

6-8-2021

A New Right-Skewed Upside Down Bathtub Shaped Heavy-tailed Distribution and its Applications

Sandeep Kumar Maurya

Central University of South Bihar, Gaya, India, sandeepmaurya.maurya48@gmail.com

Sanjay K. Singh

Banaras Hindu University, Varanasi, India

Umesh Singh

Banaras Hindu University, Varanasi, India, usingh_52@yahoo.co.in

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Maurya, S. K., Singh, S. K., & Singh, U. (2020). A New Right-Skewed Upside Down Bathtub Shaped Heavy-tailed Distribution and its Applications. *Journal of Modern Applied Statistical Methods*, 19(1), eP2888. <https://doi.org/10.22237/jmasm/1608552600>

This Emerging Scholar is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

EMERGING SCHOLAR

A New Right-Skewed Upside Down Bathtub Shaped Heavy-tailed Distribution and its Applications

Sandeep Kumar Maurya
Central University of S. Bihar
Gaya, India

Sanjay K. Singh
Banaras Hindu University
Varanasi, India

Umesh Singh
Banaras Hindu University
Varanasi, India

A one parameter right skewed, upside down bathtub type, heavy-tailed distribution is derived. Various statistical properties and maximum likelihood approaches for estimation purpose are studied. Five different real data sets with four different models are considered to illustrate the suitability of the proposed model.

Keywords: lifetime distribution, upside down bathtub, maximum likelihood estimate

Introduction

Lifetime distribution can be categorized into five broad classes according to their nature of failure rates, i.e., constant, decreasing, increasing, bathtub and upside down bathtub (UBT) types. The use of exponential distribution is very restricted, because its failure rate is constant and it is not applicable in most of the real situation, due to this reason, a number of lifetime models have been developed which have non-constant failure rate like Weibull, gamma, generalized exponential, Lindley distributions etc. The distributions have failure rate is either monotone or non-monotone according to that failure rate weather increasing/decreasing or bathtub/upside-down bathtub (Barlow et al., 1963; Barlow and Proschan, 1975; Deshpande and Suresh, 1990; Deshpande and Purohit, 2006). But generally, lifetime devices exhibit non-monotone type behavior in the real-life situations when the failure rate firstly increase and then decrease, say for example in the engineering field, quality of production of an item in a company due to untrained workers, hazard rate due to any serious disease in the medical field. Due to this reason, a

number of non-monotone type distributions with their applications have been developed by researchers. Hjorth (1980) proposed a model having all the type of failure rate except UBT, Rajarshi and Rajarsh (1988) proposed a review of bathtub failure models, Mudholkar and Srivastava (1993) proposed exponentiated Weibull distribution having all the type of failure rates, Xie and Lai (1996) proposed a bathtub nature additive Weibull model, Xie et al. (2002) proposed modified Weibull model having bathtub type of failure rate. The UBT failure rate distributions commonly appear in medical and biological fields like in lung cancer patient data (Bennett, 1983), in bladder cancer patient data (Efron, 1988) and in breast carcinoma patient data (Lamberson, 1974).

Most of the inverted probability distributions show the UBT shape of failure rate, see, generalized inverse Weibull distribution (de Gusmão et al., 2009), transmuted inverse exponential distribution (Oguntunde and Adejumo, 2014), transmuted inverse Weibull distribution (Khan et al., 2014), transmuted inverse Rayleigh distribution (Sharma et al., 2014), inverse Lindley (Sharma et al., 2015), generalized inverse Lindley distribution (Sharma et al., 2016), etc. Various new lifetime distributions were proposed by using incorporating additional parameters by any methods or transformation of variable and shown their suitability and compatibility over the existing distributions (see Marshall and Olkin, 1988; Gupta et al., 1998; Shaw and Buckley, 2009). Kumar et al. (2015) proposed a DUS transformation for generating new lifetime distribution. Motivated by them Maurya et al. (2016) proposed new method for proposing more flexible and new lifetime model having various shapes of failure rate and more applicable model. Here the noticeable point is that in all of the above-mentioned transformations, only DUS transformation does not include any additional parameter (see also Maurya et al., 2017, for more about DUS transformation).

The purpose of this study is to propose a new inverted probability model with UBT type of failure rate. A one parameter inverse exponential distribution in DUS transformation is considered. The cumulative distribution function (cdf) of inverse exponential distribution is defined as,

$$F(x) = e^{-\theta/x} \quad x; \theta > 0. \quad (1)$$

Let X be random variable having baseline cdf given in (1), then the cdf of our proposed distribution is given as:

$$F(x) = \frac{e^{e^{-\theta/x}} - 1}{e - 1} \quad x; \theta > 0 \quad (2)$$

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

and the probability density function (pdf) is

$$f(x) = \frac{\theta}{x^2(e-1)} e^{-\frac{\theta}{x}} e^{-\frac{\theta}{x}} \quad x; \theta > 0. \quad (3)$$

The associated hazard rate is

$$h(x) = \frac{\theta e^{-\theta/x} e^{-\theta/x}}{x^2(e - e^{-\theta/x})} \quad (4)$$

where e ($\exp(\cdot)$) stands for exponential function.

Shapes of the distribution and failure rate function

The shape of the distribution reflects the idea whether the distribution is symmetric or skewed. With the help of equations (2) and (3), the shape of the cdf and pdf of the proposed distribution are plotted for different value of the parameter θ , which is given in Figure 1.

