

On Size-Biased Logarithmic Series Distribution and Its Applications

Khurshid Ahmad Mir*

Department of Statistics, Govt. College (Boys), Baramulla, Kashmir, India

Abstract: In this paper, a size-biased logarithmic series distribution (SBLSD), a particular case of the weighted logarithmic series distribution, taking the weights as the variate values is defined. The moments and recurrence relation of (SBLSD) are obtained. Negative moments and inverse ascending factorial moments of the size-biased logarithmic series distribution have been derived in terms of hyper-geometric function. Recurrence relations for these moments have also been derived using properties of hyper-geometric functions. Different estimation methods for the parameter of the model are discussed. R- Software has been used for making a comparison among the three different estimation methods and with the logarithmic series distribution.

Key Words: Size-biased logarithmic series distribution, Negative moments, Inverse ascending factorial moments, Bayes' estimator, Beta distribution, R-Software.

1. INTRODUCTION

The logarithmic series distribution (LSD) characterized by a parameter α is given by

$$P(X = x) = - \frac{1}{\log(1-\alpha)} \frac{\alpha^x}{x}; x=1, 2, \dots \quad (1)$$

The model (1) is a limiting form of zero-truncated negative binomial distribution. Negative moments of discrete distributions, mainly the binomial, Poisson and negative binomial have been investigated by various authors [Stephan [1], Grab and Savage [2], Mendenhall and Lehman [3], Govindarajulu [4,5], Tiku [6], Stancu [7], Chao and Strawderman [8], Gupta [9,10], Cressie *et al.* [11], Cressie and Borkent [12] and Roohi[13]. Inverse ascending factorial moments have only been dealt with by Lepage [14] and Jones [15]. Best *et al.* [16] discussed the test of fit for the model (1). Sadinle [17] linked the negative binomial distribution with the logarithmic series and Shanumugam [18] studied the characterization of model (1). A brief list of authors and their works can be seen in Johnson, Kotz and Kemp [19].

The first four moments of LSD are given as

$$\mu'_1 = \theta(1-\alpha)^{-1} \alpha, \text{ where} \quad (2)$$

$$\mu'_2 = \theta(1-\alpha)^{-3} \alpha(1-\alpha) \quad (3)$$

$$\mu'_3 = \theta(1-\alpha)^{-5} \alpha(1-\alpha)(1-\alpha^2) \quad (4)$$

$$\mu'_4 = \theta(1-\alpha)^{-7} \alpha(1-\alpha)(\alpha^4 + 2\alpha^3 - 6\alpha^2 + 2\alpha + 1) \quad (5)$$

The variance of LSD is given as

$$\mu_2 = \theta(1-\alpha)^{-3} \alpha[1-2\alpha+\alpha^2] \quad (6)$$

In this paper, we have made an attempt to study proposed size-biased LSD, its moments and recurrence relations. Negative moments and inverse ascending factorial moments of the size-biased logarithmic series distribution have been derived in terms of hyper-geometric function. Recurrence relations for these moments have also been derived using properties of hyper-geometric functions. In order to make a comparative analysis among the three estimation methods for the parameter of the size-biased logarithmic series distribution (SBLSD), one of the standard software packages R-Software is used which is meant for data analysis and graphics.

2. SIZE-BIASED LOGARITHMIC SERIES DISTRIBUTION (SBLSD)

A size-biased logarithmic distribution (SBLSD) is obtained by taking the weight of the LSD (1) as x .

