords: two-parameter Lindley, hazard function, truncated distribution, lifetime data

Introduction

Truncated distributions are quite effective for using data analysis in various fields, including engineering, medicine, finance and demographics, when such types of truncated data arise in practical statistics in cases with the opportunity to record, or even when occurrences are limited to values which lie above or below a given threshold or within a specified range. For example, pH, grades, and humidity in environmental studies have upper and lower physical bounds, and their probability distribution functions are not necessarily symmetrical within these bounds (Dodge, 2003; Singh et al., 2014).

A truncated normal distribution had been derived from the standard normal distribution as proposed by Johnson et al. (1994). The truncated normal distributions are applied in many practical situations where there is a restriction on the variable under consideration. Several studies have considered the various aspects of the truncated normal distributions, i.e., Johnson (2001) studied the truncated normal distribution in characteristics of singly and doubly truncated populations of application in management science, Iwueze (2007) applied the truncated normal distribution to the left at zero is in descriptive modelling of time series data, Sun (2013) introduced the finite fault modelling for the Wenchuan earthquake using hybrid slip model with truncated normal distributed source parameters, and Hamasha (2017) proposed a mathematical approximation of the left-sided truncated normal distribution using the Cadwell an approximation model to data about the life of a television set. Therefore, many researchers being attracted to problems of analyzing such truncated data encountered in various disciplines, proposed the truncated versions of usual probability distributions. Ahmed et

 doi: 10.14416/j.appsci.2018.05.005

al. (2010) have discussed the application of the truncated version of the Birnbaum-Saunders distribution to improve an actuarial forecasting model and particularly for modeling data from insurance payments to establish deductibles. For instance, Aban et al. (2006), and Zaninetti & Ferraro (2008) have discussed the application of the truncated Pareto distribution to the statistical analysis of the masses of stars and the diameters of asteroids. Zhang & Xie (2011) have proposed the upper-truncated Weibull distribution and its application about time-to-failure of the turbocharger of one type of engine. Recently, Singh et al. (2014) have proposed the truncated Lindley distribution and applied its distribution to a set of real data concerning the strengths of the glass of aircraft windows.

From the above commentary as well as monitoring the wide applicability of the truncated distributions, the truncated versions of a two-parameter Lindley (TPL) distribution have been proposed as a flexible alternative for analyzing the truncation data range. The TPL distribution has been introduced by Shanke et al. (2013), which becomes a lifetime distribution by mixing the exponential distribution with a scale parameter θ and the gamma distribution with the shape parameter 2 and the scale parameter θ by using the mixed proportion with $\theta/(\theta+\alpha)$ and $\alpha/(\theta+\alpha)$. The TPL has been fitted to some data-sets relating to waiting times and survival times (i.e., the waiting times of bank customers, the survival times of guinea pigs infected with virulent tubercle bacilli, and the mortality grouped data for blackbird species). Results based on chi-square test shown that the TPL distribution provides better fits than the Lindley distribution. The Lindley distribution (Lindley, 1985) is a mixture of the exponential distribution with a scale parameter θ and the gamma distribution the shape parameter 2 and the scale parameter θ. We will find that the TPL has the shape parameter α that makes it more flexible to fit the data than the Lindley distribution.

The rest of the paper has been arranged in the following sections. In methods, the TPL distribution and the truncated distribution are introduced. For results of study, a new truncated distribution of the TPL distribution is proposed, which is classified as doubly, left and right truncated distributions. Particularly, the flexibility of the proposed distribution has been shown to demonstrate the characteristics of the probability function with different combinations of the values of its parameters. Some statistical properties such as moments, survival, hazard, and quantile functions, are also discussed. Moroever, the method of the maximum likelihood estimation is applied to obtain the parameter estimate of the proposed distribution. The set of real data is modeled through the different distributions to compare the efficiency of the proposed distribution and other distributions (i.e., exponential, Lindley, two-parameter Lindley, and truncated Lindley distributions). Finally, the conclusions are shown.

Methods

Let X be distributed as the TPL random variable with the shape parameter α and the scale parameter θ, which denoted by $\mathrm{X} \sim \mathrm{TPL}(\alpha,\theta)$, where the probability density function (pdf) and cumulative density function (cdf) of X (Shanker et al., 2013) are, respectively,

$$\mathrm{f}(x;\alpha,\theta) = \frac{\theta^2}{\theta+\alpha}(1+\alpha x)e^{-\theta x} \text{ and } \mathrm{F}(x;\alpha,\theta) = 1 - \frac{\theta+\alpha+\alpha\theta x}{\theta+\alpha}e^{-\theta x}, \qquad (1)$$

where $x>0$, $\theta>0$, and $\alpha>-\theta$.

It can easily be seen that when $\alpha=1$, the TPL distribution reduces to the one-parameter Lindley distribution (Lindley, 1985) and that when $\alpha=0$, it reduces to the exponential distribution (Gutenberg & Richter, 1944). Some plots of the pdf of the TPL are shown in Figure 1.

 doi: 10.14416/j.appsci.2018.05.005

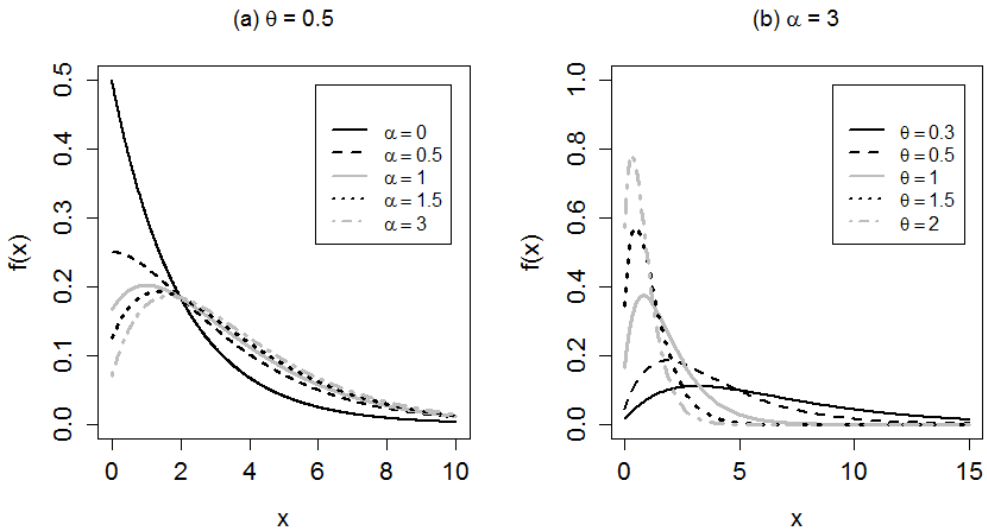

Figure 1. Some pdf plots of the TPL distribution with specified parameters θ and α