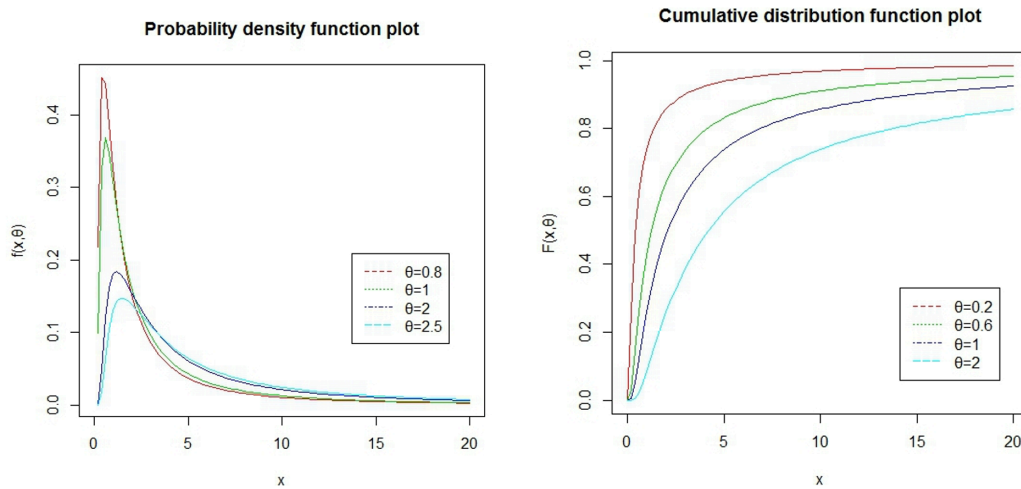


Figure 1. Probability density and cumulative distribution function plot

This figure shows that the proposed distribution exhibit right skewed model. By using equation (4) we can get an idea about the nature of its failure rate, we have plotted the shapes of failure rate for the various value of the parameter θ in Figure 2.

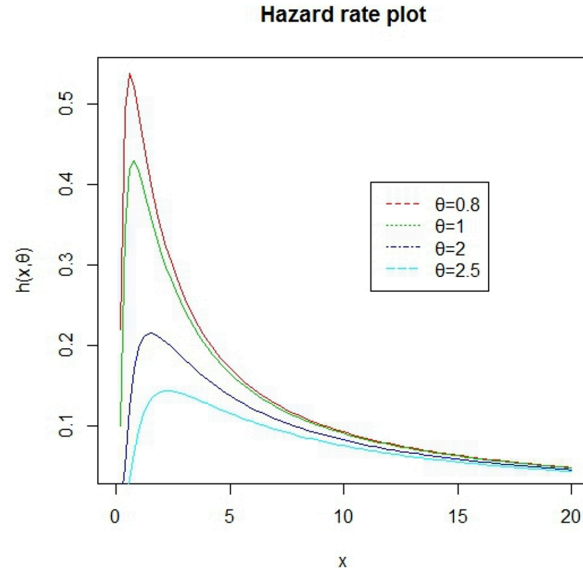


Figure 2. Hazard rate function plot for various choice of parameter

This figure shows that the proposed distribution has UBT type failure rate. This can also be verified mathematically by using the result of Glaser (1980). Glaser proved that if $\eta'(t) > 0$ for all $t \in (0, t_0)$, and $\eta'(t) = 0$ and $\eta'(t) < 0$ for all $t > t_0$ and satisfying $\lim_{x \rightarrow 0} f(t) = 0$ then distribution has upside down bathtub failure rate (UBT), where $\eta(t)$ is equal to $-f'(t)/f(t)$ and $f'(t)$ is the first order derivative of density function $f(t)$ with respect to t . In the case of our proposed distribution, we see that;

$$\eta(t) = \frac{2}{t} - \frac{\theta}{t^2} (e^{-\theta/t} + 1)$$

and

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

$$\eta'(t) = \frac{1}{(t^4)} \left[2\theta t e^{-\theta/t} + 2\theta t - 2t^2 - \theta^2 e^{-\theta/t} \right]. \quad (5)$$

Since the above equation is not in explicit form to get the solution, we use a simulation study and we find that for $t_0 = 1.227\theta$ approx $\eta'(t) > 0$ for all $t \in (0, t_0)$, and $\eta'(t) = 0$ and $\eta'(t) < 0$ for all $t > t_0$. Also, from equation (3) we can verify that $\lim_{x \rightarrow 0} f(t) = 0$ (because rate of convergence of exponential function is more than the algebraic function). Hence, the proposed distribution is right skewed distribution having UBT shape of failure rate.

Tail area property

A distribution is called heavy-tailed distribution if it possesses the heavy tail property. Klugman et al. (2012) suggested some criterion to detect the heavy tail property of a distribution (see also Nair et al., 2013). One of the important properties of heavy tailed distribution is that some or all order of the moments not exist. Therefore if the first moment i.e. arithmetic mean of the distribution does not exist then it possesses the heavy tail property of distribution. The arithmetic mean is derived for the distribution by solving

$$\frac{\theta}{e-1} \int_0^{\infty} \frac{1}{x} e^{-\theta/x} e^{e^{-\theta/x}} dx$$

by putting $t = e^{-\theta/x}$,

$$\frac{-\theta}{e-1} \int_0^1 \frac{e^t}{\log(t)} dt \quad (6)$$

which is a divergent equation. Hence, arithmetic mean of the distribution doesn't exist. Hence, by this criteria the proposed distribution is heavy-tailed distribution.

Another way is to examine the heavy tail property of distribution if the ratio of hazard rate to x goes to zero as x tends to infinity then it possesses the property of heavy-tailed distribution. For the proposed distribution,

$$\left(\frac{\theta e^{-\theta/x} e^{-\theta/x}}{e - e^{-\theta/x}} \right) \frac{1}{x^3} \rightarrow 0. \quad (7)$$

This can be proved by using L'Hôpital's rule. Thus, the proposed distribution has heavy-tailed distribution.

By the ratio of two survivals, the heavy tail property of distribution can also be checked. The distribution is said to be heavier than the other if the ratio of survivals goes to infinity as x tends to infinite. This indicates that the numerator puts highly significant value than the denominator. The limiting case of the ratio of two survival provides the limiting case of two probability density function. So, this ratio gives the ratio of two density functions i.e.

$$\lim_{x \rightarrow \infty} \frac{S_1(x)}{S_2(x)} = \lim_{x \rightarrow \infty} \frac{S_1'(x)}{S_2'(x)} = \lim_{x \rightarrow \infty} \frac{f_1(x)}{f_2(x)}$$

In this paper, we compare the density function with Pareto density (Pareto, 1964) with survival function $(\theta/x)^a$. Then the ratio of proposed density with Pareto distribution is given as

$$\frac{\theta e^{-\theta/x} e^{-\theta/x} (x)^{\alpha+1}}{x^2 (e-1) \alpha \theta^\alpha} \quad (8)$$

which goes to infinity as x tends to infinity for $a > 1$. Hence, the tail of the proposed distribution is heavier than the Pareto distribution. See also Foss et al. (2011) for more detail about heavy-tailed distributions.