We have from (1) and (2)

$$\sum_{x=1}^{\infty} x \cdot P(X = x) = \frac{\alpha\theta}{1-\alpha}, \quad \theta = - \frac{1}{\log(1-\alpha)}$$

$$\sum_{x=1}^{\infty} \theta \alpha^x = \frac{\alpha\theta}{1-\alpha}$$

This gives the size-biased logarithmic series distribution (SBLSD) as

$$P_1[X = x] = \{ {}^2F_1[1, A; A; \alpha] \}^{-1} \alpha^{x-1}; x=1, 2, \dots \quad (7)$$

Where $0 < \alpha < 1$ and $\{ {}^2F_1[1, A; A; \alpha] \}^{-1} = (1-\alpha)$

*Address correspondence to this author at the Department of Statistics, Govt. College (Boys), Baramulla, Kashmir, India; Tel: 91-1942426584; Fax: 91-19424265; E-mail: khrshdmir@yahoo.com

2.1. Moments

The r^{th} moment $\mu'_r(s)$ of SBLSD (7) about origin is obtained as

$$\mu'_r(s) = E(X^r) = \sum_{x=1}^{\infty} x^r P_1[X = x] \quad ; r=1,2,\dots$$

$$\mu'_r(s) = \sum_{x=1}^{\infty} x^r (1-\alpha)\alpha^{x-1} \tag{8}$$

obviously $\mu'_0(s) = 1$ and for $r \geq 1$

$$\begin{aligned} \mu'_r(s) &= \frac{1-\alpha}{\alpha} \sum_{x=1}^{\infty} x^r \alpha^x \\ &= \frac{1-\alpha}{\alpha\theta} \sum_{x=1}^{\infty} x^r \theta \alpha^x \\ &= \frac{1-\alpha}{\alpha\theta} \sum_{x=1}^{\infty} x^{r+1} P[X = x] \end{aligned}$$

$$\mu'_r(s) = \frac{1-\alpha}{\alpha\theta} \mu'_{r+1} \tag{9}$$

where μ'_{r+1} is the $(r + 1)^{\text{th}}$ moments about origin of LSD (1). The moments of SBLSD can be obtained by using relations (2) to (4) in (9)

$$\mu'_1(s) = \frac{1}{(1-\alpha)} \tag{10}$$

Using relation (4) in (9), we get

$$\mu'_2(s) = \frac{(1-\alpha)(1-\alpha^2)}{(1-\alpha)^4} \tag{11}$$

Which gives the variance of SBLSD (7) as

$$\mu_2(s) = \frac{1}{(1-\alpha)^2} \tag{12}$$

The higher moments of SBLSD (7) about origin can also be obtained by using (9) if so desired.

3. RECURRENCE RELATION OF MOMENTS ABOUT ORIGIN OF SIZE-BIASED LSD

The recurrence relation can be obtained by differentiating (8) as

$$\begin{aligned} \frac{\partial \mu'_r(s)}{\partial \alpha} &= \sum_{x=1}^{\infty} x^r \left[\frac{\partial}{\partial \alpha} \{ \alpha^{x-1} (1-\alpha) \} \right] \\ &= \frac{1}{\alpha} \mu'_{r+1}(s) - \frac{1}{\alpha} \mu'_r(s) - \frac{1}{(1-\alpha)} \mu'_r(s) \end{aligned}$$

$$\mu'_{r+1}(s) = \alpha \frac{\partial \mu'_r(s)}{\partial \alpha} + \frac{1}{(1-\alpha)} \mu'_r(s) \tag{13}$$

For $r = 0$, we get

$$\mu'_1(s) = \frac{1}{(1-\alpha)}, \text{ where } \mu'_0(s) = 1$$

The second moment of (7) about origin can also be obtained by using the relation (13)

4. NEGATIVE MOMENTS AND INVERSE ASCENDING FACTORIAL MOMENTS

Theorem I: Suppose the random variable X has a size-biased logarithmic series distribution with parameter α , then the relation

$$E(X + A)^{-1} = \{ (A + 1)^2 F_1[1, A; A; \alpha] \}^{-1} {}^2F_1[1, A + 1; A + 2; \alpha]$$

holds.

Proof: Since X has a size-biased logarithmic series distribution, then

$$\begin{aligned} E(X + A)^{-1} &= \{ {}^2F_1[1, A; A; \alpha] \}^{-1} \sum_{x=1}^{\infty} \frac{\alpha^{x-1}}{(x + A)} \\ &= \{ (A + 1)^2 F_1[1, A; A; \alpha] \}^{-1} {}^2F_1[1, A + 1; A + 2; \alpha] \end{aligned} \tag{14}$$

This completes the proof.

