Cairo University

The 57th Annual International Conference of Data science

Department of Mathematical Statistics

9 – 11, Dec, 2024

Index

1	**Length-biased new Sushila Distribution** M. R. Mahmoud , I. B. Abdul - Moniem and S. S. Naguib	1 - 10

Length-biased new Sushila Distribution

M. R. Mahmoud*, I. B. Abdul-Moniem and S. S. Naguib***

* Department of Mathematical Statistics Faculty of Graduate Studies for Statistical Research, Cairo University

** Department of Statistics, Higher Institute of Management Sciences in Sohag, Sohag, Egypt

Abstract:

In this paper, we introduce a new distributions called Length-Biased New Sushila Distribution (LBNSD). Some properties of this distribution will be discussed. The estimation of unknown parameters for LBNSD will be handled using Maximum Likelihood method. Finally, an application to real data sets is illustrated.

Key words: Weighted distribution - New Sushila distribution– Moments – Maximum Likelihood Estimation.

1. Introduction

The first appear of the concept "weighted distributions" can be traced to Fisher (1934). Rao (1965). identified various situations that can be modeled by weighted distributions.

Let X be a non-negative random variable with probability density function (*pdf*) $g(x)$. The *pdf* of the weighted random variable X is given by

$$f(x) = \frac{w(x)g(x)}{E\left[w(X)\right]},\ x > 0$$

where $w(x)$ be a non-negative weight function.

When $w(x) = x$, the distribution is called length-biased, whose *pdf* is

$$f(x) = \frac{xg(x)}{E(X)},\ x > 0 \tag{1}$$

The formula (1) is used by many authors. Shaban and Boudrissa (2007) discussed the Weibull length biased distribution with properties and estimation. The length biased weighted generalized Rayleigh distribution is introduced by Das and Roy (2011). Seenoi et al (2014) discussed the length biased exponentiated inverted Weibull distribution. The length biased weighted Lomax distribution, statistical properties and application are introduced by Afaq et al (2016). Abdul-Moniem and Diab (2018) introduced Length-Biased weighted Exponentiated Lomax Distribution. Length Biased Sushila distribution handled by Rather and Subramanian (2018). Atikankul et al. (2020) are studded the Length-Biased Weighted Lindley Distribution. The length-biased power hazard rate distribution introduced by Musafa and Khan (2022).

A random sample X is said to have a new Sushila distribution (NSD) if its *pdf* is in the form:

$$g(x) = \frac{\beta\left(\alpha+\beta^2 x\right)e^{-\frac{\beta}{\alpha}x}}{\alpha^2(\beta+1)}; \qquad \alpha, \beta > 0 \text{ and } x \geq 0 \tag{2}$$

The $E(X)$ corresponding (2) is

$$E(X) = \frac{\alpha(2\beta+1)}{\beta(\beta+1)} \tag{3}$$

More details on this distribution and its applications can be found in Boonthiem et al. (2022).

2. The Length-Biased new Sushila Distribution

Using (1), (2) and (3) with $\theta = \frac{\beta}{\alpha}$, we can define the *pdf* of length-biased new Sushila distribution (*LBNSD*) as follows

$$f(x;\alpha,\theta) = \frac{\theta^2 x\left(1+\alpha\theta^2 x\right)e^{-\theta x}}{(2\alpha\theta+1)}; \qquad \alpha, \theta > 0 \text{ and } x \geq 0. \tag{4}$$

Probability plots of *LBNSD* are given in Figure 1 for particular values of α and θ.