A distribution is said to be heavy-tailed distribution if and only if $\int_0^\infty e^{\theta x} f(x) dx = \infty$ for all $\theta > 0$. Since, the first moments of proposed distribution is infinity, therefore this equation goes to infinity. Hence, the proposed distribution is heavy-tailed distribution. Hence, form the above discussion we can conclude that the proposed distribution is heavy tailed distribution.

Some statistical properties

Harmonic mean of the distribution

Because the proposed distribution is heavy-tailed distribution and the first moment does not exist, it is necessary to derive its inverse moment. The harmonic mean (H) of the distribution can be obtained by solving the given equation

$$\frac{1}{H} = \frac{\theta}{e-1} \int_0^{\infty} \frac{1}{x^3} e^{-\theta/x} e^{e^{-\theta/x}} dx. \quad (9)$$

By substituting $e^{-\theta/x} = t$, and after solving equation (9),

$$\frac{1}{H} = \frac{E_i(1) - \gamma}{\theta(e-1)} = \frac{1.3179}{\theta(e-1)}$$

where $\gamma(0.5772)$ is Euler-Mascheroni constant and $E_i(1)(1.8951)$ is exponential integral, which yields

$$H = 1.3038\theta \quad (10)$$

Quantile function of the distribution

The quantile function $Q(p)$ of p^{th} quantile is obtained by solving the equation $F(Q(p)) = p$. Hence, from the equation (2),

$$Q(p) = \frac{-\theta}{\log[\log(1 + p(e-1))]} \quad (11)$$

Because the mean of the proposed distribution does not exist, the mean cannot be used in any other expressions. Also, the quantile function may be used to evaluate the coefficients of skewness and kurtosis. Thus, Bowley's coefficient of skewness (Bowley, 1907) which is defined as

$$B = \frac{Q(\frac{3}{4}) + Q(\frac{1}{4}) - 2Q(\frac{1}{2})}{Q(\frac{3}{4}) - Q(\frac{1}{4})}$$

Its value is $0.4818 > 0$, this shows that the proposed distribution is positively skewed distribution. The Moors's coefficient of kurtosis (Moors, 1988) which is given as

$$M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}.$$

Its value is 2.1481 for the proposed distribution, which shows that it is more peaked than standard normal distribution.

Median of the proposed distribution

Since the proposed distribution is right skewed heavy tailed distribution, the median is the more suitable measure of central tendency. The median of the proposed distribution can be obtained by considering $p = 1/2$ in quantile function which is given in equation (11). We get the median (Md)

$$Md = \frac{-\theta}{\log\left[\log\left(0.5(e+1)\right)\right]} = 2.0927\theta \quad (12)$$

Mode of the distribution

The expression of the mode is obtained by solving the equation

$$\frac{d}{dx} f(x) = \frac{\theta e^{-\theta/x} e^{-\theta/x}}{x^4} \left[\theta e^{-\theta/x} + \theta - 2x \right] = 0. \quad (13)$$

Because the above equation is not in explicit form, the solution cannot be obtained directly. Through computations it was found this equation has an approximate solution for $x = 0.6\theta$. The approximate mode of the distribution is 0.6θ . It can be also checked by the Figure 3. In this figure, the red vertical line is drawn at $x = 0.6\theta$ which supports the above conclusion. Thus, for the proposed distribution, mean $>$ median $>$ mode. Hence, it is positively skewed distribution.

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

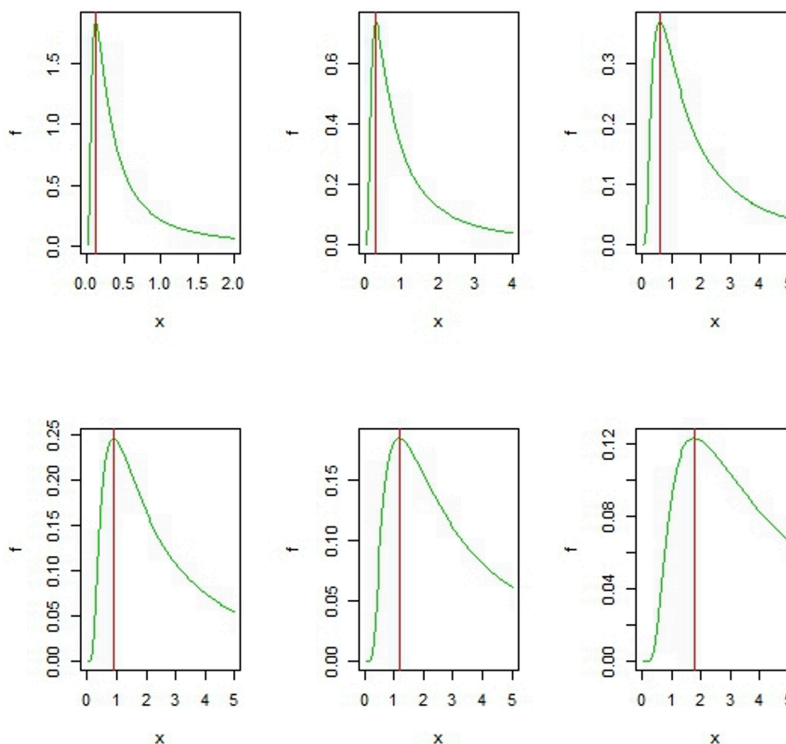


Figure 3. Mode plot for the various choice of parameter

Stochastic order

Let X_1 and X_2 be random variables having cdf and $F_2(x)$ respectively. Then X_1 is said to stochastically greater than X_2 if $F_1(x) \leq F_2(x)$ for all x (see Gupta et al., 1998 for more detail).

Theorem 1. Let X_1 and X_2 be the random variables of the proposed distribution, with parameter θ_1 and θ_2 respectively, then X_1 is stochastically greater than X_2 if $\theta_1 > \theta_2$.

