On a Size-Biased Quasi Poisson-Lindley Distribution

Rama Shanker[1,*], A. Mishra[2]

[1]Department of Statistics, Eritrea Institute of Technology, Asmara, Eritrea
[2]Department of Statistics, Patna University, Patna, India

Abstract   In this paper, a size- biased quasi Poisson-Lindley distribution, of which the size-biased Poisson-Lindley distribution of Ghitany and Al Mutairi[4] is a particular case, has been obtained by size biasing the quasi Poisson - Lindley distribution of Shanker and Mishra[14]. A general expression for the rth factorial moment of this distribution has been derived and hence its first four moments have been obtained. The method of maximum likelihood and the method of moments of estimation of its parameters have been discussed. The distribution has been fitted to some data sets to test its goodness of fit.

Keywords   Poisson-Lindley Distribution, Size-biased Distribution, Estimation of Parameters, Goodness of Fit

1. Introduction

Size-biased distributions arise in practice when observations from a sample are recorded with unequal probabilities, having probability proportional to some measure of unit size. Fisher[3] first introduced these distributions to model ascertainment bias which were later formalized by Rao[11] in a unifying theory. Van Deusen[16] discussed size-biased distribution theory and applied it to fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling (HPS). Later, Lappi and Bailey[7] used size-biased distributions to analyze HPS diameter increment data. Most of the statistical applications of these distributions, especially to the analysis of data relating to human population and ecology can be found in Patil and Rao ([9],[10]). Gove[5] reviewed some of the recent results on size-biased distributions pertaining to parameter estimation in forestry, with special emphasis on Weibull family.

If a random variable X have distribution $f(x; \theta)$ then a simple size-biased distribution is given by its probability function $f*(x; \theta) = x f(x; \theta) / \mu_1'$ where $\mu_1' = E(X)$, is the mean of the distribution.

Ghitany and Al Mutairi[4] obtained a size-biased Poisson - Lindley distribution (SBPLD) given by its probability mass function (p.m.f.)

$$P_1(x, \theta) = \frac{\theta^3}{\theta+2} \cdot \frac{x(x+\theta+2)}{(\theta+1)^{x+2}}; \theta > 0, x = 1, 2, 3, \dots (1.1)$$

by size biasing the Poisson -Lindley distribution (PLD) of Sankaran[12] having pmf

$$P_2(x; \theta) = \frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}} , \ x = 0, 1, 2, \dots, \ \theta > 0. (1.2)$$

It is to be mentioned that Sankaran[12] obtained the distribution (1.2) by mixing the Poisson distribution with the Lindley (1958) distribution having pdf

$$f_1(x, \theta) = \frac{\theta^2}{\theta+1}(1+x) e^{-\theta x}; x > 0, \ \theta > 0 \quad (1.3)$$

The first four moments about origin of the SBPLD (1.1) have been obtained as

$$\mu_1' = 1 + \frac{2(\theta+3)}{\theta(\theta+2)} \quad (1.4)$$

$$\mu_2' = 1 + \frac{6(\theta+3)}{\theta(\theta+2)} + \frac{6(\theta+4)}{\theta^2(\theta+2)} \quad (1.5)$$

$$\mu_3' = 1 + \frac{14(\theta+3)}{\theta(\theta+2)} + \frac{36(\theta+4)}{\theta^2(\theta+2)} + \frac{24(\theta+5)}{\theta^3(\theta+2)} \quad (1.6)$$

$$\mu_4' = 1 + \frac{30(\theta+3)}{\theta(\theta+2)} + \frac{126(\theta+4)}{\theta^2(\theta+2)} + \frac{240(\theta+5)}{\theta^3(\theta+2)} + \frac{120(\theta+6)}{\theta^4(\theta+2)} \quad (1.7)$$

and so its variance as

$$\mu_2 = \frac{2(\theta^3 + 6\theta^2 + 12\theta + 6)}{\theta^2(\theta+2)^2} \quad (1.8)$$