Suppose X has a parent distribution, which is a continuous random variable, $-\infty < x < \infty$, with the parameter Θ. The pdf and c.d.f of X are $\mathrm{f}(x;\Theta)$ and $\mathrm{F}(x;\Theta)$, respectively. Let $\mathrm{T} \in F$ where F is a σ-field on a real number $\Re$, such that $0 \le \mathrm{P}(\mathrm{X} \in \mathrm{T}) \le 1$. Then the conditional distribution of X, i.e., $\mathrm{P}(\mathrm{X} \le x \mid \mathrm{X} \in \mathrm{T})$, is defined for any $x \in \Re$ it is called the truncated distribution of X. We obtain the truncated distribution function with the cdf and the pdf of $\mathrm{X} \in \mathrm{T}$ (Dodge, 2003; Singh et al., 2014; Ahsamullah et al., 2016) are given as, respectively,

$$\mathrm{F}(x;\Theta,\mathrm{T}) = \mathrm{P}(\mathrm{X} \le x \mid \mathrm{X} \in \mathrm{T}) = \frac{\int_{(-\infty,x]\cap\mathrm{T}} \mathrm{f}(u;\Theta)\,du}{\int_{\mathrm{T}} \mathrm{f}(u;\Theta)\,du}, \text{ and}$$

$$\mathrm{f}(x;\Theta,\mathrm{T}) = \frac{\mathrm{f}(x;\Theta)}{\int_{\mathrm{T}} \mathrm{f}(u;\Theta)\,du}, \quad x \in \mathrm{T}. \tag{2}$$

Clearly $\mathrm{f}(x;\Theta,\mathrm{T})$ defines the pdf of X with support T, i.e., T is not necessarily a bounded set of real numbers, since $\int_{\mathrm{T}} \mathrm{f}(x;\Theta,\mathrm{T})dx = \int_{\mathrm{T}} \mathrm{f}(x;\Theta)\,dx \Big/ \int_{\mathrm{T}} \mathrm{f}(u;\Theta)\,du = 1.$

Let X is defined on $\mathrm{T} = [a,b]$, the conditional distribution of X given that $-\infty < a \le x \le b < \infty$, which is called a doubly truncated distribution on interval $[a,b]$. Then the pdf of X is

$$\mathrm{f}(x;\Theta,a,b) = \frac{\mathrm{f}(x;\Theta)}{\mathrm{F}(b;\Theta) - \mathrm{F}(a;\Theta)}; \qquad -\infty < a \le x \le b < \infty. \tag{3}$$

If X be a parent distribution, which is a lifetime distribution with the pdf $\mathrm{f}(x;\Theta)$ where $x \ge 0$ then the doubly truncated distribution of X on interval $[a,b]$, the equation (3) is consideraed in 3 cases.

1) When $a = 0$ and $b < \infty$, the doubly truncated distribution reduces to the parent distribution.

doi: 10.14416/j.appsci.2018.05.005

2) When $a>0$ and $b\to\infty$, the doubly truncated distribution reduces to the left truncated distribution at a, on the other hand, its probability distribution can be derived from the parent distributioin by bounding the random variable from above a.

3) When $a=0$ and $b<\infty$, the doubly truncated distribution reduces to the right truncated distribution at b, or its probability distribution can be derived from the parent distribution by bounding the random variable from below b.

Results and discussion

A new truncated distribution

A new distribution, namely, a truncated two-parameter Lindley (T-TPL) distribution is presented, which is classified as doubly, left and right truncated distributions. Moreover, some statistical properties are also discussed.

1) Suppose X has a two-parameter Lindley distribution with the pdf and cdf as in equation (1). We obtained the doubly truncated distribution of X on interval $[a,b]$. It is called a doubly truncated two-parameter Lindley (dT-TPL) distribution, which is denoted as X ~dT-TPL(θ,α,a,b). If $f(x;\alpha,\theta)$ and $F(x;\alpha,\theta)$ in equation (1) are replaced into equation (3) respectively, then pdf and cdf of the dT-TPL distribution are, respectively,

$$f(x;\alpha,\theta,a,b)=\frac{\theta^2(1+\alpha x)e^{-\theta x}}{(\theta+\alpha+\theta\alpha a)e^{-\theta a}-(\theta+\alpha+\theta\alpha b)e^{-\theta b}}, \quad (4)$$

$$F(x;\alpha,\theta,a,b)=\frac{(\theta+\alpha)-(\theta+\alpha+\theta\alpha x)e^{-\theta x}}{(\theta+\alpha+\theta\alpha a)e^{-\theta a}-(\theta+\alpha+\theta\alpha b)e^{-\theta b}}. \quad (5)$$

The plots of pdf in equation (4) is shown in Figures 2.

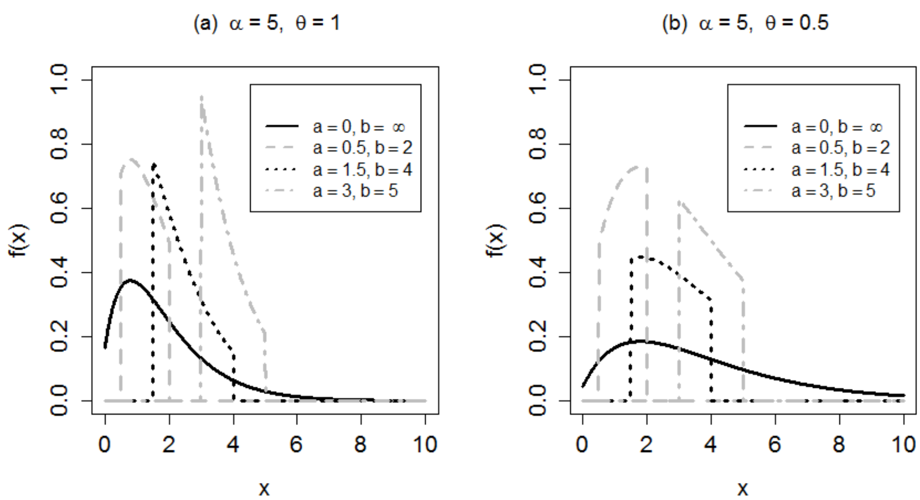

Figure 2. Plots of the pdf for the dT-PTL distribution on $[a,b]$

2) For $X \sim$dT-TPL(θ,α,a,b), when $b \to \infty$, which the distribution of X is called the left truncated two-parameter Lindley distribution, it is denoted as $X \sim$lT-TPL(θ,α,a), then the pdf and cdf of X are respectively

$$f(x;\alpha,\theta,a) = \frac{\theta^2}{\theta+\alpha+\alpha\theta a}(1+\alpha x)e^{-\theta(x-a)}, \text{ and } F(x;\alpha,\theta,a) = \frac{(\theta+\alpha)-(\theta+\alpha+\alpha\theta x)e^{-\theta x}}{(\theta+\alpha+\alpha\theta a)e^{-\theta a}}. \quad (6)$$

