International Journal of Mathematical Archive-5(1), 2014, 89-96
Available online through www.ijma.info ISSN 2229 – 5046

SOME PROPERTIES AND APPLICATION OF SIZE-BIASED POISSON-LINDLEY DISTRIBUTION

Porinita Dutta*
Department of Mathematical Sciences, Tezpur University, Napaam. City: Tezpur. State: Assam. Pin: 784028

Munindra Borah
Department of Mathematical Sciences, Tezpur University, Napaam. City: Tezpur. State: Assam. Pin: 784028

(Received on: 09-01-14; Revised & Accepted on: 30-01-14)

ABSTRACT

Size-biased distributions are a special case of the more general form known as weighted distributions. Such distributions arise naturally in practice when observations from a sample are recorded with unequal probability. In this study, size-biased Poisson-Lindley distribution (SBPL) is further investigated by working out recursive relationship of the probabilities, moments and cumulants etc. Some distributional properties of size-biased Poisson-Lindley distribution including moments, cumulants, harmonic mean, coefficient of variation, reliability function etc has been derived. The estimate of the parameter is obtained by employing method of moments and ratio of the first two relative frequencies. To justify the suitability, the distribution is fitted to a number of reported data sets. The resulting fit is found to be good in comparison to others.

Keywords: *Poisson-Lindley distribution, weighted distribution, Size-biased distribution, Recursive relations, parameter estimation, Goodness of fit.*

1. INTRODUCTION

Poisson-Lindley distribution is a well-known one parameter mixture Poisson distribution which has wide applications in the theory of accident proneness with probability mass function (p.m.f)

$$f(x;\theta)=\frac{\theta^2(x+\theta+2)}{(1+\theta)^{x+3}},\quad x=0,\,1\,,\dots,\,\theta>0. \tag{1.1}$$

The distribution arises from the Poisson distribution when its parameter λ follows a continuous Lindley distribution due to Lindley (1958).Sankaran(1970) investigated this distribution with application to errors and accidents data. Some of the difficulties in obtaining maximum likelihood estimator (MLE) of the parameter θ of Poisson-Lindley distribution is pointed out by him. Borah and Deka Nath (2000) reviewed certain properties of Poisson-Lindley distribution and also investigated two generalized distributions of Poisson-Lindley by using Gurland's generalization theorem. Two forms of geometrically infinite divisible two-parameter Poisson-Lindley distribution studied byBegum (2002). Size-biased Poisson-Lindley distribution and some of its properties are investigated by Ghitany and Al-Mutairi (2008). Recently, Mohmoudi and Zakerzadeh (2010) obtained an extended version of Poisson- Lindley distribution and estimated its parameter using the moments and maximum likelihood method.

Size-biased distributions are a special case of weighted distributions. The concept of weighted distributions can be traced to the study of the effect of methods of ascertainment upon estimation of frequencies by fisher (1934),weighted distributions were later formalized in a unifying theory by Rao (1965). Such distributions arise naturally when observations generated from a stochastic process are not given equal chance of being recorded; instead they are recorded according to some weighted function. If the random variable X has distribution $f(x;\theta)$, with unknown parameterθ, then the corresponding weighted distribution is of the form

$$f^w(x;\theta)=\frac{w(x)f(x;\theta)}{E[w(x)]}$$

where$w(x)$ is a non-negative weight function such that $E[w(x)]$ exists.

*Corresponding author: Porinita Dutta**
E-mail: porikal@tezu.ernet.in

When the weightfunction of the weighted distribution depends on the lengths of the units of the interest; the resulting distribution is called length-biased. The weighted distribution with the weight function $w(x) = x$ is known as size-biased distribution.A study of size-biased sampling and related form- invariant weighted distributions was made by Patil and Ord (1975). Van Deusen (1986) arrived at size biased distribution theory independently and applied it to fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling (HPS) (Grosenbaugh) inventories. Gove (2003) reviewed some of the more recent results on size-biased distributions pertaining to parameter estimation in forestry. Mir (2009) also discussed some of the discrete size-biased distributions. Recently, Length-biased weighted Weibull distribution and applicability of length-biased weighted Weibull distribution are studied by Das and Roy (2011).