Theorem II: Suppose the random variable X has a size-biased logarithmic series distribution with parameter α , then the k^{th} inverse ascending factorial moment is given as

$$\mu'_{-[k]} = \{ (k + 1)^2 F_1[1, A; A; \alpha] \}^{-1} {}^2F_1[1, 1; k + 2; \alpha], k = 1, 2, \dots \text{ and } 0 < \alpha < 1.$$

Proof: Here $\mu'_{-[k]} = E\left[\prod_{i=1}^k \left(\frac{1}{X + i} \right) \right]$

$$\begin{aligned} &= \{ {}^2F_1[1, A; A; \alpha] \}^{-1} \sum_{x=1}^{\infty} \frac{\alpha^{x-1}}{(x+1)(x+2)\dots(x+k)} \\ &= \{ {}^2F_1[1, A; A; \alpha] \}^{-1} \left[\frac{1}{(k+1)!} + \frac{\alpha}{(k+2)!} + \frac{\alpha^2 \cdot 2}{(k+3)!} + \dots \right] \end{aligned}$$

On simplification, we get

$$\mu'_{-[k]} = \{ (k + 1)^2 F_1[1, A; A; \alpha] \}^{-1} {}^2F_1[1, 1; k + 2; \alpha] \tag{15}$$

5. RECURRENCE RELATION FOR NEGATIVE ASCENDING FACTORIAL MOMENTS

Theorem III: Suppose the random variable X has a size-biased logarithmic series distribution with parameter α and $\mu'_{-[k]}$ is the k^{th} inverse ascending factorial moment of X, then the relation

$$\begin{aligned} (k + 1)^2 \alpha \mu'_{-[k+1]} &= [(k + 1)(\alpha - 1) + k\alpha] \mu'_{-[k]} + \\ &(1 - \alpha) \mu'_{-[k-1]}, k = 2, 3, \dots \text{ and } 0 < \alpha < 1 \end{aligned}$$

holds.

Proof: we know that

$$\begin{aligned} \mu'_{-[k]} &= \{ (k + 1)^{-1} {}^2F_1[1, A; A; \alpha] \}^{-1} {}^2F_1[1, 1; k + 2; \alpha] \\ \mu'_{-[k+1]} &= \{ (k + 2)^{-1} {}^2F_1[1, A; A; \alpha] \}^{-1} {}^2F_1[1, 1; k + 3; \alpha] \end{aligned}$$

Using the identity (see Rainville [20], page 71)

$$(a - c + 1) {}^2F_1[a, b; c; z] = a {}^2F_1[a + 1, b; c; z] - (c - 1) {}^2F_1[a, b; c - 1; z], \tag{16}$$

for $a=1, b=1, c=k+3, z=\alpha$, we get

$${}^2F_1[1, 1; k + 3; \alpha] = \frac{k + 2}{k + 1} {}^2F_1[1, 1; k + 2; \alpha] - \frac{1}{k + 1} {}^2F_1[2, 1; k + 3; \alpha] \tag{17}$$

Now again using the identity given by (Rainville [20], page 71)

$$(z - 1) {}^2F_1[a, b; c; z] = a {}^2F_1[a - 1, b; c; z] - \left(\frac{c - b}{c}\right) z {}^2F_1[a, b; c + 1; z], \text{ for } a=2, b=1, c=k+2, z=\alpha, \text{ we get}$$

$${}^2F_1[1, 2; k + 3; \alpha] = \frac{k + 2}{(k + 1)\alpha} {}^2F_1[1, 1; k + 2; \alpha] - \frac{(k + 2)(1 - \alpha)}{(k + 1)\alpha} {}^2F_1[2, 1; k + 2; \alpha] \tag{18}$$

