

On Size Biased Poisson - Lindley Distribution and Its Applications to Model Thunderstorms

Rama Shanker^{1,*}, Hagos Fesshaye², Abrehe Yemane¹

¹Department of Statistics, Eritrea Institute of Technology, Asmara, Eritrea

²Department of Economics, College of Business and Economics, Halhale, Eritrea

Abstract In this paper firstly a general expression for the r th factorial moment of size biased Poisson-Lindley distribution (SBPLD) has been obtained and hence its first four moments about origin have been given. The expression for moment generating function and the first inverse moment of SBPLD has also been obtained. To test the goodness of fit of SBPLD over size-biased Poisson distribution (SBPD) for modeling thunderstorms, SBPLD has been fitted to a number of data sets related to thunderstorms using maximum likelihood estimate and it has been found that SBPLD gives much closer fit than SBPD, and thus SBPLD can be considered as an important alternative tool for modeling thunderstorms.

Keywords Poisson-Lindley distribution, Size-Biased distribution, Inverse moment, Estimation of parameter, Thunderstorms, 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. These distributions were firstly introduced by Fisher (1934) to model ascertainment biases which were later formalized by Rao (1965) in a unifying theory. Van Deusen (1986) 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 (1987) used size-biased distributions to analyze HPS diameter increment data. Most of the statistical applications of size biased distributions, especially to the analysis of data relating to human population and ecology, can be found in Patil and Rao (1977, 1978). Some of the recent results on size-biased distributions pertaining to parameter estimation in forestry with special emphasis on Weibull family have been reviewed by Gove (2003).

If a random variable X have distribution $f(x; \theta)$ then a simple size-biased distribution is defined by its probability function $f^*(x; \theta) = x f(x; \theta) / \mu'_1$, where μ'_1 is the mean of the original distribution.

In this paper, a general expression for the r th factorial moment of size-biased Poisson-Lindley distribution

(SBPLD) has been obtained and hence its first four moments about origin has been derived. The expression for moment generating function and the first inverse moment has also been obtained. It seems that no work has been done on the applications of SBPLD to model thunderstorms. The SBPLD has been fitted to some data sets related to thunderstorms to test its goodness of fit over SBPD and it has been found that SBPLD provides much closer fit than SBPD. This shows that the SBPLD is more flexible than SBPD for modeling thunderstorms.

2. Size-Biased Poisson-Lindley Distribution (SBPLD)

A size biased Poisson-Lindley distribution (SBPLD) given by its probability mass function (pmf)

$$P_1(x, \theta) = \frac{\theta^3}{\theta + 2} \cdot \frac{x(x + \theta + 2)}{(\theta + 1)^{x+2}}; \theta > 0, x = 1, 2, 3, \dots \quad (2.1)$$

has been obtained by Ghitany and Mutairi (2008) by size biasing the Poisson -Lindley distribution of Sankaran (1970) having pmf

$$P_2(x; \theta) = \frac{\theta^2 (x + \theta + 2)}{(\theta + 1)^{x+3}}; x = 0, 1, 2, \dots, \theta > 0. \quad (2.2)$$

It is to be mentioned that Sankaran (1970) obtained the distribution (2.2) to model count data by mixing the Poisson distribution with the Lindley (1958) distribution having pdf

$$f_1(x, \theta) = \frac{\theta^2}{\theta + 1} (1 + x) e^{-\theta x}; x > 0, \theta > 0 \quad (2.3)$$

* Corresponding author:

shankerrama2009@gmail.com (Rama Shanker)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

Further, it is to be noted that SBPLD can also be obtained by mixing size biased Poisson distribution (SBPD) having pmf

$$P_3(x|\lambda) = e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!}; x=1,2,3,\dots, \lambda > 0 \quad (2.4)$$

when its parameter λ follows the size biased Lindley distribution (SBLD) with pdf

$$f_2(\lambda; \theta) = \frac{\theta^3}{\theta+2} \lambda(1+\lambda)e^{-\theta\lambda} \quad \lambda > 0, \theta > 0 \quad (2.5)$$

We have

$$\begin{aligned} P_1(x, \theta) &= \int_0^\infty P_3(x|\lambda) f_2(\lambda; \theta) d\lambda \\ &= \int_0^\infty e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \cdot \frac{\theta^3}{\theta+2} \lambda(1+\lambda)e^{-\theta\lambda} d\lambda \\ &= \frac{\theta^3}{\theta+2} \frac{1}{(x-1)!} \int_0^\infty e^{-(\theta+1)\lambda} (\lambda^x + \lambda^{x+1}) d\lambda \\ &= \frac{\theta^3}{\theta+2} \left[\frac{x}{(\theta+1)^{x+1}} + \frac{x(x+1)}{(\theta+1)^{x+2}} \right] \\ &= \frac{\theta^3}{\theta+2} \cdot \frac{x(x+\theta+2)}{(\theta+1)^{x+2}}; \theta > 0, x=1,2,3,\dots \quad (2.6) \end{aligned}$$

which is the size biased Poisson-Lindley distribution (SBPLD).