Proof. From equation (2),

$$\frac{F_1(x)}{F_2(x)} = \frac{e^{-\theta_1/x} - 1}{e^{-\theta_2/x} - 1}$$

which will be always greater than 1, showing that X_1 is stochastically greater than X_2 for $\theta_1 > \theta_2$.

Distribution of order statistic

Take n random sample from the proposed distribution say, x_1, x_2, \dots, x_n and the corresponding order statistics is, $x_{1:1}, x_{2:2}, \dots, x_{n:n}$. Let $F(x)$ and $f(x)$ be the population cdf and pdf respectively, then for $r = 1, 2, \dots, n$, the pdf $f_r(x)$ of r^{th} order statistics $X_{r:n}$ is,

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} F^{r-1}(x) [1-F(x)]^{n-r} f(x).$$

Now by using equations (2) and (3) in above equation we have,

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} \frac{\theta e^{-\theta/x} e^{e^{-\theta/x}}}{x^2 (e-1)^{n-1}} (e^{e^{-\theta/x}} - 1)^{r-1} (e - e^{e^{-\theta/x}})^{n-r}. \quad (14)$$

And corresponding r^{th} order statistic of cdf $F_r(x)$ is,

$$F_r(x) = \sum_{i=r}^n {}^n C_i F^i(x) [1-F(x)]^{n-i} \quad (15)$$

Using equation (2) in equation (15) we have,

$$F_r(x) = \sum_{i=r}^n \sum_{j=0}^{n-i} {}^n C_i {}^{n-i} C_j (-1)^j \left[\left(e^{e^{-\theta/x}} - 1 \right) / e - 1 \right]^{i+j} \quad (16)$$

Shannon entropy for the distribution

Entropy is used to measure the average amount of information contained in random variable X . Shannon entropy proposed by Shannon (1951) and defined as $E[-\log f(x)]$. For the proposed distribution,

$$-\log f(x) = -\log \frac{\theta}{(e-1)} + 2 \log x + \frac{\theta}{x} - e^{-\theta/x} \quad (17)$$

and

$$E[-\log f(x)] = -\log \frac{\theta}{(e-1)} + 2E[\log x] + \theta E\left[\frac{1}{x}\right] - E[e^{-\theta/x}] \quad (18)$$

After solving the above equation, the value of Shannon entropy is obtained and given as,

$$E[-\log f(x)] = 2.5835 + \log \theta. \quad (19)$$

Estimation procedure for the parameter

Maximum likelihood estimation

Consider the maximum likelihood estimation (MLE) method for the estimation of the parameter θ of the proposed distribution which is obtained by maximizing the logarithm of the likelihood function. Let X_1, X_2, \dots, X_n be n random sample from the proposed distribution. Then, the logarithm likelihood function of the proposed distribution is,

$$\log L = n \log \frac{\theta}{(e-1)} + \sum_{i=1}^n \log \left(\frac{1}{x_i^2} \right) - \theta \sum_{i=1}^n \frac{1}{x_i} + \sum_{i=1}^n e^{-\theta/x_i}. \quad (20)$$

Differentiating it with respect to the parameter θ ,

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n \frac{1}{x_i} + \sum_{i=1}^n \frac{\theta}{x_i^2} e^{-\theta/x_i}. \quad (21)$$

Equating the equation (22) to zero, the likelihood equation is

$$\frac{n}{\theta} = \sum_{i=1}^n \frac{1}{x_i} - \sum_{i=1}^n \frac{\theta e^{-\theta/x_i}}{x_i^2} \quad (22)$$

which is a non-linear equation. After solving this, the MLE $\hat{\theta}$ of parameter θ is obtained. Because it is not in closed form it cannot be solved analytically.

Therefore, a numerical technique, such as the Newton-Raphson method (R Core Team, 2013) can be used. The confidence intervals are obtained based on the diagonal elements of Fisher information matrix $I^{-1}(\hat{\theta})$ which provides the estimated asymptotic variance for the parameter θ . Thus, the two-sided $100(1 - \eta)\%$ confidence interval of θ can be defined as $Z_{\eta/2} \sqrt{\text{var}(\hat{\theta})}$, where $Z_{\eta/2}$ stands for the upper $\eta/2\%$ points of standard normal distribution.

The Fisher Information matrix can be estimated by,

$$I(\hat{\theta}) = \left[\frac{-\partial^2 \log L}{\partial \theta^2} \right]_{\hat{\theta}} \quad (23)$$

where

$$\frac{-\partial^2 \log L}{\partial \theta^2} = -\frac{n}{\theta^2} + \sum_{i=1}^n \frac{(x_i - \theta)}{x_i^3} e^{-\theta/x_i}. \quad (24)$$

Real data application

Five different real data sets have been considered, out of which three have UBT nature of failure rate and last two have heavy tail nature of failure rate, and four different lifetime model out of which three have UBT nature and one has heavy-tailed property, for real data illustration. Also, two models have two parameters and rest two have the single parameter like the proposed model. The cdfs of the considered models are given below:

1. Inverse Exponential (IE) distribution have UBT type of failure rate and its cdf is given as,

$$F(x) = \exp(-\theta/x) \quad x > 0, \theta > 0.$$

2. Inverse Lindley (IL) distribution (Sharma et al., 2015). It has also UBT type of failure rate.

$$F(x) = e^{-\theta/x} \left(\frac{1 + \theta + \theta x}{(1 + \theta)x} \right) \quad x > 0, \theta > 0.$$

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

3. Inverse Weibull (IW) distribution with cdf:

$$F(x) = \exp\left(-\theta / x^\alpha\right) \quad x > 0, \theta > 0, \alpha > 0.$$

4. Pareto distribution (Pareto, 1964). This distribution is possessed heavy tail property.

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^\alpha \quad x > \theta, \theta > 0, \alpha > 0.$$