3) If $X \sim$dT-TPL(θ,α,a,b), and $a=0$, then the distribution of X is called the right truncated two-parameter Lindley distribution, it is denoted as $X \sim$rT-TPL(θ,α,b), then the pdf and cdf of X are respectively

$$f(x;\alpha,\theta,b) = \frac{\theta^2(1+\alpha x)e^{-\theta x}}{(\theta+\alpha)-(\theta+\alpha+\alpha\theta b)e^{-\theta b}}, \text{ and } F(x;\alpha,\theta,b) = \frac{(\theta+\alpha)-(\theta+\alpha+\alpha\theta x)e^{-\theta x}}{(\theta+\alpha)-(\theta+\alpha+\alpha\theta b)e^{-\theta b}}. \quad (7)$$

The pdf plots of the lT-TPL and rT-TPL distributions are shown in Figures 3-4, respectively.

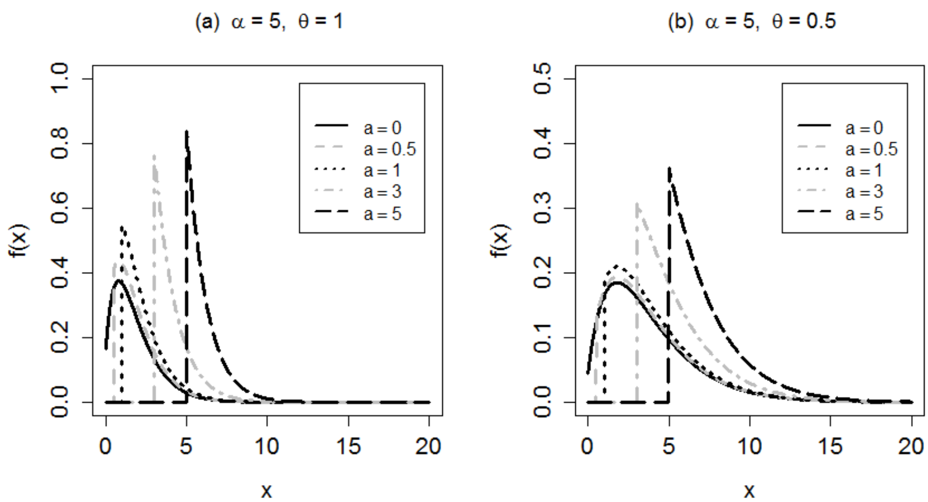

Figure 3. Plots of the pdf for the lT-TPL distribution at a

If $X \sim$T-TPL(θ,α,a,b), and $\alpha=1$, then the T-TPL distribution reduces to the truncated Lindley (T-Lindley) distribution (Singh et al., 2014) and when $\alpha=0$, it reduces to the truncated exponential (T-Exponential) distribution (Hannon & Dahiya, 1999).

Moments

The factorial moments about the origin of order kth are provided explicit expressions of $\mu'_k = E(X^k)$, where k is a positive integer. This can be used to study the most important characteristics of the distribution (e.g., mean, variance, skewness, kurtosis, etc.)

Vol. 17 No. 1: 19-32 [2018]
doi: 10.14416/j.appsci.2018.05.005

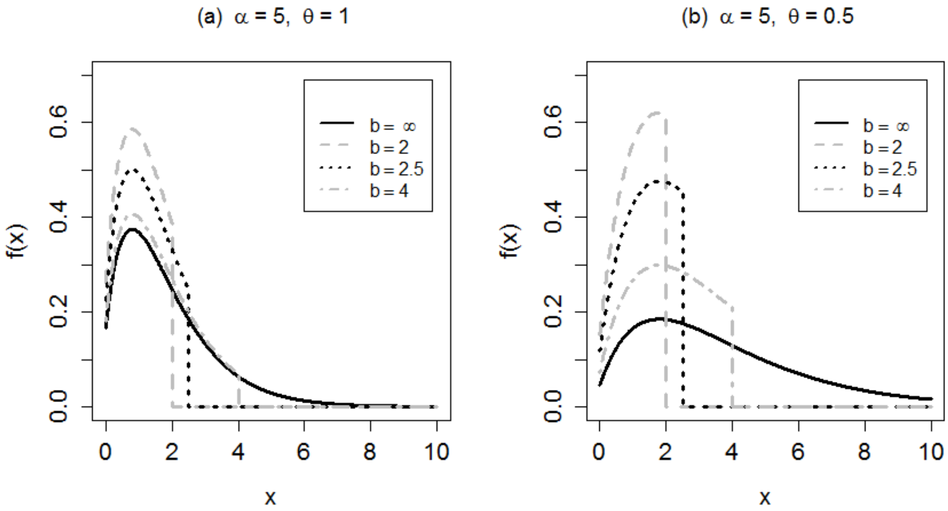

Figure 4. Plots of the pdf for the rT-TPL distribution at b

Theorem 1. If X ~dT-TPL (θ,α,a,b) then the kth moment of X is

$$\mu_k'(x;\theta,\alpha,a,b)=\frac{\theta\left[\gamma(k+1,b)-\gamma(k+1,a)\right]+\alpha\left[\gamma(k+2,b)-\gamma(k+2,a)\right]}{\theta^k\left[\left(\theta+\alpha+\alpha\theta a\right)e^{-\theta a}-\left(\theta+\alpha+\alpha\theta b\right)e^{-\theta b}\right]}, \tag{8}$$

$k=1,2,\ldots$ where $\gamma(k,b)=\int_0^b x^{k-1}e^{-x}dx$ is the lower incomplete gamma function.