The SBPLD has been extensively studied by Ghitany and Mutairi (2008) and they have discussed its various properties. The SBPLD has been generalized by many researchers. Shanker and Mishra (2015) introduced a size biased two parameter Poisson-Lindley distribution (SBTPPLD) by size biasing the two parameter Poisson-Lindley distribution of Shanker and Mishra (2014), which has been obtained by compounding Poisson distribution with a two parameter Lindley distribution introduced by Shanker and Mishra (2013 a). A size biased quasi Poisson-Lindley distribution has been introduced by Shanker and Mishra (2013 b) by size biasing a quasi Poisson-Lindley distribution of Shanker and Mishra (2015), which is the Poisson mixture of a quasi Lindley distribution introduced by Shanker and Mishra (2013 c). Shanker (2013) obtained a size biased version of discrete two parameter Poisson-Lindley distribution of Shanker *et al* (2012), which is a Poisson Mixture of a two parameter Lindley distribution for modeling waiting and survival times data of Shanker *et al* (2013). Further, Shanker *et al* (2014) obtained size biased new quasi Poisson-Lindley distribution by size biasing the new quasi Poisson-Lindley distribution of Shanker and Tekie (2014), which is a Poisson mixture of a new quasi Lindley distribution introduced by Shanker and Amanuel (2013).

3. Moments

The r th factorial moment of the SBPLD (2.1) can be obtained as

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

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

From (2.6), we get

$$\begin{aligned} \mu'_{(r)} &= \int_0^\infty \left[\sum_{x=r}^\infty x^{(r)} \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} \right] \frac{\theta^3}{\theta+2} \lambda(1+\lambda)e^{-\theta\lambda} d\lambda \\ &= \int_0^\infty \left[\lambda^{r-1} \sum_{x=r}^\infty x \frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} \right] \frac{\theta^3}{\theta+2} \lambda(1+\lambda)e^{-\theta\lambda} d\lambda \quad (3.2) \end{aligned}$$

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

$$\begin{aligned} \mu'_{(r)} &= \int_0^\infty \lambda^{r-1} \left[\sum_{x=0}^\infty (x+r) \frac{e^{-\lambda} \lambda^x}{x!} \right] \\ &\quad \frac{\theta^3}{\theta+2} \lambda(1+\lambda)e^{-\theta\lambda} d\lambda \end{aligned}$$

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

$$\mu'_{(r)} = \frac{\theta^3}{\theta+2} \int_0^\infty (\lambda+r) \lambda^r (1+\lambda) e^{-\theta\lambda} d\lambda \quad (3.3)$$

Using gamma integral and little algebraic simplification, we get finally a general expression for the r th factorial moment of the SBPLD as

$$\begin{aligned} \mu'_r &= \frac{\Gamma(r+1)}{(\theta+2)\theta^r} \left[(\theta+r+2)(r+1) + r\theta(\theta+r+1) \right]; \\ r &= 1,2,3,\dots \quad (3.4) \end{aligned}$$

Substituting $r=1,2,3$ and 4 in (3.4), the first four factorial moments can be obtained and then using the relationship between factorial moments and moments about origin, the first four moments about origin of the SBPLD were obtained as

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

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

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

$$\begin{aligned} \mu'_4 = 1 + & \frac{30(\theta+3)}{\theta(\theta+2)} + \frac{126(\theta+4)}{\theta^2(\theta+2)} \\ & + \frac{240(\theta+5)}{\theta^3(\theta+2)} + \frac{120(\theta+6)}{\theta^4(\theta+2)} \end{aligned} \quad (3.8)$$

and the variance of SBPLD is obtained as

$$\mu_2 = \frac{2(\theta^3 + 6\theta^2 + 12\theta + 6)}{\theta^2(\theta+2)^2} \quad (3.9)$$