The various criterion like p-value, AIC (Akaike Information Criterion) and BIC (Bayesian information criterion) are used to check the fitting of the distributions. Also we have calculated the negative of log likelihood value ($-\text{LogL}$) and KS (Kolmogorov-Smirnov) test statistic. The AIC and BIC are defined as,

$$AIC = 2 * k - 2 * \log \hat{L}, \quad BIC = k * \log(n) - 2 * \log \hat{L},$$

and KS test statistics (D) is defined as

$$D = \text{Sup}_x \left| F_n(x) - F(x) \right|, \quad \text{where } F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{x_i \leq x}.$$

where $F_n(x)$ is empirical distribution function, $F(x)$ is cdf, n is sample size, k is a number of parameters and \hat{L} is maximum likelihood value for the considered distribution. First, consider the p-value for checking which models are fitted to the considered data set and after that we calculate the other mention criterion to know which model is more suitable for the data set among the fitted model. The smaller value of AIC, BIC, KS test statistic and $-\text{LogL}$ values indicate a better fit of distributions. The MLEs of parameters for various distributions were also calculated and compiled in in Table 1. The real data descriptions are given below;

Dataset 1. Flood level data. This dataset contained 39 observation of annual flood discharge rates of the Floyd River (located in James, Iowa, USA) for the years 1935-1973 and the dataset was taken by Mudholkar and Hutson (1996) and Merovci and Puka (2014).

Dataset 2. Head and Neck cancer disease data using treatment RT+CT. This dataset contains 44 observations survival times of a group of patients suffering from Head and Neck cancer disease and treated using a combination of radiotherapy and chemotherapy (RT+CT) Proposed by Efron (1988). This dataset was analyzed by Shanker et al. (2015), Sharma et al. (2015) used this dataset for IL and compared with inverse Rayleigh distribution, latter by Maurya et al. (2018).

Dataset 3. Vinyl Chloride data. This dataset contains 34 observations of vinyl chloride data given by Bhaumik et al. (2009) from clean up gradient monitoring wells in mg/l. This dataset analyzed by Barreto-Souza and Bakouch, (2013) and Maurya et al. (2018) for decreasing failure rate models.

Dataset 4. Sieve Diameter data. This dataset contains 13 observations of the mass-size distribution of a sand sample, determined by sieving given in mm and obtained by Bagnold (1954). Latter used by Barndorff-Nielsen (1977).

Dataset 5. Observed number of diamond data. This dataset contains 25 observations of a size distribution of diamond from a large mining area in South Africa, proposed by Sichel (1973) and latterly used by Barndorff-Nielsen (1977).

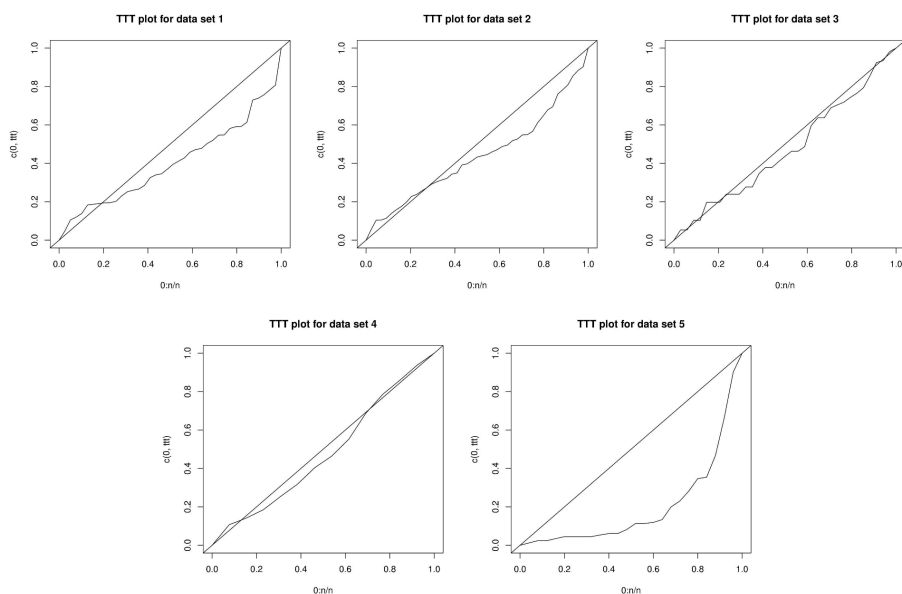


Figure 4. TTT plot of considered data sets

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

These data sets were considered because datasets 1-3 are UBT, data set 4 shows a UBT nature of failure rate and along with dataset 5 are used for heavy-tailed distributions (see Barndorff-Nielsen, 1977). The nature of datasets is plotted in Figure 4.

The curve above the abline and then below shows the UBT nature (for more detail about TTT plot see Aarset, 1987). A detailed description of Table 1 follows the table.

Table 1. MLE, -Log Likelihood, AIC, BIC and KS statistic with p-value for fitted datasets.

<i>Dataset 1.</i>		ML Estimate		KS Test		AIC	BIC
Distributions	$\hat{\alpha}$	$\hat{\theta}$	-LogL	Statistic	p-value		
IE	-	2166.2550	377.9945	0.0864	0.9089	757.9890	759.6526
Proposed	-	1627.5820	377.9722	0.1037	0.7573	757.9444	759.6080
IL	-	2167.2380	377.9945	0.0864	0.9089	757.9890	759.6526
IW	1.0145	2405.5306	377.9869	0.0862	0.9101	759.9737	763.3008
P	0.4125	318.0000	392.8099	0.3380	0.0002	789.6199	792.9470

<i>Dataset 2.</i>		ML Estimate		KS Test		AIC	BIC
Distributions	$\hat{\alpha}$	$\hat{\theta}$	-LogL	Statistic	p-value		
IE	-	76.70060	279.57730	0.08890	0.84770	561.15460	562.93880
Proposed	-	57.95910	279.43010	0.09030	0.83440	560.86020	562.64440
IL	-	77.67550	279.57840	0.08890	0.84740	561.15680	562.94090
IW	1.01330	80.76250	279.57010	0.09270	0.81040	563.14030	566.70860
P	0.42630	12.20000	294.77720	0.29380	0.00070	593.55450	597.12290

