A New Method of Estimation of Size-Biased Generalized Logarithmic Series Distribution

Khurshid Ahmad Mir[*]

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

Abstract: In this paper, a size-biased generalized logarithmic series distribution (SBGLSD) is introduced and its moments are obtained. The estimates of the parameters of SBGLSD are obtained by employing the method of moments and a proposed new method of estimation. The new proposed method of estimation uses the non-zero frequency of a variable only up to a finite value. In this method, the estimation of only one parameter is needed and of the other is obtained by the relationship among the parameters by counting the number of non-zero frequency classes. The method is found very simple and quick to apply in practice. Extensive simulations are performed to compare the performances of the proposed and the moment method of estimation mainly with respect to their biases and mean squared errors (MSE's), for different sample sizes and of different parametric values. Comparison has been made among different estimation methods by means of Pearson's Chi-square, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) techniques.

Key Words: Size-biased generalized logarithmic series distribution, Non- zero frequency classes, Chi-square AIC, BIC.

1. INTRODUCTION

The generalized logarithmic series distribution (GLSD) characterized by two parameters α and β was defined by Jain and Gupta [1]. The probability function of the GLSD model is given by

$$P[X = x] = \frac{\theta \; \Gamma(\beta x)\, \alpha^x \,(1-\alpha)^{\beta x - x}}{x!\; \Gamma(\beta x - x + 1)};$$

$$x = 1, 2, ..., \quad \beta \geq 1 \text{ and } 0 \leq \alpha \leq \beta^{-1}. \qquad (1)$$

Where $\quad \theta = \dfrac{-1}{log(1-\alpha)}.$

The model (1) reduces to the simple logarithmic series distribution when $\beta = 1$. The GLSD model is a member of Gupta's [2] modified power series distribution and of Consul and Shenton's [3] Lagrangian probability distributions. The model (1) is also a limiting form of zero-truncated form of Jain and Consul's [4] generalized negative binomial distribution. Patel [5] defined GLSD and obtained the estimates of the parameters by the method of moments. Famoye [6] showed that the GLSD is unimodal and the mode is at the point x = 1. Some methods of sampling from the model (1) are provided by Famoye [7]. Famoye [8] obtained the moment estimators, Jani and Shah [9] discussed the maximum likelihood and moment method of estimation for two parameter GLSD model. Mishra and Tiwary [10] suggested an

alternative method of estimation based on the first three moments and showed that the GLSD provides a very close fits to the observed data from various fields such as asentomolgy, medicine, engineering etc. Famoye [11] discussed the fitting of GLSD. Tripathi and Gupta [12] studied another generalization of GLSD. A brief list of authors and their works can be seen in Johnson, Kotz and Kemp [13] and Consul and Famoye [14].

The first four moments about origin of GLSD are given as

$$\mu_1' = \grave{e}(1-\alpha\,\hat{a})^{-1}\alpha. \qquad (2)$$

$$\mu_2' = \theta(1-\alpha\beta)^{-3}\alpha(1-\alpha). \qquad (3)$$

$$\mu_3' = \theta(1-\alpha\beta)^{-5}\alpha(1-\alpha)(1-2\alpha+2\alpha\beta-\alpha^2\beta). \qquad (4)$$

$$\mu_4' = \theta(1-\alpha\beta)^{-7}\alpha(1-\alpha)\begin{pmatrix} 1-6\alpha+6\alpha^2+2\alpha\beta \\ (4-9\alpha+4\alpha^2) \end{pmatrix} + \qquad (5)$$

$$\beta^2\alpha^2(6-6\alpha+\alpha^2).$$

The recurrence relation among the central moments is given as

$$\mu_{r+1} = \frac{\alpha(1-\alpha)}{1-\alpha\beta}\frac{d\mu_r}{d\alpha} + r\mu_2\mu_{r-1}. \qquad (6)$$

Which gives the first four central moments as

$$\mu_2 = \theta(1-\alpha\beta)^{-3}\alpha(1-\alpha-\theta\alpha)(1-\alpha\beta). \qquad (7)$$

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

$$\mu_3 = \theta\left(1-\alpha\beta\right)^{-5}\alpha\left(1-\alpha\right)\begin{bmatrix}1-2\alpha+\alpha\beta\left(2-\alpha\right)-\\ 3\theta\alpha\left(1-\alpha\beta\right)+2\theta^2\alpha\left(1-\alpha\beta\right)^2\end{bmatrix}. \quad (8)$$

$$\mu_4 = \theta(1-\alpha\beta)^{-7}\left(\alpha(1-\alpha)(1-6\alpha+6\alpha^2+2\alpha\beta)(4-9\alpha+4\alpha^2)\right.$$
$$+\alpha^2\beta^2(6-6\alpha+\alpha^2))-4\theta\alpha^2(1-\alpha)(1-\alpha\beta)(1-2\alpha+2\alpha\beta \quad (9)$$
$$-\alpha^2\beta)+6\theta^2\alpha^3(1-\alpha)(1-\alpha\beta)^2-3\theta^3\alpha^4(1-\alpha\beta).$$

In this paper, a size-biased generalized logarithmic series distribution (SBGLSD) taking the weights of the probabilities as the variate values, is defined. The moments of the proposed model are obtained. The estimates of the parameters of SBGLSD are obtained by employing the method of moments and a proposed new method of estimation. It is very difficult to compare the theoretical performances of different estimators proposed in this paper. Therefore, we perform extensive simulations to compare the performances of the different methods of estimation mainly with respect to their biases and mean squared errors (MSE's), for different sample sizes and of different parametric values. Goodness of fit test is done in order to see that proposed new method of estimation gives better result in comparison to the method of moments.