Proof: Using equation (4), let $\delta_{(\theta,\alpha,a,b)}=(\theta+\alpha+\theta\alpha a)e^{-\theta a}-(\theta+\alpha+\theta\alpha b)e^{-\theta b}$, one can write

$$\mathrm{E}(\mathrm{X}^k)=\frac{\theta^2}{\delta_{(\theta,\alpha,a,b)}}\int_a^b x^k(1+\alpha x)e^{-\theta x}dx=\frac{\theta^2}{\delta_{(\theta,\alpha,a,b)}}\left[\int_a^b x^k e^{-\theta x}dx+\alpha\int_a^b x^{k+1}e^{-\theta x}dx\right].$$

By setting $u=\theta x$ and $du=\theta dx$, and based on $\gamma(k,b)$, one can write $\mathrm{E}(\mathrm{X}^k)$ as

$$\mu_k'(x;\theta,\alpha,a,b)=\frac{\theta^2}{\delta_{(\theta,\alpha,a,b)}}\left[\frac{\gamma(k+1,b)-\gamma(k+1,a)}{\theta^{k+1}}+\frac{\alpha\left(\gamma(k+2,b)-\gamma(k+2,a)\right)}{\theta^{k+2}}\right].$$

From *Theorem 1*, the mean and variance of X are respectively,

$$\mathrm{E}(\mathrm{X})=\frac{1}{\theta\delta_{(\theta,\alpha,a,b)}}\left[\theta\left(\gamma(2,b)-\gamma(2,a)\right)+\alpha\left(\gamma(3,b)-\gamma(3,a)\right)\right],$$

$$\mathrm{V}(\mathrm{X})=\frac{1}{\theta^2\delta^2_{(\theta,\alpha,a,b)}}\Big[\left(\theta\left(\gamma(3,b)-\gamma(3,a)\right)+\alpha\left(\gamma(4,b)-\gamma(4,a)\right)\right)\delta_{(\theta,\alpha,a,b)}$$
$$-\left(\theta\left(\gamma(2,b)-\gamma(2,a)\right)+\alpha\left(\gamma(3,b)-\gamma(3,a)\right)\right)^2\Big].$$

Theorem 2. If X ~lT-TPL (θ,α,a), then the kth moment of X is

 doi: 10.14416/j.appsci.2018.05.005

$$\mu_k'(x;\theta,\alpha,a)=\frac{\theta\Gamma(k+1,a)+\alpha\Gamma(k+2,a)}{\theta^k\left(\theta+\alpha+\alpha\theta a\right)e^{-\theta a}},\ k=1,2,... \qquad (9)$$

where $\Gamma(k,a)=\int_a^\infty x^{k-1}e^{-x}dx$ is the upper incomplete gamma function.

Proof: Using equation (6) and letting $\delta_{(\theta,\alpha,a)}=\left(\theta+\alpha+\alpha\theta a\right)e^{-\theta a}$, one can write

$$\mathrm{E}(\mathrm{X}^k)=\frac{\theta^2}{\delta_{(\theta,\alpha,a)}}\int_a^\infty x^k(1+\alpha x)e^{-\theta x}dx=\frac{\theta^2}{\delta_{(\theta,\alpha,a)}}\left[\int_a^\infty x^k e^{-\theta x}dx+\alpha\int_a^\infty x^{k+1}e^{-\theta x}dx\right].$$

By setting $u=\theta x,$ and based on $\Gamma(k,a),$ the $\mathrm{E}(\mathrm{X}^k)$ can be written as

$$\mu_k'(x;\theta,\alpha,a)=\frac{\theta^2}{\delta_{(\theta,\alpha,a)}}\left[\frac{\Gamma(k+1,a)}{\theta^{k+1}}+\frac{\alpha\Gamma(k+2,a)}{\theta^{k+2}}\right].$$

From *Theorem 2*, the mean and variance of X are, respectively,

$$\mathrm{E}(\mathrm{X})=\frac{1}{\theta\delta_{(\theta,\alpha,a)}}\left[\theta\Gamma(2,a)+\alpha\Gamma(3,a)\right],$$

$$\mathrm{V}(\mathrm{X})=\frac{1}{\theta^2\delta^2_{(\theta,\alpha,a)}}\left[\left(\theta\Gamma(3,a)+\alpha\Gamma(4,a)\right)\delta_{(\theta,\alpha,a)}-\left(\theta\Gamma(2,a)+\alpha\Gamma(3,a)\right)^2\right].$$

Theorem 3. If X ~rT-TPL (θ,α,b) then the $k\mathrm{th}$ moment of X is

$$\mu_k'(x;\theta,\alpha,b)=\frac{\theta\gamma(k+1,b)+\alpha\gamma(k+2,b)}{\theta^k\left[\left(\theta+\alpha\right)-\left(\theta+\alpha+\alpha\theta b\right)e^{-\theta b}\right]},\ k=1,2,... \qquad (10)$$

where $\gamma(k,b)$ is the lower incomplete gamma function.

Proof: Using equation (7) and letting $\delta_{(\theta,\alpha,b)}=\left(\theta+\alpha\right)-\left(\theta+\alpha+\alpha\theta b\right)e^{-\theta b}$, one can write

$$\mathrm{E}(X^k)=\frac{\theta^2}{\delta_{(\theta,\alpha,b)}}\int_0^b x^k(1+\alpha x)e^{-\theta x}dx=\frac{\theta^2}{\delta_{(\theta,\alpha,b)}}\left[\int_0^b x^k e^{-\theta x}dx+\alpha\int_0^b x^{k+1}e^{-\theta x}dx\right].$$

By setting $u=\theta x$ and based on $\gamma(k,b),$ then $\mathrm{E}(\mathrm{X}^k)$ is

$$\mu_k'(x;\theta,\alpha,b)=\frac{\theta^2}{\delta_{(\theta,\alpha,b)}}\left[\frac{\gamma(k+1,b)}{\theta^{k+1}}+\frac{\alpha\gamma(k+2,b)}{\theta^{k+2}}\right].$$

From *Theorem 3*, the mean and variance of X are

$$\mathrm{E}(\mathrm{X})=\frac{1}{\theta\delta_{(\theta,\alpha,b)}}\left[\theta\gamma(2,b)+\alpha\gamma(3,b)\right],$$

$$\mathrm{V}(\mathrm{X})=\frac{1}{\theta^2\delta^2_{(\theta,\alpha,b)}}\left[\left(\theta\gamma(3,b)+\alpha\gamma(4,b)\right)\delta_{(\theta,\alpha,b)}-\left(\theta\gamma(2,b)+\alpha\gamma(3,b)\right)^2\right].$$