In this paper, Size-biased Poisson-Lindley distribution is further investigated by working out some recursive relationship of probabilities, factorial moments, raw moments, cumulants etc. Other distributional properties of the distribution including coefficient of variation, harmonic mean, reliability function etc. has been also derived. Estimation of the parameter has been also discussed. Finally, fitted the distribution to two reported data sets for empirical comparison.

2. MATERIALS AND METHODS

Two reported data sets such as immunogold assay data and animal abundance data from Ghitany and Al-Mutairi (2008) are considered for the study to empirical comparison of the distribution.

To introduce the concept of a weighted distribution, suppose X is a non-negative random variable with its natural Probability mass function (p.m.f)$f(x;\theta)$, where the natural parameter is θ. Let the weight function be $w(x)$ is a non-negative function. Then the weighted p.m.fis obtained as

$$f^{w}(x;\theta) = \frac{w(x)f(x;\theta)}{E[w(x)]}, \quad x = 1,2,3,... \tag{2.1}$$

Assuming that $E[(x)]$ i.e. the first moment of (x) exists.

By taking weight $w(x) = x$ we obtain size-biased distribution.For example, when $w(x) = x$, in that case $E(x) = \mu$. Then the distribution is called size-biased distribution with p.m.f

$$f^{*}(x;\theta) = \frac{xf(x;\theta)}{\mu}$$

3. SIZE-BIASED POISSON-LINDLEY DISTRIBUTION

A size-biased Poisson-Lindley distribution is obtained by applying weights $w(x) = x$ to the Poisson-Lindley distribution. The p.m.f of Size-biased Poisson-Lindley distribution can be obtained as

$$f^{*}(x;\theta) = \frac{xf(x;\theta)}{\mu} = \frac{\theta^3}{\theta+2}\frac{x(x+\theta+2)}{(1+\theta)^{x+2}}, \quad x = 1,\ 2,\ \dots, \quad \theta > 0 \tag{3.1}$$

where$\mu = \frac{\theta+2}{\theta(1+\theta)}$ is the mean of the Poisson-Lindley distribution with p.m.f (1.1).

Ghitany and Al-Mutairi (2008) investigated this distribution with application and also investigated some distributional properties of size-biased Poisson-Lindley distribution. They also showed that Size-biased Poisson-Lindley (SBPL) distribution also arises from the size-biased Poisson (SBP) distribution with p.m.f

$$g(x\backslash\lambda) = e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!}\ , x = 1,2,3,\dots., \qquad \lambda > 0.$$

when its parameter λ follows a size-biased Lindley (SBL) model with p.d.f.

$$h(\lambda\,;\,\theta) = \frac{\theta^3}{\theta+2}\lambda\,(1+\lambda)e^{-\theta\lambda}, \quad \lambda > 0, \quad \theta > 0.$$

They proved that, the method of moments and the maximum likelihood estimators of size-biased Poisson-Lindley distribution are positively biased (asymptotically unbiased) for small (large) sample size, consistent and asymptotically normal with almost equal asymptotic variances.

4. DISTRIBUTIONAL PROPERTIES OF SIZE-BIASED POISSON-LINDLEY DISTRIBUTION

The p.mf of SBPL distribution is

$$p_x(\theta) = \frac{\theta^3}{\theta+2}\frac{x(x+\theta+2)}{(1+\theta)^{x+2}}, \qquad x = 1,2,3,\ldots, \quad \theta > 0. \tag{4.1}$$

where θ is known as parameter of the distribution and denote the distribution by SBPL (θ). Note that,

$$\sum_{x=1}^{\infty} p_x(\theta) = \frac{\theta^3}{(\theta+2)(1+\theta)^2}\sum_{x=1}^{\infty}\frac{x(x+\theta+2)}{(1+\theta)^x} = 1.$$

Distribution function

The corresponding distribution function is

$$F(x) = P(X \le x) = \sum_{m=1}^{x} p(m) = \frac{\theta^3}{(\theta+2)(1+\theta)^2}\sum_{m=1}^{x}\frac{m(m+\theta+2)}{(1+\theta)^m} \quad , x = 1,2,3, \ldots.$$

Probability generating function (p. g. f)