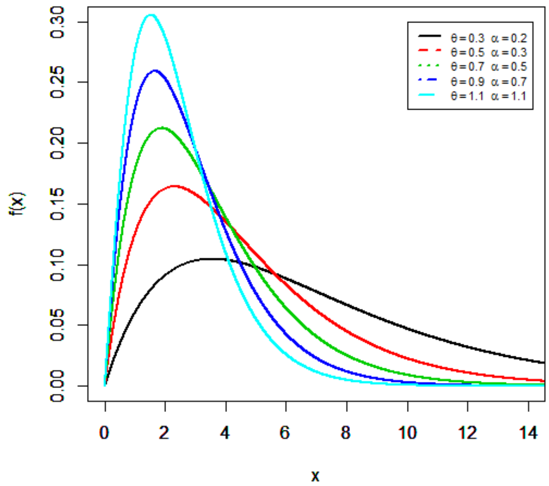

Figure 1: Plots of the *pdf* of *LBNSD* for various parameter values.

The cumulative distribution function (*CDF*) $F(x)$, survival (reliability) function $S(x)$, the hazard rate function (*HRF*) $h(x)$ and the reversed hazard rate function (*RHRF*) $h^*(x)$ for *LBNSD* are in the following forms:

$$F(x) = 1 - \left[\frac{\alpha\theta^3 x^2}{2\alpha\theta + 1} + \theta x + 1\right] e^{-\theta x}, \tag{5}$$

$$S(x) = \left[\frac{\alpha\theta^3 x^2}{2\alpha\theta + 1} + \theta x + 1\right] e^{-\theta x}, \tag{6}$$

$$h(x) = \frac{\theta^2 x\left(1 + \alpha\theta^2 x\right)}{\alpha\theta^3 x^2 + \theta(2\alpha\theta + 1)x + (2\alpha\theta + 1)}, \tag{7}$$

and

$$h^*(x) = \frac{\theta^2 x\left(1 + \alpha\theta^2 x\right) e^{-\theta x}}{(2\alpha\theta + 1) - \left[\alpha\theta^3 x^2 + \theta(2\alpha\theta + 1)x + (2\alpha\theta + 1)\right] e^{-\theta x}}. \tag{8}$$

Hazard function plots of *LBNSD* are given in Figure 2 for particular values of α and θ.

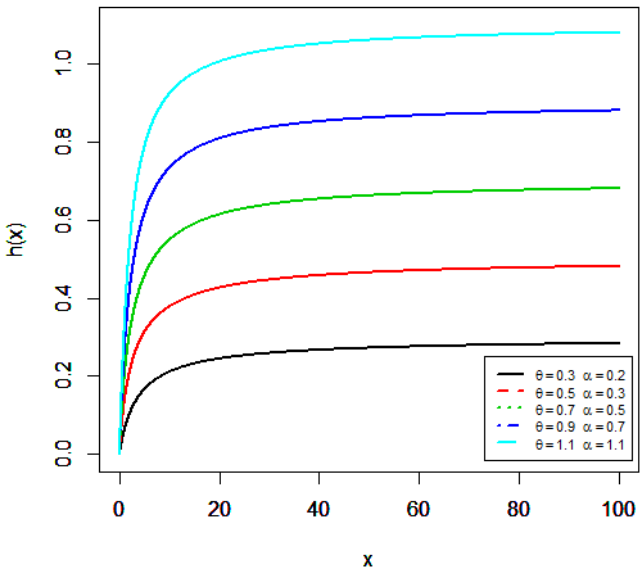

Figure 2: Plots of the *HRF* of *LBNSD* for various parameter values.

From Figure 2, we note that

- $\lim\limits_{x \to 0} h(x) = 0,$
- $\lim\limits_{x \to \infty} h(x) = \theta.$

Then the *HRF* is increasing failure rate and tends to constant as x tends to infinity.