Recently, Shanker and Mishra[14] obtained a quasi Poisson-Lindley distribution (QPLD) given by its pmf

$$P_3\left(x;\theta,\alpha\right) = \frac{\theta}{\left(\alpha+1\right)\left(\theta+1\right)^{x+2}}\left(\alpha + \frac{\theta\left(x+1\right)}{\theta+1}\right); \quad (1.9)$$

$$x = 0,1,2,...; \ \theta > 0, \alpha > -1$$

It can be seen that the PLD (1.2) is a particular case of it at $\alpha = \theta$. Shanker and Mishra[14] have shown that (1.9) is a better model than the PLD of Sankaran[12] for analyzing different types of count data. This distribution arises from the Poisson distribution when its parameter λ follows the Shanker and Mishra[13] quasi Lindley distribution having its probability density function (p.d.f)

$$f_2\left(x;\theta,\alpha\right) = \frac{\theta}{\alpha+1}\left(\alpha+x\theta\right)e^{-\theta x}; x>0, \theta>0, \alpha>-1 \quad (1.10)$$

In this paper, a size-biased quasi Poisson-Lindley distribution (SBQPLD) of which the SBPLD (1.1) is a particular case, has been obtained. A general expression for the rth factorial moment of this distribution has been obtained and hence its first four moments about origin of this distribution have been obtained. The method of maximum likelihood and the method of moments of estimation of its parameters have been discussed. The distribution has been fitted to some data- sets and it has been found that to almost all these data-sets it provides closer fits than the SBPLD. This shows that the SBQPLD is more flexible than the

Ghitany et al[4] SBPLD for analyzing different types of count data.

2. A Size-Biased Quasi Poisson-Lindley Distribution (Sbqpld)

The pmf of size-biased quasi Poisson-Lindley distribution (SBQPLD) with parameters α and θ is obtained as

$$P_4\left(x;\theta,\alpha\right) = \frac{x\,P_3\left(x;\theta,\alpha\right)}{\mu_1'} \quad (2.1)$$

Taking the pmf $P_3\left(x;\theta,\alpha\right)$ from (1.9) and its mean $\mu_1' = \left(\alpha+2\right)\big/\theta\left(\alpha+1\right)$ of Shanker and Mishra[14], we get

$$P_4\left(x;\theta,\alpha\right) = \frac{\theta^2}{\alpha+2}\frac{x\left(\theta x+\theta\alpha+\theta+\alpha\right)}{\left(\theta+1\right)^{x+2}};$$

$$x=1,2,3,.....; \ \theta>0, \ \alpha>-2 \quad (2.2)$$

It can be easily seen that (2.2) reduces to (1.1) at $\alpha=\theta$.

The SBQPLD can also be obtained from the size-biased Poisson distribution when its parameter λ follows a size-biased quasi Lindley distribution with p.d.f.

$$f_3\left(\lambda;\theta,\alpha\right) = \frac{\theta^2}{\alpha+2}\lambda\left(\alpha+\lambda\theta\right)e^{-\theta\lambda} \quad ;\lambda>0, \theta>0, \alpha>-2 \quad (2.3)$$

We have

$$P\left(X=x\right) = \int_0^\infty \frac{e^{-\lambda}\lambda^{x-1}}{\left(x-1\right)!}\cdot\frac{\theta^2}{\alpha+2}\lambda\left(\alpha+\lambda\theta\right)e^{-\theta\lambda}d\lambda \quad (2.4)$$

$$= \frac{\theta^2}{\alpha+2}\frac{1}{\left(x-1\right)!}\int_0^\infty e^{-\left(\theta+1\right)\lambda}\left(\alpha\lambda^x+\theta\lambda^{x+1}\right)d\lambda$$

$$= \frac{\theta^2}{\alpha+2}\left[\frac{\alpha x}{\left(\theta+1\right)^{x+1}}+\frac{\theta\left(x+1\right)x}{\left(\theta+1\right)^{x+2}}\right]$$

$$= \frac{\theta^2}{\alpha+2}\frac{x\left(\theta x+\theta\alpha+\theta+\alpha\right)}{\left(\theta+1\right)^{x+2}} \quad ;x=1,2,3,... \quad (2.5)$$

which is the SBQPLD. To know the behaviors of the pmf of the SBQPLD, plots of its pmf have been drawn for different values of its parameters in figure 1.