Generating Function: The probability generating function of SBPLD is given by

$$\begin{aligned} P_X(t) = E(t^X) &= \frac{\theta^3}{\theta+2} \sum_{x=1}^{\infty} t^x \frac{x(x+\theta+2)}{(\theta+1)^{x+2}} \\ &= \frac{\theta^3}{(\theta+1)^2(\theta+2)} \\ &\quad \left[\sum_{x=1}^{\infty} x^2 \left(\frac{t}{\theta+1} \right)^x + (\theta+2) \sum_{x=1}^{\infty} x \left(\frac{t}{\theta+1} \right)^x \right] \\ &= \frac{\theta^3}{(\theta+1)^2(\theta+2)} \frac{t(\theta+2-t)}{(\theta+1-t)^2}. \end{aligned}$$

The moment generating function of SBPLD is thus obtained as

$$M_X(t) = E(e^{tX}) = \frac{\theta^3}{(\theta+1)^2(\theta+2)} \frac{e^t(\theta+2-e^t)}{(\theta+1-e^t)^2}.$$

First Inverse Moment: The first inverse moment of SBPLD is given by

$$\begin{aligned} \mu'_{-1} = E\left(\frac{1}{X}\right) &= \sum_{x=1}^{\infty} \frac{1}{x} \frac{\theta^3}{\theta+2} \frac{x(x+\theta+2)}{(\theta+1)^{x+2}} \\ &= \frac{\theta^3}{(\theta+2)(\theta+1)^2} \sum_{x=1}^{\infty} \left[x \left(\frac{1}{\theta+1} \right)^x + (\theta+2) \left(\frac{1}{\theta+1} \right)^x \right] \\ &= \frac{\theta(\theta^2 + 3\theta + 1)}{(\theta+2)(\theta+1)^2}. \end{aligned}$$

4. Estimation of Parameters

4.1. Maximum Likelihood (ML) Estimates

Let (x_1, x_2, \dots, x_n) be a random sample of size n from the SBPLD (2.1). Let f_x be the observed frequency in the sample corresponding to $X = x$ ($x = 1, 2, 3, \dots, k$) 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 SBPLD (2.1) is given by

$$L = \left(\frac{\theta^3}{\theta+2} \right)^n \frac{1}{(\theta+1)^{\sum_{x=1}^k f_x(x+2)}} \prod_{x=1}^k [x^2 + x(\theta+2)]^{f_x}$$

The log likelihood function is given by

$$\begin{aligned} \log L &= n \log \left(\frac{\theta^3}{\theta+2} \right) - \sum_{x=1}^k f_x (x+2) \log(\theta+1) \\ &\quad + \sum_{x=1}^k f_x \log [x^2 + x(\theta+2)] \end{aligned}$$

The maximum likelihood estimate, $\hat{\theta}$ of θ is the solution of the equation $\frac{d \log L}{d\theta} = 0$ and is given by solution of the following non-linear equation

$$\frac{3n}{\theta} - \frac{n}{\theta+2} - \frac{n(\bar{x}+2)}{\theta+1} + \sum_{x=1}^k \frac{f_x}{(x+\theta+2)} = 0 \quad (4.1.1)$$

where \bar{x} is the sample mean. It has been shown by Ghitany and Mutairi (2008) that the ML estimator $\hat{\theta}$ of θ is consistent and asymptotically normal.

4.2. Estimates from Moments

Let x_1, x_2, \dots, x_n be a random sample of size n from the SBPLD (2.1). Equating the first moment about origin to the sample mean, the method of moment (MOM) estimate, $\tilde{\theta}$, of θ is given by

$$\tilde{\theta} = \frac{(2 - \bar{x}) + \sqrt{\bar{x}^2 + 2\bar{x} - 2}}{\bar{x} - 1}; \bar{x} > 1 \quad (4.2.1)$$

where \bar{x} is the sample mean. It has been shown by Ghitany and Mutairi (2008) that the MOM estimator $\hat{\theta}$ of θ is positively biased, consistent and asymptotically normal.

5. Applications of SBPLD to Model Thunderstorms

The Poisson distribution is a suitable model for the situations where events seem to occur at random such as the number of customers arriving at a service point, the number of telephone calls arriving at an exchange, the number of fatal traffic accidents per week in a given state, the number of radioactive particle emissions per unit of time, the number of meteorites that collide with a test satellite during a single orbit, the number of organisms per unit volume of some fluid, the number of defects per unit of some materials, the number of flaws per unit length of some wire, etc. However, the Poisson distribution requires events to be independent- a

condition which is rarely satisfied completely. In thunderstorm activity, the occurrence of successive thunderstorm events (THE's) is often dependent process meaning that the occurrence of a THE indicates that the atmosphere is unstable and the conditions are favorable for the formation of further thunderstorm activity. The negative binomial distribution (NBD) is a possible alternative to the Poisson distribution when successive events are possibly dependent [see Johnson *et al*, 2005]. The theoretical and empirical justification for using the NBD to describe THE activity has been fully explained and discussed by Falls *et al* (1971). Further, for fitting Poisson distribution to the count data equality of mean and variance should be satisfied. Similarly, for fitting NBD to the count data, mean should be less than the variance. In THE, these conditions are not fully satisfied.