3. Statistical properties

3.1 The mode

The mode of LBNSD is

$$x^* = \frac{1}{2\alpha\theta^2}\left[2\alpha\theta + \sqrt{1+4\alpha^2\theta^2} - 1\right]$$

Proof:

$$\ln f(x) = 2\ln(\theta) - \ln(2\alpha\theta+1) + \ln\left(x + \alpha\theta^2 x^2\right) - \theta x$$

$$\frac{\partial \ln f(x)}{\partial x} = \frac{1+2\alpha\theta^2 x}{x+\alpha\theta^2 x^2} - \theta$$

$$\frac{\partial \ln f(x)}{\partial x} = 0$$

$$\Rightarrow 1 + 2\alpha\theta^2 x = \theta x + \alpha\theta^3 x^2$$

$$\Rightarrow \alpha\theta^3 x^2 + \theta(1 - 2\alpha\theta)x - 1 = 0$$

$$\Rightarrow x = \frac{1}{\theta} + \frac{\sqrt{1 + 4\alpha^2\theta^2} - 1}{2\alpha\theta^2}$$

Note that: the *LBNSD* has mode for all values of θ but the base distribution don't has mode for $\theta \le 1$.

3.2.1 Moments about zero

The r^{th} moments of *LBGD* is

$$\mu_r' = E\left[X^r\right] = \frac{\theta^2}{2\alpha\theta + 1}\int_0^\infty x^{r+1}\left(1 + \alpha\theta^2 x\right)e^{-\theta x}\,dx$$

$$= \frac{\theta^{-r}}{(2\alpha\theta + 1)}\left[1 + \alpha\theta(r+2)\right]\Gamma(r+2).$$

The first four moments are

$$\mu_1' = \frac{2(1+3\alpha\theta)}{\theta(2\alpha\theta+1)},\ \mu_2' = \frac{6(1+4\alpha\theta)}{\theta^2(2\alpha\theta+1)},\ \mu_3' = \frac{24(1+5\alpha\theta)}{\theta^3(2\alpha\theta+1)} \text{ and } \mu_4' = \frac{120(1+6\alpha\theta)}{\theta^4(2\alpha\theta+1)}.$$

3.2.2 Moments about mean

The r^{th} moments about mean of *LBGD* is

$$\mu_r = E\left(X - \mu_1'\right)^r = \int_0^\infty \left(x - \mu_1'\right)^r f(x)\,dx$$

$$= \sum_{i=0}^{r}\binom{r}{i}(-1)^i\left(\mu_1'\right)^i \int_0^\infty x^{r-1} f(x)\,dx$$

$$= \sum_{i=0}^{r}\binom{r}{i}(-1)^i\left(\mu_1'\right)^i \mu_{r-i}'$$

The first four moments about mean are

$$\mu_1 = \mu_1' - \mu_1' = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + \left(\mu_1'\right)^3$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3\left(\mu_1'\right)^4$$

3.3 Quantiles

The Quantiles can be obtained from solving Eq. (9) numerically.

$$\ln\left[\frac{\alpha\theta^3 x^2}{2\alpha\theta+1}+\theta x+1\right]-\theta x=\ln(1-q) \tag{9}$$

Table (1) contain some results for first quintal, median and third quintal with different values of parameters.

Table (1): Some results for q_1, median and q_3 with different values of α and θ.

Parameters \ Quintal	q_1	Median	q_3
$\theta=0.3,\ \alpha=0.2$	3.386	5.905	9.459
$\theta=0.5,\ \alpha=0.3$	2.173	3.772	6.010
$\theta=0.7,\ \alpha=0.5$	1.723	2.950	4.635
$\theta=0.9,\ \alpha=0.7$	1.465	2.462	3.809
$\theta=1.1,\ \alpha=0.9$	1.281	2.116	3.233
$\theta=1.3,\ \alpha=1.1$	1.136	1.852	2.803
$\theta=1.5,\ \alpha=1.3$	1.019	1.644	2.471

From Table (1), we not that the values of q_1, median and q_3 are decreasing as α and θ increasing.