Substituting (18) in (17) we get

$${}^2F_1[1, 1; k + 3; \alpha] = \left[\frac{k + 2}{k + 1}\right] \left[1 - \frac{1}{\alpha(k + 1)}\right] {}^2F_1[1, 1; k + 2; \alpha] + \frac{(k + 2)(1 - \alpha)}{(k + 1)^2 \alpha} {}^2F_1[2, 1; k + 2; \alpha]$$

Again using (5.1), for $a=1, b=1, c=k+2$ and $z=\alpha$, we have

$${}^2F_1[2, 1; k + 2; \alpha] = \left[\frac{k + 2}{k + 1}\right] \left[1 - \frac{1}{\alpha(k + 1)} - \frac{k(1 - \alpha)}{\alpha(k + 1)}\right] {}^2F_1[1, 1; k + 2; \alpha] + \frac{(k + 2)(1 - \alpha)}{(k + 1)\alpha} {}^2F_1[1, 1; k + 1; \alpha]$$

$$\mu'_{-[k+1]} = \{(k + 2)^{-1} {}^2F_1[1, A; A; \alpha]\}^{-1} \left[\frac{k + 2}{k + 1}\right] \left[1 - \frac{1}{\alpha(k + 1)} - \frac{k(1 - \alpha)}{\alpha(k + 1)}\right] {}^2F_1[1, 1; k + 2; \alpha] + \frac{(k + 2)(1 - \alpha)}{(k + 1)\alpha} {}^2F_1[1, 1; k + 1; \alpha]$$

This gives

$$\hat{\alpha} = 1 - \frac{1}{\bar{x}} \tag{20}$$

$$\mu'_{-[k+1]} = \left[\frac{2k\alpha - k - 1 + \alpha}{\alpha(k + 1)^2}\right] \mu'_{-[k]} + \frac{(1 - \alpha)}{\alpha(k + 1)^2} \mu'_{-[k-1]}$$

$$(k + 1)^2 \alpha \mu'_{-[k+1]} = [(k + 1)(\alpha - 1) + k\alpha] \mu'_{-[k]} + (1 - \alpha) \mu'_{-[k-1]}, k = 2, 3, \dots \tag{19}$$

Where ${}^2F_1[1, 1; k + 2; \alpha] = (k + 1)! {}^2F_1[1, A; A; \alpha] \mu'_{-[k]}$ and

$${}^2F_1[1, 1; k + 1; \alpha] = (k)! {}^2F_1[1, A; A; \alpha] \mu'_{-[k-1]}$$

6. ESTIMATION METHODS

In this section, we discuss the various estimation methods for size-biased logarithmic series distribution and verify their efficiencies.

6.1. Method of Moments

In the method of moments replacing the population mean $\mu'_1 = \frac{1}{(1 - \alpha)}$ by the corresponding sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ we get}$$

6.2. Method of Maximum Likelihood

Let X_1, X_2, \dots, X_n be a random sample from size-biased logarithmic series distribution, then corresponding likelihood function is given as

$$L = (1 - \alpha)^n \alpha \sum_{x_i=1}^{x_i=n} \tag{21}$$

$$L = (1 - \alpha)^n \alpha^{y-n}, \text{ where } y = \sum x_i \tag{22}$$

The log likelihood function can be written as $\log L = n \log(1 - \alpha) + (y - n) \log \alpha$ and the likelihood equation as

$$\frac{\partial \log L}{\partial \alpha} = \frac{-n}{(1 - \alpha)} + \frac{(y - n)}{\alpha} = 0$$

On solving, we get the maximum likelihood estimate as

$$\hat{\alpha} = 1 - \frac{1}{\bar{x}}, \text{ which coincides with the moment estimate.}$$

6.3. Bayesian Method of Estimation

Since $0 < \alpha < 1$, therefore we assume that prior information about α is from beta distribution.

$$\text{Thus } f(\alpha) = \frac{\alpha^{a-1} (1-\alpha)^{b-1}}{B(a,b)}; 0 < \alpha < 1, a > 0, b > 0. \quad (23)$$