Survival and hazard functions

The survival function, $\mathrm{S}(x),$ is the probability that a subject survives longer than time $x.$ Suppose X is a lifetime random variable representing the time until a spectified event of interest is occured, then the survival function of X is defined as $\mathrm{S}(x)=\mathrm{P}(\mathrm{X}>x)=1-\mathrm{F}(x).$ From the cdf in equation (5), the survival function of the dT-TPL distribution is

 doi: 10.14416/j.appsci.2018.05.005

$$\mathrm{S}(x) = 1 - \frac{(\theta+\alpha) - (\theta+\alpha+\theta\alpha x)e^{-\theta x}}{(\theta+\alpha+\theta\alpha a)e^{-\theta a} - (\theta+\alpha+\theta\alpha b)e^{-\theta b}},\ 0 < a \le x < b \le \infty. \quad (11)$$

Consequently, the ratio of the pdf, $\mathrm{f}(x)$, and the survival function, $\mathrm{S}(x)$, is given by $\mathrm{h}(x) = \mathrm{f}(x)/\mathrm{S}(x)$, which is called the hazard function. From the pdf in equation (4) and the survival function in equation (11), the hazard function of X is given by

$$\mathrm{h(x)} = \frac{\theta^2(1+\alpha x)e^{-\theta x}}{(\theta+\alpha+\theta\alpha a)e^{-\theta a} + (\theta+\alpha+\theta\alpha t)e^{-\theta t} - (\theta+\alpha+\theta\alpha b)e^{-\theta b} - (\theta+\alpha)}. \quad (12)$$

Plots of hazard in equation (12) and survival functions in equation (11) are shown in Figure 5.

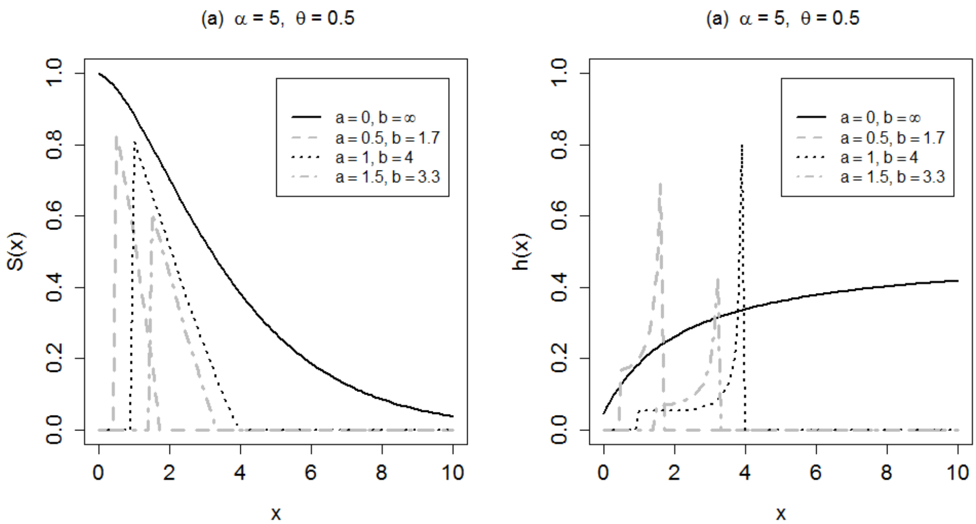

Figure 5. Plots of the survival and hazard functions for the dT-TPL distribution on $[a, b]$

Quantile function

The quantile function is specified by $\mathrm{q}(p)$ as the value at which the probability of the random variable is not more than the given probability p. It is also called the percent-point function or inverse cumulative distribution function, i.e., $\mathrm{q}(p) = \mathrm{F}^{-1}(x)$. We solve this equation by using the *Lambert-W* function (Jorda, 2010), which is a multivalued complex function defined as $\mathrm{W}(z)e^{W(z)} = z$ where z is a complex number. From the cdf in equations (5)-(7) and by defining $t = \theta/\alpha + 1$, the quantile function of X are as follows;

1) For X ~dT-TPL (θ, α, a, b),

$$\mathrm{q}(p) = -\frac{1}{\theta}\left\{W\left[\left(p(t+\theta a)e^{-\theta a} - p(t+\theta b)e^{-\theta b} - t\right)e^{-t}\right] + t\right\}. \quad (13)$$

2) For X ~lT-TPL (θ, α, a),

$$\mathrm{q}(p) = -\frac{1}{\theta}\left\{W\left[\left(p(t+\theta a)e^{-\theta a} - t\right)e^{-t}\right] + t\right\}. \quad (14)$$

3) For X ~rT-TPL (θ, α, b),

 doi: 10.14416/j.appsci.2018.05.005

$$\mathrm{q}(p) = -\frac{1}{\theta}\left\{W\left[\left(pt - p(t+\theta b)e^{-\theta b} - t\right)e^{-t}\right] + t\right\}. \quad (15)$$

The generation of a T-TPL random variate

The inverse transformation technique is used to generate a random variable for the TP_2L distribution by setting $x_i = \mathrm{F}^{-1}(u_i),$ where u is a random variable of the uniform distribution on (0,1), which denoted as $\mathrm{U}(0,1).$ The *LamW* package in R (Adler, 2016) is employed to solve the *Lambert-W* function in equations (13)-(15). Then $\mathrm{X}_i, i = 1,2,...,n,$ can be generated as follows:

1) Generate $u_i, i = 1,2,...,n$ from $\mathrm{U}(0,1).$
2) Set $t = \theta / \alpha + 1$.
3) Set w_i, when w_i is $W\{z\}$ and can be obtained by using "lambertWm1(z)" in R;
   3.1) $w_i = W\left[\left(u_i(t+\theta a)e^{-\theta a} - u_i(t+\theta b)e^{-\theta b} - t\right)e^{-t}\right]$ for dT-TPL $(\theta, \alpha, a, b),$
   3.2) $w_i = W\left[\left(u_i(t+\theta a)e^{-\theta a} - t\right)e^{-t}\right]$ for lT-TPL $(\theta, \alpha, a),$
   3.3) $w_i = W\left[\left(u_i t - u_i(t+\theta b)e^{-\theta b} - t\right)e^{-t}\right]$ for rT-TPL $(\theta, \alpha, b).$
4) Then, $x_i = -\frac{1}{\theta}\{w_i + t\}$.

The parameter estimation

In this section, author describes the procedure to obtain the maximum likelihood estimates (MLE) of the parameters of the dT-TPL, as well as, lT-TPL and rT-TPL distributions based on the random sample $\tilde{x} = (x_1, x_2, ..., x_n)$ of n observations so that these distributions can be effectively used to model real problems depending upon the nature of the data.