<i>Dataset 3.</i>		ML Estimate		KS Test		AIC	BIC
Distributions	$\hat{\alpha}$	$\hat{\theta}$	-LogL	Statistic	p-value		
IE	-	0.5725	59.1930	0.1470	0.4544	120.3860	121.9124
Proposed	-	0.4462	57.9264	0.1081	0.8222	117.8528	119.3792
IL	-	0.8774	61.8136	0.1908	0.1683	125.6272	127.1535
IW	0.8804	0.6539	58.6266	0.1134	0.7745	121.2532	124.3059
P	0.4177	0.1000	66.7972	0.3219	0.0017	137.5943	140.6471

<i>Dataset 4.</i>		ML Estimate		KS Test		AIC	BIC
Distributions	$\hat{\alpha}$	$\hat{\theta}$	-LogL	Statistic	p-value		
IE	-	0.1871	4.6046	0.1137	0.9890	11.2092	11.7742
Proposed	-	0.1433	4.5903	0.1314	0.9569	11.1807	11.7456
IL	-	0.3279	5.4029	0.1990	0.6131	12.8059	13.3708
IW	1.0257	0.1772	4.5976	0.1230	0.9756	13.1951	14.3251
P	0.5689	0.0540	5.2402	0.2073	0.5622	14.4803	15.6102

<i>Dataset 5.</i>		ML Estimate		KS Test		AIC	BIC
Distributions	$\hat{\alpha}$	$\hat{\theta}$	-LogL	Statistic	p-value		
IE	-	2.8704	103.2788	0.2003	0.2683	208.5576	209.7765
Proposed	-	2.2910	102.3038	0.1677	0.4833	206.6077	207.8265
IL	-	3.5073	103.6401	0.2066	0.2361	209.2803	210.4992
IW	0.7091	2.4559	100.5521	0.1160	0.8898	205.1042	207.5420
P	0.3589	0.5000	102.9349	0.1677	0.4833	209.8698	212.3076

For Dataset 1. Observed that among all the considered model only Pareto not fit this data set at 5% level of significance. The KS statistics is minimum for IW distribution but the value of $-\text{LogL}$ is least for our proposed distribution. The model selection criterion i.e. AIC and BIC both are least for our proposed model. Also, Merovci and Puka (2014) analyzes this data set for transmuted Pareto (TP) distribution (with $\text{AIC} = 776.698$, $\text{BIC} = 778.025$) and compared it with Pareto, Generalized Pareto, exponentiated Pareto distributions. Therefore, we can say that our proposed model is best among the considered models along with TP and others.

For Dataset 2. The table shows that only Pareto model is not fitted at 5% level of significance and KS statistics is least for IE. But the $-\text{LogL}$ value and model selection criterion (AIC and BIC both) are least for our proposed model among all the fitted models. Maurya et al. (2018) used this data set (with $\text{AIC} = 564.8111$ and $\text{BIC} = 566.5953$) and shown that their proposed model is fitted best to this data set in comparison to other seven lifetime models like Exponential Poisson Lindley, Extension of Exponential, Generalized Lindley, Weibull, Gamma, Exponentiated Exponential and exponentiated Binomial distributions. Hence, it can be concluded the proposed model is best among these models for this data set.

For Dataset 3. In the consideration of all the models, again only Pareto not fitted at desired level of significance i.e. 5%. The proposed distribution has minimum KS statistics along with $-\text{LogL}$. Also, the model selection criterion AIC and BIC are least for our proposed model. It can also be concluded the proposed model explains this data set better than the other models.

For Dataset 4. All the considered models fitted this data set at 5% level of significance and IE has minimum KS statistics. But the model selection criterion and $-\text{LogL}$ values are least for our proposed model. The model is best among all of the competitors for this data set.

For Dataset 5. Again all the considered models fit this data set at a desired level of significance. The $-\text{LogL}$, AIC, BIC is least for IW model and also for the proposed model at second place because our proposed model has only UBT nature and the data set shows decreasing nature. The proposed model is comparable to the IW model. The reason behind the fitting of Pareto model to last two datasets and not fitting in first three datasets is the dataset 4 and 5 are suitable for the heavy-tailed distributions.

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

Table 2. Interval estimates of the parameter for real datasets.

Distribution	Parameter	Data 1	Data 2	Data 3	Data 4	Data 5
IE	θ	1486.306, 2846.205	54.035, 99.366	0.380, 0.765	0.085, 0.289	1.745, 3.996
Proposed	θ	1058.445, 2196.719	38.926, 76.992	0.282, 0.610	0.056, 0.231	1.292, 3.291
IL	θ	1487.305, 2847.206	55.006, 100.345	0.656, 1.098	0.199, 0.456	2.343, 4.672
IW	θ	1650.451, 3160.536	56.896, 104.628	0.434, 0.874	0.081, 0.274	1.494, 3.419
	α	0.972, 1.057	0.941, 1.086	0.7020, 1.059	0.807, 1.244	0.502, 0.916
P	θ	317.890, 318.110	11.639, 12.761	0.0, 6.298	0.0, 8.489	0.0, 3.272
	α	0.283, 0.542	0.301, 0.552	0.277, 0.558	0.260, 0.878	0.218, 0.499

Table 2 shows the 95 % confidence interval of the parameter(s) for the real data sets and the considered models. The non-parametric empirical cdf (ecdf) plot, kernel density (KD) plot, histogram plot, and fitted density plot were used to show the suitability of the model graphically. The ecdf plot for all the considered models and datasets are given in Figure 5 and these figures also support our conclusions.