As

$$\frac{P_4\left(x+1;\theta,\alpha\right)}{P_4\left(x;\theta,\alpha\right)} = \frac{1}{\theta+1}\left(1+\frac{1}{x}\right)\left(1+\frac{\theta}{\theta x+\theta\alpha+\theta+\alpha}\right) \quad (2.6)$$

is a decreasing function of x, $P_4\left(x;\theta,\alpha\right)$ is log-concave. Therefore, the SBQPLD is unimodal, has an increasing failure rate (IFR) and hence increasing failure rate average (IFRA). It is new better than used (NBU), new better than used in expectation (NBUE) and has decreasing mean residual life (DMRL). Details about the definitions of these aging concepts can be seen in Barlow and Proschan[1].
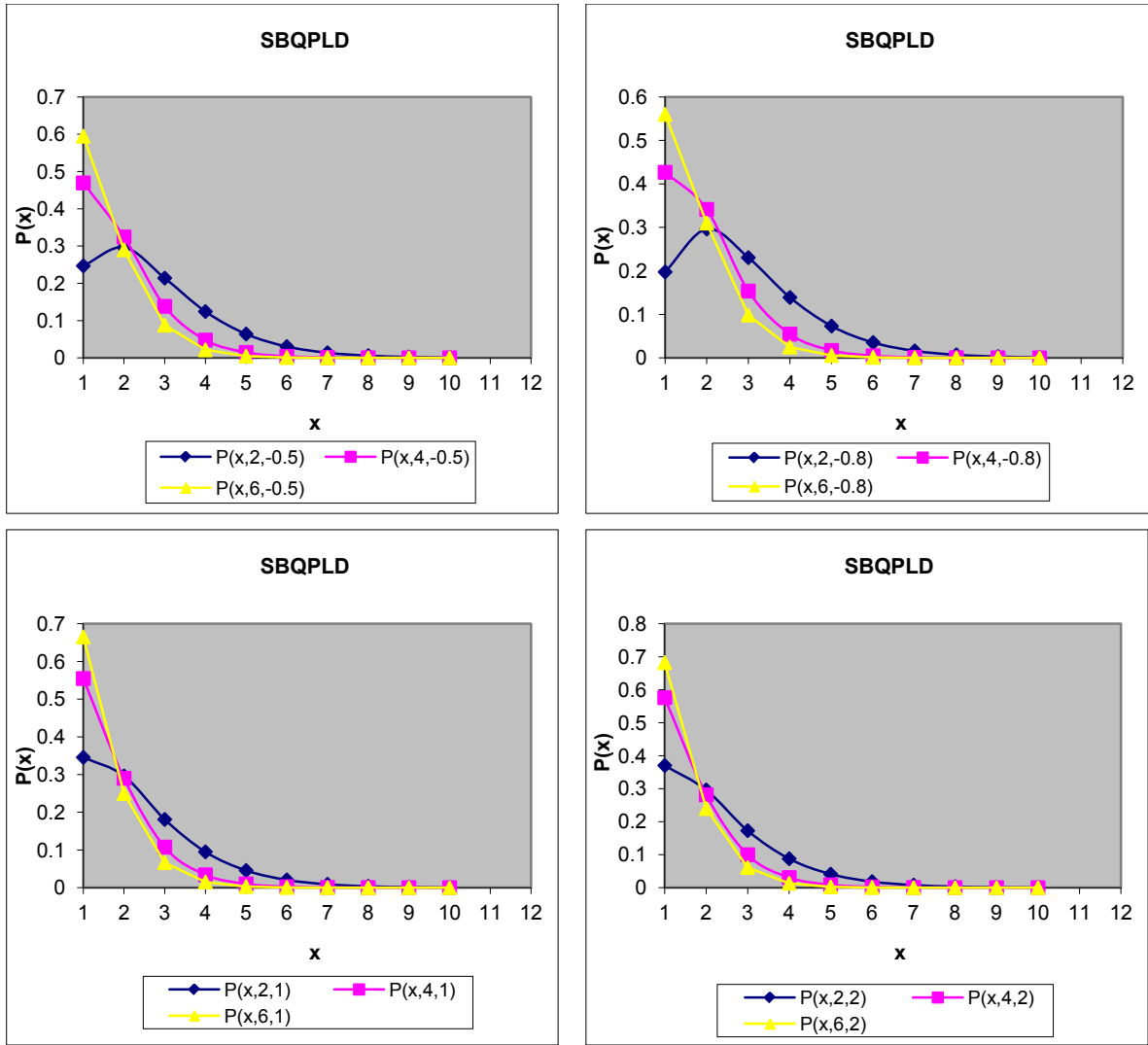
Figure 1.   Plots of SBQPLD for different values of parameters θ and α