1) Let $\mathrm{X}_i, i = 1,2,...,n$ be an iid (independent and identically distributed). The likelihood function of X ~dT-TPL (θ, α, a, b) on the observed sample $\tilde{x}$ is given by

$$L(\alpha, \theta \mid \tilde{x}, a, b) = \left(\frac{\theta^2}{(\theta+\alpha+\theta\alpha a)e^{-\theta a} - (\theta+\alpha+\theta\alpha b)e^{-\theta b}}\right)^n \prod_{i=1}^{n}(1+\alpha x_i)\cdot e^{-\theta\sum_{i=1}^{n} x_i} .$$

The corresponding log-likelihood equation, $L_1 = \log L(\alpha, \theta \mid \tilde{x}, a, b)$ for $\hat{a} = \min(x)$ and $\hat{b} = \max(x)$ is

$$L_1 = 2n\log(\theta) - n\log[(\theta+\alpha+\theta\alpha a)e^{-\theta a} - (\theta+\alpha+\theta\alpha b)e^{-\theta b}] + \sum_{i=1}^{n}\log(1+\alpha x_i) - \theta\sum_{i=1}^{n} x_i.$$

To estimate the unknown parameters θ and α from the differentiating of the log-likelihood function, we obtained the partial derivatives of L_1 with respect to θ and $\alpha,$ i.e.,

$$\frac{\partial L_1}{\partial \theta} = \frac{2n}{\theta} - [\frac{(1-\theta a-\theta\alpha a^2)e^{-\theta a} - (1-\theta b-\theta\alpha b^2)e^{-\theta b}}{(\theta+\alpha+\theta\alpha a)e^{-\theta a} - (\theta+\alpha+\theta\alpha b)e^{-\theta b}}]n - \sum_{i=1}^{n} x_i, \quad (16)$$

$$\frac{\partial L_1}{\partial \alpha} = \sum_{i=1}^{n}\frac{\partial \log(1+\alpha x_i)}{\partial \alpha} - n[\frac{(1+\theta a)e^{-\theta a} - (1+\theta b)e^{-\theta b}}{(\theta+\alpha+\theta\alpha a)e^{-\theta a} - (\theta+\alpha+\theta\alpha b)e^{-\theta b}}]. \quad (17)$$

Similarly, the estimators of the lT-TPL and rT-TPL distributions can be obtained as follows.

 doi: 10.14416/j.appsci.2018.05.005

2) Let $\tilde{x}=(x_1,x_2,...,x_n)$ be a random sample of size n from the lT-TPL distribution. The log-likelihood function, $L_2=\log L(\alpha,\theta\,|\,\tilde{x},a),$ of X_i for $i=1,2,...,n,$ and $\hat{a}=\min(x)$ is

$$L_2=2n\log(\theta)-n\log(\theta+\alpha+\theta\alpha a)+\sum_{i=1}^{n}\log(1+\alpha x_i)-\theta\sum_{i=1}^{n}x_i+n\theta a.$$

Partial derivatives of L_2, this log likelihood function with respect to θ and α can be given by

$$\frac{\partial L_2}{\partial\theta}=\frac{2n}{\theta}-\frac{n(1+\alpha a)}{\theta+\alpha+\theta\alpha a}-\sum_{i=1}^{n}x_i+na, \tag{18}$$

$$\frac{\partial L_2}{\partial\alpha}=-\frac{n(1+\theta a)}{\theta+\alpha+\theta\alpha a}+\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\log(1+\alpha x_i). \tag{19}$$

3) Let X_i ~rT-TPL$(\theta,\alpha,b),$ $i=1,2,...,n$ be an iid random variable. The log-likelihood function of X_i is $L_3=\log L(\alpha,\theta\,|\,\tilde{x},b)$ for $\hat{b}=\max(x)$ given by

$$L_3=2n\log(\theta)-n\log[(\theta+\alpha)-(\theta+\alpha+\alpha\theta b)e^{-\theta b}]+\sum_{i=1}^{n}\log(1+\alpha x_i)-\theta\sum_{i=1}^{n}x_i.$$

Partial derivatives of this log likelihood function with respect to θ and α are given by

$$\frac{\partial L_3}{\partial\theta}=\frac{2n}{\theta}-\frac{n[1-(1-\theta b-\alpha\theta b^2)e^{-\theta b}]}{(\theta+\alpha)-(\theta+\alpha+\alpha\theta b)e^{-\theta b}}-\sum_{i=1}^{n}x_i, \tag{20}$$

$$\frac{\partial L_3}{\partial\alpha}=-\frac{n[1-(1-\alpha b-\alpha\theta b^2)e^{-\theta b}]}{(\theta+\alpha)-(\theta+\alpha+\alpha\theta b)e^{-\theta b}}+\sum_{i=1}^{n}\frac{\partial}{\partial\alpha}\log(1+\alpha x_i). \tag{21}$$

These differential equations in equations (16)-(21) are not in a closed-form expression. Therefore, the MLE estimator of the parameter estimate, i.e., $\hat{\theta}$ and $\hat{\alpha},$ can be obtained by using numerical optimization with the *nlm* function in the R program (R Core Team, 2016).

Table 1. Maximum likelihood estimates, SE, and MSE values under models based on simulated data (Parameters: $\theta=0.5$ and $\alpha=5$).

Distributions	n	$\theta=0.5$			$\alpha=5$		
		Estimates	SE	MSE	Estimates	SE	MSE
dT-TPL	20	0.2171	0.0513	0.1326	2.3787	2.1801	101.9256
(θ,α,a,b)	50	0.3591	0.0280	0.0592	3.6834	1.2353	78.0323
$a=1,b=5$	100	0.3936	0.0163	0.0378	4.2341	0.7496	56.7701
	200	0.4385	0.0083	0.0174	4.7218	0.4514	40.8250
	500	0.4780	0.0030	0.0049	4.9526	0.1494	11.1658
	1000	0.4893	0.0015	0.0025	5.0420	0.0860	7.3910
lT-TPL	20	0.6312	0.0202	0.0254	5.7802	3.9994	320.5091
(θ,α,a)	50	0.6280	0.0122	0.0238	6.1164	1.8965	181.0903
$a=1,b=\infty$	100	0.6326	0.0076	0.0234	6.4807	1.2412	156.2369
	200	0.6381	0.0038	0.0220	6.7860	0.7084	103.5645
	500	0.6427	0.0015	0.0214	7.3975	0.3425	64.3930
	1000	0.6447	0.0004	0.0211	7.5028	0.1533	29.7569
tT-TPL	20	0.2719	0.0506	0.1032	4.1138	2.7196	148.7085
(θ,α,b)	50	0.3691	0.0252	0.0489	4.7495	1.6198	131.2441
	100	0.4391	0.0127	0.0198	5.0256	1.0620	112.7745
$a=0,b=5$	200	0.4737	0.0060	0.0079	5.3574	0.6547	85.8650
	500	0.4938	0.0021	0.0023	5.6355	0.2384	28.8174
	1000	0.4971	0.0011	0.0012	5.8564	0.1027	11.2816