The probability generating function (p.g.f) of SBPL distribution is

$$g(t) = \frac{\theta^3 t(3+\theta-t)}{(2+\theta)(1+\theta-t)^3}, \quad \theta > 0 \tag{4.2}$$

The expression for probability to obtain higher order probabilities of the distribution is given by

$$P_{r+1} = \frac{(4r+2\theta r+4)P_r - rP_{r-1}}{r(3+\theta)(1+\theta)}, \quad r \ge 1 \tag{4.3}$$

where,

$$P_1 = \frac{\theta^3(\theta+3)}{(\theta+2)(\theta+1)^3}$$

$$P_2 = \frac{\theta^3 2(\theta+4)}{(\theta+2)(1+\theta)^4}$$

Factorial moment generating function (f. m. g. f)

The factorial moment generating function (f. m. g. f) is

$$g(\mathrm{t}+1) = \frac{\theta^3}{(\theta+2)}\frac{(1+t)(\theta+2-t)}{(\theta-t)^3}, \quad \theta > 0 \tag{4.4}$$

The expression for factorial moments of SBPL distribution is given by

$$\mu_{(r)} = \frac{1}{\theta^3} r\left[3\theta^2\mu_{(r-1)} - (r-1)\{3\theta\mu_{(r-2)} + (r-2)\mu_{(r-3)}\}\right], \quad r \ge 3. \tag{4.5}$$

where,

$$\mu_{(1)} = \frac{\theta^2+4\theta+6}{\theta(\theta+2)}$$

$$\mu_{(2)} = \frac{2(2\theta^2+9\theta+12)}{\theta^2(\theta+2)}$$

Moment generating function (m. g. f)

The moment generating function (m.g.f) is

$$m(t) = \frac{\theta^3 e^t(3+\theta-e^t)}{(2+\theta)(1+\theta-e^t)^3}, \quad \theta > 0 \tag{4.6}$$

The expression for recursive relation of raw moments of the distribution is

$$\mu'_{r+1} = \frac{1}{A}\left[B + (1+\theta)(3+\theta)\mu'_r + \sum_{j=1}^{r}\left\{\binom{r}{j}(4+2\theta-2^j) + \binom{r}{j-1}(4-2^{j-1})\right\}\mu'_{r+1-j}\right], \quad r > 1 \tag{4.7}$$

where $A = \theta(\theta+2),\ B = (4-2^r)$

$\mu'_1 = \frac{\theta^2+4\theta+6}{\theta(\theta+2)}$ (Mean)

$\mu'_2 = \frac{\theta^4+10\theta^3+40\theta^2+72\theta+48}{\theta^2(1+\theta)^2}$

$\mu'_3 = \frac{\theta^6+20\theta^5+146\theta^4+564\theta^3+1104\theta^2+1052\theta+480}{\theta^3(1+\theta)^3}$ etc.

Mean and variance of SBPL distribution

The mean of SBPL distribution is given by

$$\mu = \frac{\theta^2+4\theta+6}{\theta(\theta+2)} \tag{4.8}$$

The variance of SBPL distribution is

$$\sigma^2 = \frac{2(\theta^3+6\theta^2+12\theta+6)}{\theta^2(\theta+2)^2} \tag{4.9}$$

Third and fourth order central moments of SBPL distribution

$\mu_3 = \frac{2(\theta^5+10\theta^4+42\,\theta^3+72\theta+24)}{\theta^3(1+\theta)^3}$

$\mu_4 = \frac{2(\theta^7+22\theta^6+184\theta^5+780\theta^4+1800\theta^3+225\theta^2+1440\theta+360)}{\theta^4(1+\theta)^4}$

Coefficient of variation of SBPL distribution

$$CV = \frac{\sqrt{V(X)}}{\mu} = \frac{\sqrt{2(\theta^3+6\theta^2+12\theta+6)}}{\theta^2+4\theta+6} \tag{4.10}$$

The cumulant generating function of SBPL distribution

The c. g. f of SBPL distribution is

$$K(t) = \log m(t) = log\left[\frac{\theta^3 e^t(3+\theta-e^t)}{(2+\theta)(1+\theta-e^t)^3}\right], \tag{4.11}$$

The expression for recursive relation of cumulants is given by

$$K_{r+1} = \frac{1}{\theta(\theta+2)}\left[(1+\theta)(3+\theta) + 4 - 2^r + \sum_{j=1}^{r}(4+2\theta-2^j)\binom{r}{j}K_{r+1-j}\right], \ r > 1 \tag{4.12}$$

where
$K_1 = \frac{\theta^2+4\theta+6}{\theta(\theta+2)}$

$K_2 = \frac{2(\theta^3+6\theta^2+12\theta+6)}{\theta^2(\theta+2)^2}$ etc.