3. Moments

The rth factorial moment of the SBQPLD (2.2) can be obtained as

$$\mu'_{(r)} = E\left[E\left(X^{(r)} | \lambda \right) \right] \tag{3.1}$$

where $X^{(r)} = X(X-1)(X-2)......(X-r+1)$.

From (2.4), we get

$$\mu'_{(r)} = \int_0^\infty \left[\sum_{x=0}^\infty x^{(r)} \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} \right] \frac{\theta^2}{\alpha+2} \lambda(\alpha + \lambda\theta) e^{-\theta\lambda} \, d\lambda \tag{3.2}$$

$$= \int_0^\infty \left[\lambda^{r-1} \sum_{x=r}^\infty x \; \frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} \right] \frac{\theta^2}{\alpha+2} \lambda(\alpha + \lambda\theta) e^{-\theta\lambda} \, d\lambda$$

Taking (x+r) in place of x, we get

$$\mu'_{(r)} = \int_0^\infty \lambda^{r-1} \left[\sum_{x=0}^\infty (x+r) \; \frac{e^{-\lambda} \lambda^{x}}{x!} \right] \frac{\theta^2}{\alpha+2} \lambda(\alpha + \lambda\theta) e^{-\theta\lambda} \, d\lambda$$

The expression within bracket is clearly $(\lambda + r)$ and hence we have

$$\mu'_{(r)} = \frac{\theta^2}{\alpha+2} \int_0^\infty (\lambda+r)\lambda^r (\alpha+\lambda\theta)e^{-\theta\lambda} \, d\lambda \tag{3.3}$$

Using gamma integral, we get finally, after a little simplification, a general expression for the rth factorial moment of the SBQPLD as

$$\mu'_{(r)} = \frac{\Gamma(r+1)}{(\alpha+2)\theta^r}\left[(\alpha+r+2)(r+1)+r\theta(\alpha+r+1)\right]; \; r = 1,2,3,... \tag{3.4}$$

After obtaining the first four factorial moments, by substituting $r = 1,2,3$ and 4 in (3.4) and then using the relationship between factorial moments and moments about origin, the first four moments about origin of the SBQPLD were obtained as

$$\mu'_1 = 1+\frac{2(\alpha+3)}{\theta(\alpha+2)} \tag{3.5}$$

$$\mu'_2 = 1+\frac{6(\alpha+3)}{\theta(\alpha+2)}+\frac{6(\alpha+4)}{\theta^2(\alpha+2)} \tag{3.6}$$

$$\mu'_3 = 1+\frac{14(\alpha+3)}{\theta(\alpha+2)}+\frac{36(\alpha+4)}{\theta^2(\alpha+2)}+\frac{24(\alpha+5)}{\theta^3(\alpha+2)} \tag{3.7}$$

$$\mu'_4 = 1+\frac{30(\alpha+3)}{\theta(\alpha+2)}+\frac{126(\alpha+4)}{\theta^2(\alpha+2)}+\frac{240(\alpha+5)}{\theta^3(\alpha+2)}+\frac{120(\alpha+6)}{\theta^4(\alpha+2)} \tag{3.8}$$