 doi: 10.14416/j.appsci.2018.05.005

Simulation study

Simulation study is illustrated to show the efficiency of maximum likelihood estimation (MLE). As an illustration, the sample data generated from a T-TPL random variable with the specified parameters of 3 cases has been displayed in Table 1 for the sample sizes (n) are 20, 50, 100, 200, 500, and 1,000. In each situation, the parameters are estimated via the MLE with *nlm* function in R (R Core Team, 2016) of 1000 replications. Suppose the estimated parameter of ω is $\hat{\omega}$, then the bias, standard deviation, and standard error of $\hat{\omega}$, are computed by the formulas: $\hat{\omega}=\sum_{t=1}^{1000}\hat{\omega}_t\big/1000$, $\mathrm{Bias}(\omega)=\hat{\omega}-\omega$, $\mathrm{SD}(\hat{\omega})=\sum_{t=1}^{1000}(\hat{\omega}_t-\hat{\omega})^2/999$, and $\mathrm{SE}(\hat{\omega})=\mathrm{SD}(\hat{\omega})/$, respectively. The mean squared error (MSE), i.e., $\mathrm{MSE}(\hat{\omega})=\mathrm{Var}(\hat{\omega})+\mathrm{Bias}^2(\hat{\omega})$, is used for the criteria to compare the efficiency of the MLE of each parameter.

From the results in Table 1, the MLE for estimating θ and α gives the minimum MSE value. The MSE of each estimated parameter trends to decrease when the sample size is increasing. Moreover, the estimate values of each parameter by using MLE method gives the biased estimate value from the parameter. However, the MLE method gives the estimate value that close to the parameter. Thus, the MLE method is used to estimate the parameters of the distributions for a real data in application study.

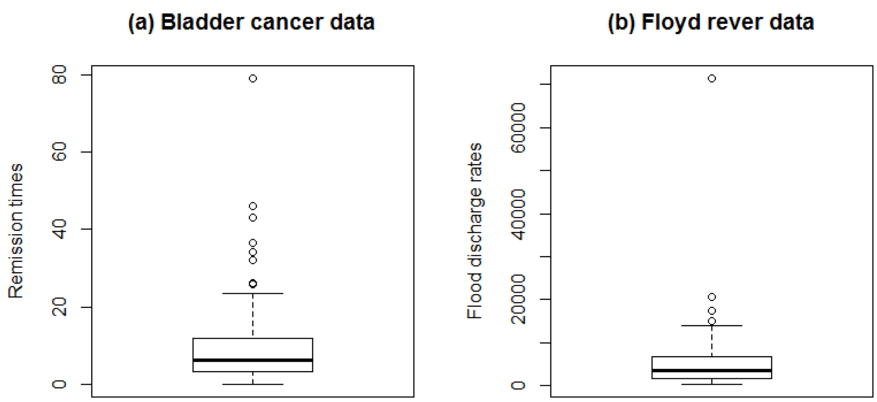

Figure 6. Box plots of the real datasets for application study

Application study

The first data set is the cancer patient data set, which is an uncensored data set corresponding to the remission times (in months) of a random sample of 128 bladder cancer patients. Thse data sets were previously studied by Lee and Wang in 2003 (Zea et al., 2012). The values of the minimum, mean, and maximum of the remission times are 0.09, 9.37, and 79.05, respectively. Another, real data set is the Floyd River flood rates for the years 1935–1973, which appeared in Akinsete et al. (2008). The minimum, mean, and maximum values of flood discharge rates are 318, 6771.10, and 71500, respectively. The summary of this data set is shown in Figure 6. Estimates of the parameters of the T-TPL distribution, Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (AICC), Bayesian Information Criterion (BIC), and KS test by using value of D are, respectively, given as follows:

$$\mathrm{AIC} = -2log(L) + 2p,$$
$$\mathrm{AICC} = \mathrm{AIC} + (2p(p+1))/(n-p-1),$$
$$\mathrm{BIC} = -2\log(L) + p\log(n),$$
$$\mathrm{D} = \sup_x |F_n(x) - F_0(x)|,$$

where n is the sample size and p is the number of parameters.

Based on the results in Tables 2-3, the exponential, Lindley, TPL, truncated (doubly, left, right) Lindley, and truncated (doubly, left, right) TPL distributions have been compared with the KS test, and it has been found that the rT-TPL $(\hat{\theta} = 0.1066,\ \hat{\alpha} = -3.580 \times 10^{-6},\ b = 79.05)$ gives a reasonable fit for the data of bladder cancer patients better than other distributions. And for real data of the Floyd River flood rates, the dT-TPL $(\hat{\theta} = 1.536 \times 10^{-4},\ \hat{\alpha} = -4.289 \times 10^{-7},$ $a = 318,\ b = 71{,}500)$ is the best fit for the flood discharge rates. Moreover, the KS-test of the exponential, Lindley, and TPL distributions are shown that they can not fit with the data.

Table 2. Maximum likelihood estimates and statistic values under models based on real data on the bladder cancer data (Zea et al., 2012).