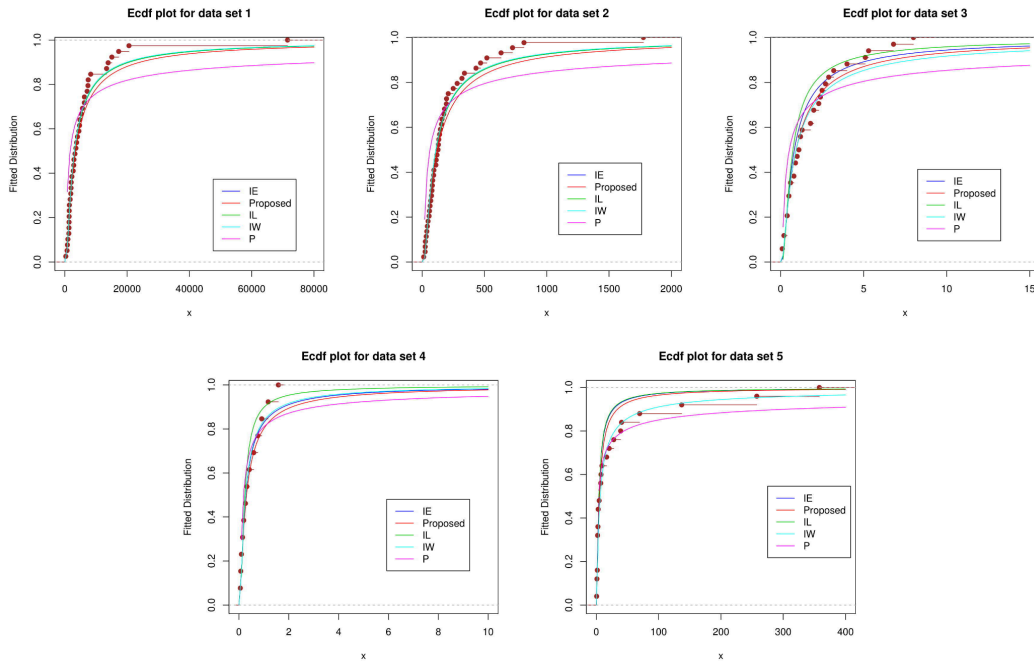


Figure 5. Ecdf plot of considered data sets

The fitted density plot, KD plot and relative histogram plots for all the five data sets are plotted in Figure 6.

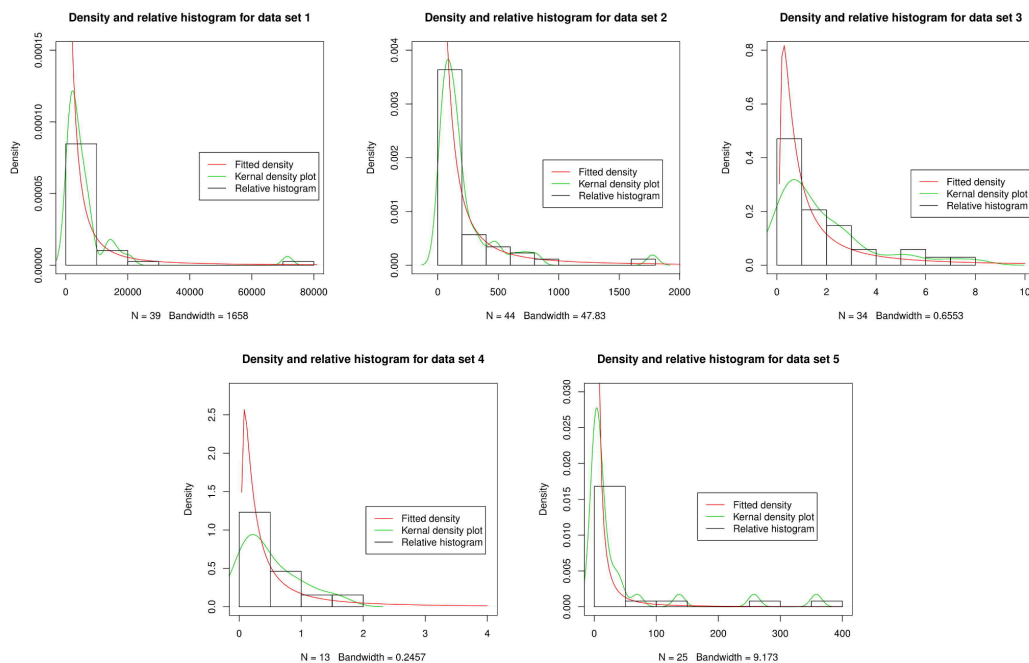


Figure 6. Fitted density, Relative histogram and Kernel density plots for proposed distribution.

Conclusion

A new right skewed lifetime model having upside down failure rate nature, which also has heavy-tailed nature. Some statistical properties like median, harmonic mean, model, Shannon entropy, pdf and cdf of r^{th} order statistics, quantile function were derived. The parameters of the proposed model were estimated through the maximum likelihood methods and $100(1 - \eta)\%$ confidence interval. For the real data illustration, the inverse exponential, inverse Lindley, inverse Weibull and Pareto distributions were considered, and five different real data sets in which three are UBT type nature and two have heavy-tailed, and the proposed model fitted well for all them. Therefore, this model can be recommended.

References

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, R-36(1), 106-108. <https://doi.org/10.1109/tr.1987.5222310>

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

- Bagnold, R. A. (1954). *The physics of blown sand and desert dunes*. Courier Corporation.
- Barlow, R. E., Marshall, A. W., and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. *The Annals of Mathematical Statistics*, 34(2), 375-389. <https://doi.org/10.1214/aoms/1177704147>
- Barlow, R. E. and Proschan, F. (1975). *Statistical theory of reliability and life testing: probability models*. (Technical report). Tallahassee, FL: Florida State University.
- Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 353(1674), 401-419. <https://doi.org/10.1098/rspa.1977.0041>
- Barreto-Souza, W. and Bakouch, H. S. (2013). A new lifetime model with decreasing failure rate. *Statistics*, 47(2), 465-476. <https://doi.org/10.1080/02331888.2011.595489>
- Bennett, S. (1983). Log-logistic regression models for survival data. *Applied Statistics*, 32(2), 165. <https://doi.org/10.2307/2347295>
- Bhaumik, D. K., Kapur, K., and Gibbons, R. D. (2009). Testing parameters of a gamma distribution for small samples. *Technometrics*, 51(3), 326-334. <https://doi.org/10.1198/tech.2009.07038>
- Bowley, A. L. (1907). *Elements of statistics* (3rd ed.). P. S. King & Son. <https://doi.org/10.1037/13657-000>
- Deshpande, J. V. and Purohit, S. G. (2006). *Lifetime Data: Statistical Models and Methods*. (Series on Quality, Reliability and Engineering Statistics, Vol. 11). World Scientific Publishing. <https://doi.org/10.1142/5988>
- Deshpande, J. V. and Suresh, R. P. (1990). Non-monotonic ageing. *Scandinavian Journal of Statistics*, 17(3), 257-262.
- de Gusmão, F. R. S., Ortega, E. M. M., and Cordeiro, G. M. (2009). The generalized inverse Weibull distribution. *Statistical Papers*, 52(3), 591-619. <https://doi.org/10.1007/s00362-009-0271-3>
- Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American Statistical Association*, 83(402), 414-425. <https://doi.org/10.1080/01621459.1988.10478612>
- Foss, S., Korshunov, D., Zachary, S. (2011). *An introduction to heavy-tailed and subexponential distributions*. New York: Springer. <https://doi.org/10.1007/978-1-4419-9473-8>

Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75(371), 667-672.