The posterior distribution from (22) and (23) can be written as

$$\prod(\alpha/y) = \frac{(1-\alpha)^{n+b-1} \alpha^{y+a-n-1}}{\int_0^1 (1-\alpha)^{n+b-1} \alpha^{y+a-n-1} d\alpha} \quad (24)$$

The Bayes' estimator of α is given as

$$\begin{aligned} \hat{\alpha} &= \int_0^1 \alpha \prod(\alpha/y) d\alpha \\ \hat{\alpha} &= \frac{\int_0^1 (1-\alpha)^{n+b-1} \alpha^{y+a-n} d\alpha}{\int_0^1 (1-\alpha)^{n+b-1} \alpha^{y+a-n-1} d\alpha} \\ &= \frac{y+a-n}{y+a+b} \end{aligned} \quad (25)$$

For $a=b=0$, we get moment and mle estimate. Thus Bayes' estimator works as a general estimate in comparison to moment and mle estimator.

7. COMPUTER SIMULATION AND CONCLUSIONS

It is very difficult to compare the theoretical performances of different estimators proposed in the previous section. Therefore, we perform extensive simulations to compare the performances of different methods of estimation mainly with respect to their biases and the mean squared errors (MSE's), for different sample sizes and different parametric values. Regarding the choice of values of (a, b) in Bayes' estimator $\hat{\alpha}$, there was no information about their values except that they are real and positive numbers. Therefore, 25 combinations of values of (a, b) were considered for a, b=1, 2,3,4,5 and those values of a, b were selected for which the Bayes' estimator has minimum variance. It was found that for a=b=5, the Bayes' estimator has minimum variance and χ^2 values between the simulated sample frequencies and the estimated Bayes' frequencies were least. In Table 1.1, we have fitted the LSD (1), SBLSD (7) to some zero-truncated distribution of 1534 biologists according to the number of research papers to their credit in the review of Applied Entomology, vol 24, 1936 (see Williams [21]). Data given in Table 1.2 is on species frequency distribution of insect catches from Kempton [22]. As evident from the data, the number of moths per species is 128+, hence the data has a very long tail. Consequently, the fit by simple logarithmic series distribution is poor as is evident from the value of the Pearson's chi-square. However, the fit given by size-biased logarithmic series distribution reflects that the size-biased phenomenon is working in the sense that if species with more and more moths are included in the study, then those species will have a higher probability of being represented in the sample.

Table 1.1. Distribution of 1534 Biologists According to the Number of Research Papers to their Credit in The Review of Applied Entomology, Vol. 24, 1936

No. of Papers Per Author	No. of Authors	Expected Frequency		
		LSD	SBLSD	
			MLE	Bayes'
1	1062	1058.69	1061.26	1061.95
2	263	262.36	262.54	262.91
3	120	118.25	119.25	120.00
4	50	51.36	49.21	49.89
5	22	21.54	21.75	22.00
6	7	8.54	7.54	6.98
7	6	7.36	6.21	6.51
8	2	3.25	2.68	1.98
9	0	1.25	0.24	0.21
10	1	1.65	1.65	0.97
11	1	0.44	1.67	0.60
Total	1534	1534	1534	1534
χ^2		1.48	0.542	0.0074
$\hat{\alpha}$		0.35525	0.35525	0.36

Table 1.2. Species Frequency Distribution of Insect Catches from the Trap at Rathamsted

Moths per Species	Observed Frequency	Expected Frequency		
		LSD	SBLSD	
			MLE	Bayes'
1	33	36.25	32.4	32.78
2+	24	23.01	23.3	23.58
4+	22	16.98	21.7	21.95
8+	9	8.26	8.34	8.93
16+	5	8.25	4.97	4.23
32+	4	4.58	4.28	3.98
64+	0	3.59	2.62	0.23
128+	5	1.08	4.39	6.32
Total	102	102	102	102
χ^2		21.05	2.812	0.656
$\hat{\alpha}$		0.458	0.589	0.621

ACKNOWLEDGEMENT

The author is highly thankful to the editor and the two anonymous referees for their valuable suggestions.