It can be easily verified that at $\alpha = \theta$, these moments reduce to the respective moments, (1.4) to (1.7) of the SBPLD.

4. Estimation of Parameters

4.1. Maximum Likelihood Estimates

Let $(x_1, x_2,\ldots, x_n)$ be a random sample of size n from the SBQPLD (2.2) and let f_x be the observed frequency in the sample corresponding to $X = x$ $(x = 1,2,...,k)$ such that $\sum_{x=1}^k f_x = n$, where k is the largest observed value having non-zero frequency. The likelihood function, L of the SBQPLD is given by

$$L = \left(\frac{\theta^2}{\alpha+2}\right)^n \frac{1}{(\theta+1)^{\sum_{x=1}^k (x+2)f_x}} \prod_{x=1}^k \left[\theta x^2 + x(\theta\alpha+\theta+\alpha)\right]^{f_x} \tag{4.1}$$

and so the log likelihood function is obtained as

$$\log L = n\log\left(\frac{\theta^2}{\alpha+2}\right) - \sum_{x=1}^k f_x(x+2)\log(\theta+1) + \sum_{x=1}^k f_x\log\left[\theta x^2 + x(\theta\alpha+\theta+\alpha)\right] \tag{4.2}$$

The two log likelihood equations are thus obtained as

$$\frac{\partial \log L}{\partial \theta} = \frac{2n}{\theta} - \sum_{x=1}^k \frac{(x+2)f_x}{\theta+1} + \sum_{x=1}^k \frac{f_x\left[x^2 + x(\alpha+1)\right]}{\left[\theta x^2 + x(\theta\alpha+\theta+\alpha)\right]} = 0 \tag{4.3}$$

$$\frac{\partial \log L}{\partial \alpha} = \frac{-n}{\alpha+2} + \sum_{x=1}^k \frac{x(\theta+1)f_x}{\left[\theta x^2 + x(\theta\alpha+\theta+\alpha)\right]} = 0 \tag{4.4}$$

The two equations (4.3) and (4.4) are difficult to solve directly. However, the Fisher's scoring method can be applied to solve these equations. For, we have

$$\frac{\partial^2 \log L}{\partial \theta^2} = \frac{-2n}{\theta^2} + \sum_{x=1}^{k} \frac{(x+2) f_x}{(\theta+1)^2} - \sum_{x=1}^{k} \frac{f_x \left[x^2 + x(\alpha+1) \right]^2}{\left[\theta x^2 + x(\theta\alpha + \theta + \alpha) \right]^2} \tag{4.5}$$

$$\frac{\partial^2 \log L}{\partial \theta \, \partial \alpha} = -\sum_{x=1}^{k} \frac{x(x^2 + x) f_x}{\left[\theta x^2 + x(\theta\alpha + \theta + \alpha) \right]^2} \tag{4.6}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \frac{n}{(\alpha+2)^2} - \sum_{x=1}^{k} \frac{x^2 (\theta+1)^2 f_x}{\left[\theta x^2 + x(\theta\alpha + \theta + \alpha) \right]^2} \tag{4.7}$$

using which the equations

$$\begin{bmatrix} \dfrac{\partial^2 \log L}{\partial \theta^2} & \dfrac{\partial^2 \log L}{\partial \theta \partial \alpha} \\[3mm] \dfrac{\partial^2 \log L}{\partial \theta \partial \alpha} & \dfrac{\partial^2 \log L}{\partial \alpha^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\alpha}=\alpha_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\[2mm] \hat{\alpha} - \alpha_0 \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \log L}{\partial \theta} \\[3mm] \dfrac{\partial \log L}{\partial \alpha} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\alpha}=\alpha_0}} \tag{4.8}$$