<https://doi.org/10.1080/01621459.1980.10477530>

Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modeling failure time data by lehmann alternatives. *Communications in Statistics - Theory and Methods*, 27(4), 887-904. <https://doi.org/10.1080/03610929808832134>

Hjorth, U. (1980). A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates. *Technometrics*, 22(1), 99–107.

<https://doi.org/10.2307/1268388>

Khan, M. S., King, R., and Hudson, I. (2014). Characterizations of the transmuted inverse Weibull distribution. *ANZIAM Journal*, 55(EMAC2013), 197–217.

<https://doi.org/10.21914/anziamj.v55i0.7785>

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions* (4th Ed.). NY: John Wiley & Sons.

Kumar, D., Singh, U., and Singh, S. K. (2015). A method of proposing new distribution and its application to bladder cancer patient data. *Journal of Statistics Applications and Probability Letters*, 2(3), 235–245.

Lamberson, L. R. (1974). An evaluation and comparison of some tests for the validity of the assumption that the underlying distribution of life is exponential. *AIIE Transactions*, 6(4), 327–337. <https://doi.org/10.1080/05695557408974971>

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403), 834-841.

<https://doi.org/10.1080/01621459.1988.10478671>

Maurya, S. K., Kaushik, A., Singh, S. K., and Singh, U. (2016). A new class of distribution having decreasing, increasing, and bathtub-shaped failure rate.

Communications in Statistics - Theory and Methods, 46(20), 10359-10372.

<https://doi.org/10.1080/03610926.2016.1235196>

Maurya, S. K., Kaushik, A., Singh, S. K., and Singh, U. (2017). A new class of exponential transformed Lindley distribution and its application to yarn data.

International Journal of Statistics & Economics, 18(2), 135–151.

Maurya, S. K., Kumar, D., Singh, S. K., and Singh, U. (2018). One parameter decreasing failure rate distribution. *International Journal of Statistics & Economics*, 19(1), 120–138.

Merovci, F. and Puka, L. (2014). Transmuted Pareto distribution. *ProbStat Forum*, 7(1), 1–11. <https://doi.org/>

A UBT SHAPED HEAVY-TAILED DISTRIBUTION

- Moors, J. J. A. (1988). A quantile alternative for kurtosis. *The Statistician*, 37(1), 25–32. <https://doi.org/10.2307/2348376>
- Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated Weibull family: some properties and a flood data application. *Communications in Statistics - Theory and Methods*, 25(12), 3059-3083. <https://doi.org/10.1080/03610929608831886>
- Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2), 299-302. <https://doi.org/10.1109/24.229504>
- Nair, J., Wierman, A., and Zwart, B. (2013). The fundamentals of heavy-tails: properties, emergence, and identification. *ACM SIGMETRICS Performance Evaluation Review*, 41(1), 387-388. <https://doi.org/10.1145/2494232.2466587>
- Oguntunde, P. and Adejumo, O. (2014). The transmuted inverse exponential distribution. *International Journal of Advanced Statistics and Probability*, 3(1), 1–7. <https://doi.org/10.14419/ijasp.v3i1.3684>
- Pareto, V. (1964). *Cours d'économie politique* (Vol. 1). Librairie Droz. <https://doi.org/10.3917/droz.paret.1964.01>
- Rajarshi, S. and Rajarshi, M. B. (1988). Bathtub distributions: A review. *Communications in Statistics - Theory and Methods*, 17(8), 2597-2621. <https://doi.org/10.1080/03610928808829761>
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. [Computer software]. R Foundation for Statistical Computing, Vienna, Austria.
- Shanker, R., Hagos, F., and Sujatha, S. (2015). On modeling of lifetimes data using exponential and Lindley distributions. *Biometrics and Biostatistics International Journal*, 2(5), 1–9. <https://doi.org/10.15406/bbij.2015.02.00042>
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 30(1), 50-64. <https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>
- Sharma, V. K., Singh, S. K., Singh, U., and Agiwal, V. (2015). The inverse lindley distribution: A stress-strength reliability model with application to head and neck cancer data. *Journal of Industrial and Production Engineering*, 32(3), 162-173. <https://doi.org/10.1080/21681015.2015.1025901>
- Sharma, V. K., Singh, S. K., Singh, U., and Merovci, F. (2016). The generalized inverse Lindley distribution: A new inverse statistical model for the study of upside-down bathtub data. *Communications in Statistics - Theory and Methods*, 45(19), 5709-5729. <https://doi.org/10.1080/03610926.2014.948206>

- Sharma, V. K., Singh, S. K., and Singh, U. (2014). A new upside-down bathtub shaped hazard rate model for survival data analysis. *Applied Mathematics and Computation*, 239(15 July 2014), 242–253. <https://doi.org/10.1016/j.amc.2014.04.048>
- Shaw, W. T. and Buckley, I. R. C. (2009). The alchemy of probability distributions: Beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv.org*. <https://arxiv.org/abs/0901.0434/>
- Sichel, H. S. (1973). Statistical valuation of diamondiferous deposits. *Journal of the Southern African Institute of Mining and Metallurgy*, 73(7), 235–243.
- Xie, M. and Lai, C. D. (1996). Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 52(1), 87-93. [https://doi.org/10.1016/0951-8320\(95\)00149-2](https://doi.org/10.1016/0951-8320(95)00149-2)
- Xie, M., Tang, Y., and Goh, T. N. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 76(3), 279-285. [https://doi.org/10.1016/s0951-8320\(02\)00022-4](https://doi.org/10.1016/s0951-8320(02)00022-4)