4. Maximum Likelihood Estimators

In this section, we consider maximum likelihood estimators (*MLE*) of *LBNSD*. Let $x_1, x_2, ..., x_n$ be a random sample of size n from *LBNSD*, then the log-likelihood function $L(\alpha,\theta)$ can be written as

$$L(\alpha,\theta)\propto n\left[2\ln(\theta)-\ln(2\alpha\theta+1)\right]+\sum_{i=1}^{n}\ln\left(x_i+\alpha\theta^2(x_i)^2\right)-\theta\sum_{i=1}^{n}(x_i) \tag{10}$$

The normal equations become

$$\frac{\partial L}{\partial\alpha}=\frac{-2n\theta}{2\alpha\theta+1}+\sum_{i=1}^{n}\frac{\theta^2(x_i)^2}{x_i+\alpha\theta^2(x_i)^2} \tag{11}$$

$$\frac{\partial L}{\partial\theta}=\frac{2n}{\theta}-\frac{2n\alpha}{2\alpha\theta+1}+2\alpha\theta\sum_{i=1}^{n}\frac{(x_i)^2}{x_i+\alpha\theta^2(x_i)^2}-\sum_{i=1}^{n}(x_i) \tag{12}$$

The MLE of α and θ can be obtain by solving the equations (11) and (12) using $\frac{\partial L}{\partial \alpha}=0$, and $\frac{\partial L}{\partial \theta}=0$.

5. Application

In this section, we compare the *LBNSD* with some other known competitive models to demonstrate its importance in data modeling. *MLE* method is used to estimate the parameters of the competitive models. the K-S (Kolmogorov-Smirnov) statistic, -2LL, AIC (Akaike Information Criterion), AICC (Akaike Information Criterion Corrected), BIC (Bayesian Information Criterion). The best distribution corresponds to lower -2LL, AIC, BIC, AICC statistics value. and P-value model selection goodness-of fit tests are used to choose best model.

Where,

AIC=2m -2LL, AICC=AIC+ $\frac{2m(m+1)}{n-m-1}$, BIC=mln(n)-2LL

m is the number of parameters in the statistical model, n the sample size.

Data I. First dataset represents the survival times of 72 guinea pigs (in days) infected with virulent tubercle bacilli, and these data were introduced by Bjerkedal (1960).
Data II. Second dataset is obtained from Hinkley (1977) and represents thirty successive values of March precipitation (in inches) in Minneapolies/St Paul.

Table 2: Descriptive statistics for both datasets.

	N	Mean	Median	Var	SK	KU
Data I	72	1.768	1.495	1.055	1.371	2.225
Data II	30	1.675	1.47	1.001	1.145	1.665

Table 3. Fitted estimates for data set I

Model	Parameter estimates	-2LL	K-S	AIC	BIC	AICC	P-value
LBNSD	$\hat{\theta}=1.697,\ \hat{\alpha}=4602$	188.488	0.109	192.488	197.041	192.662	0.3551
New Sushila dist.	$\hat{\theta}=14220,\ \hat{\alpha}=12570$	195.050	0.182	199.050	203.603	199.224	0.0170
Freshet dist.	$\hat{\theta}=1.069,\ \hat{\alpha}=1.173$	238.077	0.210	242.077	246.630	242.251	0.0034
Lomax dist.	$\hat{\theta}=520.83474,\ \hat{\alpha}=0.00109$	226.166	0.308	230.166	234.719	230.340	0.0000
Power function dist.	$\hat{\theta}=0.761,\ \hat{\alpha}=5.550$	240.871	0.295	244.871	249.424	245.045	0.0000

Table 4. Fitted estimates for data set II

Model	Parameter estimates	-2LL	K-S	AIC	BIC	AICC	P-value
LBNSD	$\hat{\theta}=1.791,\ \hat{\alpha}=459.127$	76.201	0.087	80.201	83.003	80.645	0.9760
New SD	$\hat{\theta}=2.989\times10^{4},\ \hat{\alpha}=2.504\times10^{4}$	78.478	0.145	82.478	85.280	82.922	0.5572
Freshet dist.	$\hat{\theta}=1.025,\ \hat{\alpha}=1.550$	94.847	0.186	98.847	101.649	99.291	0.2522
Lomax dist.	$\hat{\theta}=534.06124,\ \hat{\alpha}=0.00112$	90.986	0.269	94.986	97.788	95.430	0.0264
Power function dist.	$\hat{\theta}=0.819,\ \hat{\alpha}=4.750$	92.211	0.234	96.211	99.013	96.655	0.0741

Figures 3 and 4 shows the empirical, estimated *pdf* and *CDF* of the fitted *LBNSD* for the two real data sets.