Harmonic mean of SBPL distribution

The harmonic mean of SBPL distribution is obtained as

$$\frac{1}{H} = \sum_{x=1}^{\infty} \frac{1}{x} p(x)$$

$$\frac{1}{H} = \sum_{x=1}^{\infty} \frac{1}{x} \left(\frac{\theta^3}{\theta+2}\right) \frac{x(x+\theta+2)}{(1+\theta)^{x+2}}, \qquad x = 1, 2, \ldots..$$

$$\frac{1}{H} = \frac{\theta^3}{(\theta+2)(1+\theta)} \frac{(1+2\theta)}{\theta^2}$$

$$H = \frac{(\theta+2)(1+\theta)}{\theta(1+2\theta)} \tag{4.13}$$

Harmonic mean of SBPL distribution is less than the mean of this distribution for all values of $\theta > 0$.

The survival function or reliability function of SBPL distribution

The reliability function of SBPL distribution is

$$R(x) = 1 - F(x) = \frac{(\theta+2)(1+\theta) - \theta^3 \sum_{m=1}^{x} \frac{m(m+\theta+2)}{(1+\theta)^m}}{(\theta+2)(1+\theta)}, x = 1,2, \ldots. \tag{4.14}$$

The hazard rate function of SBPL distribution

The hazard rate function of SBPL distribution is given by

$$h(x) = \frac{\theta^3 x(x+\theta+2)}{(1+\theta)^x [(\theta+2)(1+\theta) - \theta^3 \sum_{m=1}^{x} \frac{m(m+\theta+2)}{(1+\theta)^m}]}, \quad x = 1, 2, 3, \ldots. \tag{4.15}$$

Generalized form of SBPL distribution:

Let X is a random variable with generalized Poisson-Lindley (GPL) distribution with parameter α and θ . Then the generalized size-biased PL distribution of X is given by

$$f^*(x;\alpha,\theta) = \frac{x}{\mu} f(x;\alpha,\theta)$$
$$= \frac{\Gamma(x+\alpha)}{(x-1)!\Gamma(\alpha+1)} \frac{\theta^{\alpha+2}}{(1+\theta)^{x+\alpha}\{1+\alpha(1+\theta)\}} \left(\alpha + \frac{x+\alpha}{1+\theta}\right), \quad x = 1,2, \ldots, \tag{4.16}$$

where $\theta > 0$ and $\alpha > 0$ be the two parameter of the distribution and $\mu = \frac{1+\alpha(1+\theta)}{\theta(1+\theta)}$ be the mean of GPL distribution.

The probability generating function (p. g. f) of (4.19) is

$$G_{SBGPL}(t) = \left(\frac{\theta}{\theta-t+1}\right)^{\alpha+2} \frac{t\{\alpha(\theta-t+2)+1\}}{\alpha(1+\theta)+1} \tag{4.17}$$

Remark 4.1: Note that, in (4.16) as$\alpha \to 0$ then it reduces to one parameter size-biased Poisson-Lindley (SBPL) distribution with parameter θ as a limiting form.

5. ESTIMATION OF PARAMETER IN THE SBPL DISTRIBUTION

In this section, we obtain estimates of the parameter for the SBPL distribution by employing method of moment and ratio of the first two relative frequencies.

5.1. Method of moment estimator

Let x_1,x_2,…,x_n be a random sample of size n from the SBPL distribution with p.m.f (4.1), then the moment estimator of $\hat{\theta}$ is obtained by setting the mean of the distribution equal to the sample mean, that is $E(x) = \mu$.