As a model to describe the frequencies of thunderstorms (TH's), given an occurrence of THE, the size biased Poisson distribution (SBPD) or zero truncated Poisson distribution (ZTPD) can be considered. But ZTPD does not give satisfactory fit to THE due to the basic assumption that events are independent with stable probabilities. Further,

SBPD also does not give satisfactory fit due to the reason that it is under-dispersed ($\mu > \sigma^2$).

The theoretical and empirical justification for the selection of the SBPLD to describe THE activity is that SBPLD is over-dispersed ($\mu < \sigma^2$), equi-dispersed ($\mu = \sigma^2$), and under-dispersed ($\mu > \sigma^2$) for $\theta < (=) > \theta^* = 1.671162$ respectively.

The data for thunderstorms activity at Cape Kennedy, Florida and immediate surroundings for the period January 1957 to December 1967, reported by Carter (2001), has been used for showing the applications of SBPLD and SBPD to model thunderstorms.

The expected frequencies according to the SBPD have also been given in these tables for ready comparison with those obtained by the SBPLD. The estimate of the parameter has been obtained by the maximum likelihood estimation.

It can be seen that the SBPLD gives much closer fits than the SBPD and thus provides a better alternative to the SBPD and it can be recommended for modeling thunderstorm events.

Table 5.1. Frequencies of thunderstorm events containing X thunderstorms at Cape Kennedy for May

X	Observed frequency	Expected frequency	
		SBPD	SBPLD
1	87	83.2	85.6
2	25	30.5	26.6
3	5	5.6	6.2
4	3	0.7	1.6
Total	120	120.0	120.0
Estimate of parameter		$\hat{\theta} = 0.366667$	$\hat{\theta} = 6.129082$
χ^2		1.624	0.123
d.f.		1	1
p-value		0.2025	0.7258

Table 5.2. Frequencies of thunderstorm events containing X thunderstorms at Cape Kennedy for June

X	Observed frequency	Expected frequency	
		SBPD	SBPLD
1	182	178.2	182.5
2	55	58.9	52.0
3	6	9.7	11.0
4	5	1.2	2.5
Total	248	248.0	248.0
Estimate of parameter		$\hat{\theta} = 0.330645$	$\hat{\theta} = 6.741400$
χ^2		0.407	0.637
d.f.		1	1
p-value		0.5235	0.4248

Table 5.3. Frequencies of thunderstorm events containing X thunderstorms at Cape Kennedy for August

X	Observed frequency	Expected frequency	
		SBPD	SBPLD
1	201	194.2	199.5
2	60	68.3	59.9
3	10	12.0	13.3
4	3	1.4	2.6
5	2	0.1	0.7
Total	276	276.0	276.0
Estimate of parameter		$\hat{\theta} = 0.351449$	$\hat{\theta} = 6.374771$
χ^2		1.414	0.165
d.f.		1	1
p-value		0.2344	0.6846

Table 5.4. Frequencies of thunderstorm events containing X thunderstorms at Cape Kennedy for September

X	Observed frequency	Expected frequency	
		SBPD	SBPLD
1	122	114.0	117.9
2	35	44.1	38.4
3	5	8.5	9.3
4	4	1.1	2.0
5	2	0.3	0.4
Total	168	168.0	168.0
Estimate of parameter		$\hat{\theta} = 0.386905$	$\hat{\theta} = 5.839181$
χ^2		2.561	0.485
d.f.		1	1
p-value		0.1095	0.4862

Table 5.5. Frequencies of thunderstorm events containing X thunderstorms at Cape Kennedy for Fall

X	Observed frequency	Expected frequency	
		SBPD	SBPLD
1	170	161.7	166.2
2	47	56.9	49.9
3	7	10.0	11.1
4	4	1.2	2.2
5	2	0.2	0.6
Total	230	230.0	230.0
Estimate of parameter		$\hat{\theta} = 0.352174$	$\hat{\theta} = 6.365473$
χ^2		2.372	0.313
d.f.		1	1
p-value		0.1235	0.5758

Table 5.6. Frequencies of thunderstorm events containing X thunderstorms at Cape Kennedy for Spring