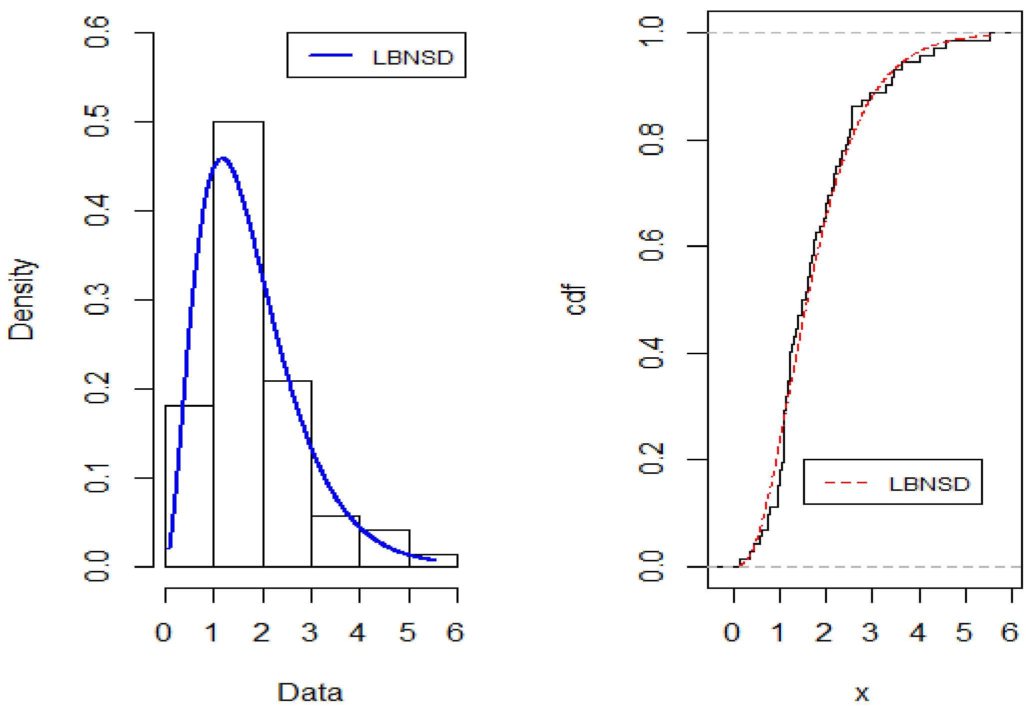

Figure 3: Empirical and estimated *pdf* and *CDF* of the fitted *LBNSD*

distribution for first data

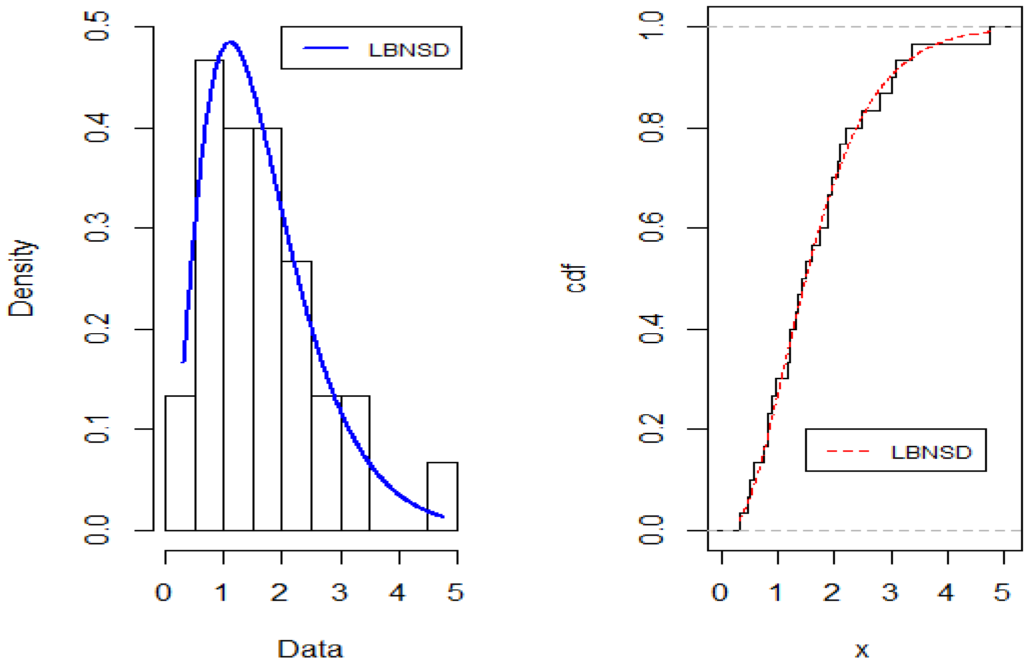

Figure 4: Empirical and estimated *pdf* and *CDF* of the fitted *LBNSD* distribution for second data

From Tables 3 and 4 and Figures 3 and 4, we can say that the real data sets indicates that the new model is superior to the fits than the other existing distributions.

References

[1] Abdul-Moniem, I. B., and Diab, L. S., (2018) "The Length-Biased weighted Exponentiated Lomax Distribution". *Int. J. Res. Math. & Stat.*, 4(1), 1-14.

[2] Afaq, A., Ahmad S.P., and Ahmed A., (2016). Length-biased weighted Lomax distribution: statistical properties and application, Pak. J. Stat. Oper. Res., XII (2), 245-255.

[3] Atikankul, Y., Thongteeraparp, A., Bodhisuwan, W. and Volodin, A. (2020). The Length-Biased Weighted Lindley Distribution with Applications. Lobachevskii Journal of Mathematics, 41(3), 308–319.

[4] Bjerkedal, T. (1960). Acquisition of Resistance in Guinea Pies infected with Different Doses of Virulent Tubercle Bacilli. American Journal of Hygiene, 72(1), 130-48.

[5] Boonthiem, S. Moumeesri, A. Klongdee, W. and Ieosanurak, W. (2022). A New Sushila Distribution: Properties and Applications. Eur. J. Pure Appl. Math, 15(3), 1280-1300