The MoM estimate $\hat{\theta}$ of θ is obtained as

$$\bar{x} = \mu$$

$$\Rightarrow \bar{x} = \frac{\theta^2 + 4\theta + 6}{\theta(\theta + 2)}$$

$$\Rightarrow \hat{\theta} = \frac{2 - \bar{x} + \sqrt{\bar{x}^2 + 2\bar{x} - 2}}{\bar{x} - 1}, \quad \bar{x} > 1$$

5.2. Ratio of the first two relative frequencies

For SBPL distribution, θmay be estimated by taking ratio of the first two relative frequencies as

$$\hat{\theta} = \frac{-(4f_2 - 2f_1) + \sqrt{(4f_2 - 2f_1)^2 - 4f_2(3f_2 - 8f_1)}}{2f_2}$$

where $\frac{f_1}{N} = \frac{\theta^3(\theta+3)}{(\theta+2)(\theta+1)^3}$and $\frac{f_2}{N} = \frac{\theta^3 2(\theta+4)}{(\theta+2)(1+\theta)^4}$be the first relative frequencies of SBPL distribution.

6. APPLICATION

To illustrate the applications and to justify suitability of SBPL distribution in a practical application, fitting this distribution to some data sets. In Table 1, we have considered immunogold assay data and animal abundance data in Table 2for fitting of the SBPL distribution. [Data sets are from Ghitany and Al-Mutairi (2008).]

Fitting of distribution to immunogold assay data

Cullen *et al.* (1990) gave counts of sites with 1, 2, 3, 4 and 5 particles from immunogold assay data. The counts were 122, 50, 18, 4, and 4.

For the problem chosen, $N = 198 \; and \; \bar{x} = 1.576$.

The null hypothesisH_0:" Distribution of the data is SBPL." againstH_1:"Distribution of data is not SBPL."

Table 1: Chi-square goodness-of-fit test for the SBP and SBPL distributions to immunogold assay data.

		Expected frequencies (E_i)		
x_i	Observed frequencies (O_i)	SBP(MoM)	SBPL(MoM)	SBPL(Ratio of first two relative frequencies)
1	122	111.3	118.9	125.5
2	50	64.1	54.0	51.4
3	18	18.5	18.2	15.7
4	4	3.5	5.4	4.3
5	4	0.6	1.5	1.1
Total	198	198.0	198.0	198.0
		$\hat{\lambda}$=0.576	$\hat{\theta}$=4.048	$\hat{\theta} = 4.528$
χ^2		4.642	0.554	2.067
d.f.= $(n - p - 1)$		1	2	2

Table 1 shows that, for SBP distribution MoM estimate of $\hat{\lambda}$=0.576 [Ghitany and Al-Mutairi (2008)] and for SBPL distribution Mom estimate of $\hat{\theta}$=4.048 and $\hat{\theta} = 4.528$ obtained by using ratio of the first two relative frequencies.

It is observed from Table 1, the expected frequencies of fitted SBPL distribution by using both of the estimation procedure provides an excellent fit of the observed data. But MoM gives better result than the method of ratio of the first two relative frequencies. It is also clear from the expected SBPL frequencies that SBPL distribution gives a "good fit" to immunogold assay data.

Fitting of distribution to animal abundance data

In a study carried out by Keith and Meslow (1968), snowshoe hares were captured over 7 days. There were 261 hares caught over 7 days. Of these, 188 were caught once, 55 were caught twice, 14 were caught three times, 4 were caught four times, and 4 were five times.

For the problem chosen, $N = 261 \; and \; \bar{x} = 1.425$.

The null hypothesisH_0:" Distribution of the data is SBPL." againstH_1:"Distribution of data is not SBPL."

Table 2: Chi-square goodness-of-fit test for the SBP and SBPL distributions to animal abundance data

		Expected frequencies (E_i)		
x_i	Observed frequencies (O_i)	SBP[MoM]	SBPL[MoM]	SBPL(Ratio of relative frequencies)
1	184	170.6	177.2	188.9
2	55	72.5	62.6	56.4
3	14	15.4	16.3	12.5
4	4	2.2	3.8	2.4
5	4	0.3	1.1	0.5
Total	261	261.0	261.0	260.7
		$\hat{\lambda}$=0.425	$\hat{\theta}$=5.343	$\hat{\theta}$=6.402
χ^2 d.f.= $(n - p - 1)$		6.216 1	1.214 1	2.991 1

In Table 2, the observed frequencies together with the expected frequencies of SBP [Ghitany and Al-Mutairi (2008)], SBPL distributions are considered.