Distributions	Maximum likelihood estimates				-logL	AIC	AICC	BIC	D (p-value)
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{a}$	$\hat{b}$					
Exponential (θ)	0.1068	-	-	-	414.342	830.684	830.716	833.536	0.9453 (<0.0001)
Lindley (θ)	0.1960	-	-	-	419.530	841.060	841.092	843.912	0.9453 (<0.0001)
dT-Lindley (θ, a, b)	0.1965	-	0.08	79.05	419.189	844.378	844.572	852.934	0.1144 (0.0703)
lT-Lindley (θ, a)	0.1965	-	0.08	-	422.530	849.060	849.156	854.764	0.1165 (0.0619)
rT-Lindley (θ, b)	0.1960	-	-	79.05	419.185	842.370	842.466	848.074	0.1141 (0.0713)
TPL (θ, α)	0.1068	1.1701×10^{-6}	-	-	414.342	832.684	832.78	838.388	0.9453 (<0.0001)
dT-TPL (θ, α, a, b)	0.1075	1.022×10^{-6}	0.08	79.05	413.318	834.636	834.961	846.044	0.0875 (0.2805)
lT-TPL (θ, α, a)	0.1077	-4.193×10^{-6}	0.08	-	415.844	837.688	837.882	846.244	0.0878 (0.2767)
rT-TPL (θ, α, b)	0.1066	-3.580×10^{-6}	-	79.05	413.314	832.628	832.822	841.184	0.0844 (0.3217)

Table 3. Maximum likelihood estimates and statistic values under models based on real data about the Floyd river flood rates for the years 1935-1973 (Akinsete et al., 2008)

Distributions	Maximum likelihood estimates				-logL	AIC	AICC	BIC	D (p-value)
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{a}$	$\hat{b}$					
Exponential (θ)	0.0001	-	-	-	382.996	767.992	768.100	769.656	1 (<0.0001)
Lindley (θ)	0.0003	-	-	-	392.649	787.298	787.406	788.962	1 (<0.0001)
dT-Lindley (θ, a, b)	0.0003	-	318	71500	392.485	790.970	791.656	795.961	0.2395 (0.0187)
lT-Lindley (θ, a)	0.0003	-	318	-	394.485	792.970	793.303	796.297	0.2395 (0.0187)
rT-Lindley (θ, b)	0.0003	-	-	71500	394.649	793.298	793.631	796.625	0.2406 (0.0180)
TPL (θ, α)	0.0017	0.3000	-	-	634.588	1,273.2	1,273.5	1,276.5	1 (<0.0001)
dT-TPL (θ, α, a, b)	1.536×10^{-4}	-4.289×10^{-7}	318	71500	381.121	770.242	771.419	776.896	0.1011 (0.7831)
lT-TPL (θ, α, a)	0.0003	0.0104	318	-	391.840	789.680	790.366	794.671	0.2288 (0.0282)
rT-TPL (θ, α, b)	0.0003	0.1001	-	71500	392.191	790.382	791.068	795.373	0.2395 (0.0188)

Conclusion

In this work, the doubly, left, and right truncated two-parameter Lindley distributions are proposed. Some characteristics of the proposed distributions (i.e., moments, and functions of survival, hazard, and quantile) are discussed. The unknown parameters of the proposed distributions are estimated by the maximum likelihood estimation. Two real data sets such as the bladder cancer patients and the flood discharge rates have been considered to show the usefulness of the proposed distributions. The results show that the proposed distributions provide a consistently better fit than the other distributions (i.e., the exponential, Lindley, two-parameter Lindley, and truncated-Lindley distributions). We hope that the proposed distributions will attract wider application in many areas such as engineering, economics, survival, lifetime data analysis, etc.

Acknowledgement

The author would like to express thanks to the Rajamangala University of Technology Thanyaburi for research funding.

References

Aban, I. B., Meerschaert, M. M., & Panorska, A. K. (2006). Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, *101*(473), 270-277. doi: 10.1198/016214505000000411

Adler, A. (2016). *Lambert-W Function*. Retrieved October 10, 2016, from https://bitbucket.org/aadler/lamw

 doi: 10.14416/j.appsci.2018.05.005

Ahmed, S. E., Castro-Kuriss, C., Flores, E., Leiva, V., & Sanhueza, A. (2010). A truncated version of the birnbaum-saunders distribution with an application in financial risk. *Pakistan Journal of Statistics*, *26*, 293-311.

Ahsanullah, M., Shakil, M., & Kibria, B. M. (2016). Characterizations of continuous distributions by truncated moment. *Journal of Modern Applied Statistical Methods*, *15*(1), 316-331.

Akinsete, A. Famoye, F., & Lee, C. (2008). The beta-Pareto distribution. *A Journal of Theoretical and Applied Statistics*, *42*(6), 547-563.

Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. Oxford: The Oxford University Press.

Gutenberg, R., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America, 34*(4), 185-188.

Hamasha M. M. (2017), A mathematical approximation of the left-sided truncated normal distribution using the cadwell approximation model. *Engineering, Technology & Applied Science Research, 7*(1), 1382-1386.

Hannon P. M. & Dahiya R. C. (1999), Estimation of parameters for the truncated exponential distribution. *Communications in Statistics-Theory and Methods, 28*(11), 2591-2612.

Iwueze I. S. (2007), Some implications of truncating the $N(1, \sigma^2)$ distribution to the left at zero, *Journal of Applied Sciences*, *7*(2), 189-195.

Johnson, A. C. (2001), *On the truncated normal distribution: Characteristics of singly- and doubly-truncated populations of application in management science* (Ph.D. Dissertation) Stuart Graduate School of Business, Illinois Institute of Technology, Illinois.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1 (2nd Ed.). New York: Wiley-Interscience.

Jorda, P. (2010). Computer generation of random variables with Lindley or Poisson−Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, *81*, 851-859.

Lindley, D. V. (1985). 'Fiducial distributions and Bayes' theorem. *Journal of the Royal Staistical Society, Series B, 20*(1), 102-107.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. Vienna, Austria. Retrieved October 10, 2016, from https://www.R-project.org/

Shanker, R., Sharma, S., & Shanker R. (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. *Applied Mathematics, 4*(2), 363-368. doi: 10.4236/am.2013.42056

Singh, S. K., Singh, U., & Sharma, V. K. (2014). The truncated Lindley distribution: inference and application. *Journal of Statistics Applications & Probability, 3*(2), 219-228.

Sun X. D., Tao X. X & Liu C. Q. (2013), Finite fault modelling for the Wenchuan earthquake using hybrid slip model with truncated normal distributed source parameters. *Applied Mechanics and Materials*, 256-259, 2161-2167.

Zaninetti, L., & Ferraro, M. (2008). On the truncated Pareto distribution with applications. *Central European Journal of Physics*, *6*(1), 1-6. doi: 10.2478/s11534-008-0008-2

Zea, L. M., Silva, T. B., Bourguignon, M., & Santos, A. M. (2012). The beta exponentiated Pareto distribution with application to bladder cancer susceptibility. *International Journal of Statistics and Probability, 1*(2), 8-19. doi:10.5539/ijsp.v1n2p8

Zhange, T., & Xie, M. (2011). On the upper truncated Weibull distribution and its reliability implications. *Reliability Engineering & System Safety, 96*(1), 194-200. doi: 10.1016/ j.ress.2010.09.004