[6] Das, K. K. and Roy, T. D. (2011). Applicability of Length Biased Weighted Generalized Rayleigh distribution, Advances in Applied Science Research, 2(4), 320-327.

[7] Fisher, R.A (1934). The effects of methods of ascertainment upon the estimation of frequencies. Ann. Eugenics, 6, 13-25.

[8] Hinkley (1977) On quick choice of power transformation. Applied Statistics, 26, 67-69.

[9] Musafa, A., Khan, M. I. (2022). The length-biased power hazard rate distribution: Some properties and applications. STATISTICS IN TRANSITION new series, 23(2), 1–16.

[10] Rao, C.R. (1965). On discrete distributions arising out of method of ascertainment, in classical and Contagious Discrete, G.P. Patil. ed; Pergamum Press and Statistical publishing Society, Calcutta, 320-332.

[11] Rather, A. A. and Subramanian, C (2018). Length Biased Sushila distribution, Universal Review, 7(xii), 1010-1023.

[12] Seenoi, P., Supapakorn, T. and Bodhisuwan, W. (2014). The length-biased exponentiated inverted Weibull distribution. *Int. J. of Pure and Appl. Math.*, 92(2), 191-206.

[13] Shaban, S. A. and Boudrissa, N. A. (2007). The Weibull length-biased distribution: properties and estimation. InterStat (available for download at: http://interstat.statjournals.net/YEAR/2007/articles/0701002.pdf).