X	Observed frequency	Expected frequency	
		SBPD	SBPLD
1	174	168.6	173.1
2	50	58.1	51.1
3	10	10.0	11.2
4	4	1.3	2.6
Total	238	238.0	238.0
Estimate of parameter		$\hat{\theta} = 0.344538$	$\hat{\theta} = 6.489747$
χ^2		1.947	0.03
d.f.		1	1
p-value		0.1629	0.8625

6. Conclusions

In this paper, a general expression for the r th factorial moment of SBPLD has been obtained and hence first four moments about origin has been given. The expressions for moment generating function and the first inverse moment has also been derived. To test the applicability and superiority of SBPLD over SBPD for modeling thunderstorms, SBPLD has been fitted to a number of data sets related to thunderstorms and it has been found that it provides a much closer fit than SBPD and hence it is recommended to be an important tool for modeling thunderstorms.

ACKNOWLEDGEMENTS

The authors are grateful for the constructive suggestions of the anonymous reviewer.

REFERENCES

- [1] Carter, M.C. (2001): A model for Thunderstorm Activity: Use of the Compound Negative Binomial-Positive Binomial Distribution, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 2, 196- 201.
- [2] Falls, L.W., Williford, W.O. and Carter, M.C. (1971): Probability distributions for thunderstorm activity at Cape Kennedy, Florida, *J. of Appl Meteor*, 10, 97- 104.
- [3] Fisher, R.A. (1934): The effects of methods of ascertainment upon the estimation of frequency, *Ann. Eugenics*, 6, 13-25.
- [4] Ghitany, M.E., Al-Mutairi, D.K. (2008): Size- biased Poisson-Lindley distribution and its applications, *Metron*, Vol. LXVI (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] Johnson, N.L., Kemp, A.W. and Kotz, S. (2005): *Univariate Discrete Distributions*, 3rd edition, John Wiley & Sons Inc.
- [7] Lappi, J. and Bailey, R. L. (1987): Estimation of diameter increment function or other relations using angle-count samples, *Forest Science*, 33, 725 – 739.
- [8] Lindley, D.V. (1958): Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B*, 20, 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*, 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, 320-332.
- [12] Sankaran, M. (1970): The discrete Poisson-Lindley distribution, *Biometrics*, 26, 145 – 149.
- [13] Shanker, R. (2013): Size biased discrete two parameter Poisson-Lindley distribution and its applications, *Global Journal of Pure and Applied Mathematics*, 9 (4), 325-334.
- [14] Shanker, R. and Mishra, A. (2013 a): A two-parameter Lindley distribution, *Statistics in Transition new Series*, 14 (1), 45-56.
- [15] 15. Shanker, R. and Mishra, A. (2013 b): On size biased quasi Poisson-Lindley distribution, *International Journal of Probability and Statistics*, Vol. 2(2), pp. 28-34.
- [16] Shanker, R. and Mishra, A. (2013 c): A quasi Lindley distribution, *African journal of Mathematics and Computer Science Research*, 6(4), 64-71.
- [17] Shanker, R. and Mishra, A. (2014): A two-parameter Poisson-Lindley distribution, *International Journal of Statistics and Systems*, 9 (1), 79-85.

- [18] Shanker, R. and Mishra, A. (2015): A quasi Poisson-Lindley distribution (submitted).
- [19] Shanker, R. and Mishra, A. (2015): On size biased two parameter Poisson-Lindley distribution (Submitted).
- [20] Shanker, R., Sharma, S., and Shanker, R. (2012): A Discrete two-Parameter Poisson Lindley Distribution, *Journal of Ethiopian Statistical Association*, 21, 15-22.
- [21] Shanker, R., Sharma, S., and Shanker, R. (2013): A two-parameter Lindley distribution for modeling waiting and survival times data, *Applied Mathematics*, 4, 363-368.
- [22] Shanker, R. and Amanuel, A. G. (2013): A new quasi Lindley distribution, *International Journal of Statistics and Systems*, 8(2), 143-156.
- [23] Shanker, R. and Tekie, A.L. (2014): A new quasi Poisson-Lindley distribution, *International Journal of Statistics and Systems*, 9(1), 87-94.
- [24] Shanker, R., Shanker, R., Tekie, A.L. and Abraham, Z. (2014): Size-biased new quasi Poisson-Lindley distribution and its applications, *International Journal of Statistics and Systems*, 9(2), 109-118.
- [25] Van Deusen, P.C. (1986): Fitting assumed distributions to horizontal point sample diameters, *Forest Science*, 32, 146 – 148.