can be solved for the estimates, $\hat{\theta}$ and $\hat{\alpha}$. where θ_0 and α_0 are the initial values of θ and α respectively. These equations are solved numerically and iteratively till sufficiently close estimates $\hat{\theta}$ and $\hat{\alpha}$ are obtained.

4.2. Method of Moments

The SBQPLD has two parameters to be estimated and so its first two moments are required to get the estimates of its parameters by the method of moments.

From (3.5) and (3.6) we have

$$\frac{(\mu_2' - 1) - 3(\mu_1' - 1)}{(\mu_1' - 1)^2} = \frac{3(\alpha+4)(\alpha+2)}{2(\alpha+3)^2} = K \, (\text{say}) \tag{4.9}$$

which gives a quadratic equation in α as

$$g(\alpha) = (3 - 2K)\alpha^2 + (18 - 12K)\alpha + (24 - 18K) = 0. \tag{4.10}$$

Replacing the first two population moments by the respective sample moments in (4.9) an estimate k of K can be obtained

and using it in (4.10), an estimate $\hat{\alpha}$ of α can be obtained.

Again, substituting the value of $\hat{\alpha}$ in (3.3) and replacing the population mean by the sample mean $\bar{x}$, an estimate of θ is obtained as

$$\hat{\theta} = \frac{2(\alpha+3)}{(\alpha+2)(\bar{x}-1)} \tag{4.11}$$

5. Goodness of Fit

The SBQPLD has been fitted to a number of data-sets related to a number of observations of the size distribution of 'freely –forming' small groups in various public situations reported by James[6], Coleman and James[2], and Simonoff[15] and it was found that to all these data -sets, the SBQPLD provides closer fits than the SBPLD. The fittings of the SBQPLD to only three such data-sets have been presented in the following tables.

The expected frequencies according to the SBPLD have also been given in these tables for ready comparison with those obtained by the SBQPLD. The estimates of the parameters have been obtained by the method of maximum likelihood using the Fisher's scoring method.

It can be seen that the SBQPLD gives much closer fits than the SBPLD and thus provides a better alternative to the SBPLD.

Table 1. Counts of groups of people in public places on a spring afternoon in Portland

Size of Groups	Observed frequency	Expected frequency	
		SBPLD	SBQPLD
1	1486	1532.5	1485.4
2	694	630.6	697.3
3	195	191.9	189.7
4	37	51.3	40.6
5	10	12.8 }	8.3 }
6	1	3.9 }	1.7 }
Total	2423	2423.0	2423.0

$$\hat{\theta} = 4.5082 \qquad \hat{\theta} = 7.14227,\ \hat{\alpha} = -0.79159$$

χ^2		13.766	0.710
d.f		3	2
$P(\chi^2)$		< 0.01	0.70

Table 2. Counts of Shopping Groups-Eugene, Spring, Department Store and Public Market

Size of Groups	Observed frequency	Expected frequency	
		SBPLD	SBQPLD
1	316	323.0	315.7
2	141	132.5	142.9
3	44	40.2	40.5
4	5	10.7 }	8.9 }
5	4	3.6 }	2.0 }
Total	60	510.0	510.0

$$\hat{\theta} = 4.5224 \qquad \hat{\theta} = 6.57872,\ \hat{\alpha} = -0.52229$$