REFERENCES

- [1] F.F. Stephan, "The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate," *Annals of Mathematical Statistics*, vol. 16, pp. 50-61, 1945.
- [2] E.L. Grab and I.R. Savage, "Tables of the expected value of 1/X for positive Bernoulli and Poisson variables", *Journal of American Statistical Association*, vol. 49, pp.169-177, 1954.
- [3] W. Mendenhall, and E.H. Jr. Lehman, "An approximation to the negative moments of the positive binomial useful in life testing", *Technometrics*, vol. 2, pp. 227-241, 1960.
- [4] Z. Govindarajulu, "The reciprocal of the decapitated negative binomial variable", *Journal of American Statistical Association*, vol. 57, pp. 906-913, 1962.
- [5] Z. Govindarajulu, "Recurrence relations for the inverse moments of the positive binomial variable", *Journal of American Statistical Association*, vol. 58, pp. 467-473, 1963.
- [6] M.L. Tiku, "A note on the negative moments of a truncated Poisson variate", *Journal of American Statistical Association*, vol. 59, pp. 1220-1224, 1964.
- [7] D.D. Stancu, "On the moments of negative order of the positive Bernoulli and Poisson variables", *Studia Universitatis Babeş Bolyai Series, Mathematics and Physics*, vol. 1, pp. 29-31, 1968.
- [8] M.T. Chao and W.E. Strawderman, "Negative moments of positive random variables". *Journal of American Statistical Association*, vol. 67, pp. 429-431, 1972.
- [9] R.C.Gupta, "Modified power series distribution and some of its applications", *Sankhya*, vol. 36, pp. 288-298, 1974.
- [10] R.C.Gupta, "Estimating the probability of winning (losing) in a gamblers ruin problem with applications", *Journal of Statistical Planning and Inference*, vol. 9, pp. 55-62, 1984.
- [11] N. Cressie, A.S. Davids, J.L. Folks and G.E. Policello, "The moment generating function and negative integer moments". *American Statistician*, vol. 35, pp.148-150, 1981.
- [12] N. Cressie and M. Borkent, "The moment generating function has its moments". *Journal of American Statistical Association*, vol. 13, pp. 337-344, 1986.
- [13] A. Roohi, *Unpublished M.Phil dissertation*, University of Lahore, Lahore, 2002.
- [14] Y. Lepage, "Negative factorial moments of positive random variables", *Industrial Mathematics*, vol. 28, pp. 95-100, 1978.
- [15] M.C. Jones, "Inverse factorial moments", *Statistics and Probability Letters*, vol. 6, pp. 37-42, 1987.
- [16] D.J. Best, J.C.W. Rayner and O. Thas, "Test of fit for the logarithmic distributions". *Journal of Applied Mathematics and Decision Sciences*, vol. 2008, pp. 1-8, 2008.
- [17] M. Sadinle, "Linking the negative binomial and logarithmic series distributions via their associated series". *Revista Colombiana de Estadística*, vol. 31, pp. 311-319, 2008.
- [18] R. Shanumugam and J. Singh, "A characterization of the logarithmic series distribution and its applications." *Communications in Statistics*, vol. 13, pp. 865-875, 1984.
- [19] N.L. Johnson, S. Kotz and A. W. Kemp, "Univariate discrete distributions", 3rd ed. John Wiley & Sons inc; Hoboken, New Jersey, 2005.
- [20] E.D. Rainville, *Special Functions*, Chelsea publishing company, Bronx New York, 1960.
- [21] C.B. William, "The number of publications written by biologists", *Annals of Eugenics*, vol. 12, pp. 143-146, 1944.
- [22] R.A. Kempton, "A generalized form of Fisher's logarithmic series". *Biometrika*, vol. 62, pp. 29-38, 1975.

Received: June 11, 2009

Revised: August 13, 2009

Accepted: August 29, 2009

© Khurshid Ahmad Mir; licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.