From Table 2, we have seen thatthe tabulation of expected frequencies and chi-square goodness of fit test of SBPL distribution provide a "better fit" to the data set as compared to the other one.So, we accept the hypothesis that the given data came from a SBPL distribution. Also it is seen that, MoM gives better result in fitting of SBPL distribution as compared to ratio of the first two relative frequencies.

7. CONCLUSION

In this paper, SBPL distribution is further investigated. Several properties of the distribution such as moments, cumulants, co-efficient of variation, harmonic mean, survival function, generalized form, estimation of parameter by the method of moment and ratio of the first two relative frequenciesetc. have been discussed.

Finally, the investigated distribution has been fitted to a number of reported data sets relating to bio-medical and foresty research to test its goodness of fit to which earlier the SBP distribution has been fitted and it is found that SBPL provides better fits than those by the SBP distribution. Thus, SBPL distribution found to be suitable use in foresty and bio-medical research.

ACKNOWLEDGEMENT

The first author is very much grateful to faculty members, Department of Mathematical Sciences, Tezpur University for their encouragement and valuable suggestions.

REFERENCES

1. Borah, M. & Begum, R.A.Some properties of Poisson-Lindley and its derived distribution. *Journal of the Indian Statistical Association*, 40(2002), 13-25.
2. Borah, M. & Nath Deka, A. Poisson Lindley and Some of its mixture distributions, *Pure and AppliedMathematika Sciences*,53(2000), 1-8.

3. Cullen, M. J., Walsh, J., Nicholson, L. V., and Harris, J. B. Ultrastructural localization of dystrophin in human muscle by using gold immunolabelling, *Proceedings of the Royal Society of London*, 20(1990), 197-210.
4. Das, K.K and Roy .On Some Length-Biased Weighted Weibull Distribution, *Advances in Applied Science Research*, 2(5)(2011), 465-475.
5. Fisher, R.A.The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugenics*, 6(1934), 13-25.
6. Ghitany, M. E. et al.Lindley distribution and its application, *Mathematics and Computer in Simulation*, 78(2008), 493-506.
7. Ghitany, M. E. & Al-Mutairi, D.K. Size-biased Poisson-Lindley distribution and its application, *Metron-International Journal of Statistics*, LXVI (3)(2008), 299-311.
8. Gove, H.J.Estimation and application of size-biased distributions in forestry, *Modeling Forest systems, A.Amaro,D. Reed and P.Soares CAB International Wallingford UK*,.(2003),201-212.
9. Keith, L. B. and Meslow, E. C. Trap response by snowshoe hares, *Journal of Wildlife Management*, 20(1968), 795-801.
10. Lindley, D. V.Fiducial distributions and Bayes' theorem, *Journal of the RoyalStatistical Society, Series B*, Vol.20.(1)(1958), 102-107.
11. Mahmoudi, E and Zakerzadeh, H. Generalized Poisson–Lindley Distribution, *Communications in Statistics—Theory and Methods*, 39(2010), 1785–1798.
12. Mir, K.A. Size-Biased Distributions and their Applications. *Pakistan Journal of Statistics*, 25(3)(2009). 283-294.
13. Patil, G. P. &Ord, J. K. On size-biased sampling and related form invariant weighted distributions, *Sankhya*, 38(1975), 48-61.
14. Patil, G.P. &Rao, C.R. Weighted distributions: a survey of their applications, in: P. R. krishnaiah. (Ed.), *Application of Statistics*, *Amsterdam*, *North Holland*, (1977), 383-405.
15. Patil, G.P. &Rao, C.R. Weighted distribution and size-biased sampling with applications to wildlife populations and human families, *Biometrics*, 34(1978), 179-189.
16. Sankaran, M. The discrete Poisson-Lindley distribution, *Biometrics*, 26(1970), 145-149.
17. Van Deusen, P.C. Fitting assumed distributions to horizontal point sample diameters.*Forest Science* 32(1986), 146-148.

Source of support: Nil, Conflict of interest: None Declared