χ^2		3.021	0.659
d.f		2	1
$P(\chi^2)$		0.40	0.44

Table 3. Counts of Play Groups-Eugene, Spring, Public Playground D

Size of Groups	Observed frequency	Expected frequency	
		SBPLD	SBQPLD
1	305	314.4	304.2
2	144	134.4	148.5
3	50	42.5	42.8
4	5	11.8 }	9.2 }
5	2	3.1 }	1.9 }
6	1	0.8 }	0.4 }
Total	507	507.0	507.0

$$\hat{\theta} = 4.3179 \qquad \hat{\theta} = 6.74857,\ \hat{\alpha} = -0.76547$$

χ^2		6.415	2.415
d.f		2	1
$P(\chi^2)$		0.043	0.130

6. Conclusions

In this paper, a size-biased quasi Poisson-Lindley distribution (SBQPLD), of which the size-biased Poisson-Lindley distribution (SBPLD) is a particular case, has been introduced to model count data. A general expression for the rth factorial moment of this distribution has been obtained and hence its first four moments about origin of this distribution have been obtained. The method of maximum likelihood and the method of moments of estimation of its parameters have been discussed. The distribution has been fitted to some data- sets and it has been found that to all these data- sets it provides closer fits than the SBPLD. The SBQPLD has been found more general in nature and wider in scope than SBPLD. It provides closer fits to observed data-sets than those provided by the SBPLD and hence it should be preferred to the SBPLD while modeling count data-sets.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Barlow,R.E. and Proschan, F. (1981): *Statistical Theory of Reliability and Life   Testing*, Silver Spring, MD

[2]   Coleman, J.S. and James, J. (1961): The equilibrium size distribution of freely   forming groups, *Sociometry*, Vol. 24, pp. 36 – 45

[3]   Fisher, R.A. (1934) : The effects of methods of ascertainment upon the estimation of frequency, *Ann. Eugenics,* Vol. 6, pp. 13-25.

[4]   Ghitany, M.E., Al-Mutairi, D.K. (2008): Size-biased Poisson -Lindley distribution and its applications, *Metron*, Vol. LXVI, No. 3, pp. 299 – 311

[5]   Gove, H. J. (2003): Estimation and application of size-biased distributions in forestry, *In Modeling Forest systems,* A. Amaro, D. Reed and P. Soares, CAB International, Wallingford, U.K., 201-212.

[6]   James, J. (1953): The Distribution of Free-Forming Small Group Size, *American Sociological Review*, Vol. 18, pp. 569 – 570.

[7]   Lappi, J. and Bailey, R. L. (1987): Estimation of diameter increment function or other relations using angle-count samples, *Forest Science,* Vol. 33, 725-739.

[8]   Lindley, D.V.(1958) Fiducial distributions and Bayes' theorem*, Journal of the Royal Statistical Society, Series B*, Vol. 20, pp. 102- 107

[9]   Patil, G. P. and Rao, C. R. (1977): The weighted distributions: A survey and their Applications, *In Applications of Statistics,* (*Ed. P. R. Krishnaiah*), 383-405, North Holland Publications Co., Amsterdam.

[10] Patil, G. P. and Rao, C. R. (1978): Weighted distributions and size-biased sampling with applications to wild life populations and human families, *Biometrics*, Vol. 34 179-189.

[11] Rao, C. R. (1965): On discrete distributions arising out of methods on ascertainment, *Classical and Contagious Discrete Distribution,* Patil, G.P. (Ed), Statistical   Publishing Society, Calcutta, pp. 320-332

[12] Sankaran, M (1970): The discrete Poisson-Lindley distribution, *Biometrics*, Vol. 26, pp. 145 – 149.

[13] Shanker, R. and Mishra, A. (2013 a): A quasi Lindley distribution, *African Journal of Mathematics and Computer Science Research,* Vol 6 (4), 64 – 71.

[14] Shanker, R. and Mishra, A. (2013 b): A quasi Poisson-Lindl ey distribution (Communicated)

[15] Simonoff, J.S. (2003): *Analyzing Categorical Data*, New York Springer.

[16] Van Deusen, P.C. (1986): Fitting assumed distributions to horizontal point sample Diameters, *Forest Science,* Vol. 32, 146-148.