2. SIZE-BIASED GENERALIZED LOGARITHMIC SERIES DISTRIBUTION (SBGLSD)

Size-biased distributions are a special case of the more general form known as weighted distributions. Fisher [15] introduced these distributions to model ascertainment bias and were later formalized in a unifying theory by Rao [16]. These distributions arise in practice when observations from a sample are recorded with unequal probability and provide a unifying approach for the problems where the observations fall in the non-experimental, non- replicated, and non-random categories.

If the random variable X has distribution $f(x;\theta)$, with unknown parameter θ, then the corresponding size-biased distribution is of the form

$$f^*(x;\theta) = \frac{x f(x;\theta)}{E(x)}. \quad (10)$$

where $E(x) = \int x f(x;\theta)dx$ for continuous case and $E(x) = \sum x\, p(X=x)$ for discrete case.

Using the criteria defined in equation (10) and by using the equations (1) and (2), the probability function of size-biased generalized logarithmic distribution (SBGLSD) is obtained as

$$\sum_{x=1}^{\infty} x \cdot P[X=x] = \frac{\alpha\theta}{1-\alpha\beta}, \text{ where } \theta = -\frac{1}{log(1-\alpha)} \quad (11)$$

$$\sum_{x=1}^{\infty} \frac{x\,\theta\,\Gamma(\beta x)\alpha^x(1-\alpha)^{\beta x-x}}{x!\,\Gamma(\beta x-x+1)} = \frac{\alpha\theta}{1-\alpha\beta}. \quad (12)$$

On simplification, the above equation is reduced to

$$\sum_{x=1}^{\infty}(1-\alpha\beta)\begin{pmatrix}\beta x-1\\ x-1\end{pmatrix}\alpha^{x-1}(1-\alpha)^{\beta x-x} = 1. \quad (13)$$

Since the above sum equals to 1, therefore, it represents a probability distribution and we name it as size-biased generalized logarithmic series distribution (SBGLSD) and is represented as

$$P_1[X=x] = (1-\alpha\beta)\begin{pmatrix}\beta x-1\\ x-1\end{pmatrix}\alpha^{x-1}(1-\alpha)^{\beta x-x}\;;\;x=1,2\ldots$$

$$= 0 \text{ for } x \geq t \text{ if } \beta\,t\text{-}t\text{-}1 \leq 0 \quad (14)$$

$\beta \geq 1$ and $0 < \alpha < \beta^{-1}$

When $\beta = 1$, the SBGLSD reduces to size- biased logarithmic series distribution (SBLSD) with probability function as

$$P_2[X=x] = (1-\alpha)\,\alpha^{x-1}\;;\;x=1,2\ldots \quad (15)$$

2.1. Moments

The rth moment $\mu'_r(s)$ of SBGLSD about origin is obtained as

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

$$= \sum_{x=1}^{\infty} x^r(1-\alpha\beta)\begin{pmatrix}\beta x-1\\ x-1\end{pmatrix}\alpha^{x-1}(1-\alpha)^{\beta x-x} \quad (16)$$

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

$$\mu'_r(s) = \frac{1-\alpha\beta}{\alpha\theta}\sum_{x=1}^{\infty} x^r\theta\begin{pmatrix}\beta x-1\\ x-1\end{pmatrix}\alpha^x(1-\alpha)^{\beta x-x}$$

$$= \frac{1-\alpha\beta}{\alpha\theta}\sum_{x=1}^{\infty} x^{r+1}P[X=x]$$

$$\mu'_r(s) = \frac{1-\alpha\beta}{\alpha\theta}\mu'_{r+1}. \quad (17)$$

where μ'_{r+1} is the $(r+1)^{\text{th}}$ moments about origin of GLSD (1).

The moments of SBGLSD can be obtained by using equations (3) and (4) in (17) as

$$\mu'_1(s) = \text{Mean} = \frac{(1-\alpha)}{(1-\alpha\beta)^2}. \quad (18)$$

$$\mu'_2(s) = \frac{(1-\alpha)\,(1-2\alpha+2\alpha\beta-\alpha^2\beta)}{(1-\alpha\beta)^4} \quad (19)$$

$$\mu_2(s) = Variance = \frac{(1-\alpha)(2\alpha\beta - \alpha - \alpha^2\beta)}{(1-\alpha\beta)^4}. \tag{20}$$

The higher moments of SBGLSD about origin can be obtained similarly using equation (17) if so desired.

3. ESTIMATION OF SIZE-BIASED GENERALIZED LOGARITHMIC SERIES DISTRIBUTION

In this section, we study the estimation of the parameters of SBGLSD by the method of moments and a new proposed method. Also comparison is made between these two estimators.

3.1. Method of Moments

Replacing sample moments with population moments, we get

$$\overline{x} = \frac{(1-\alpha)}{(1-\alpha\beta)^2}. \tag{21}$$

$$S^2 = \frac{(1-\alpha)(2\alpha\beta - \alpha - \alpha^2\beta)}{(1-\alpha\beta)^4}. \tag{22}$$

From above two equations, we get

$$\frac{S^2}{\overline{x}^2} = \frac{\left[1 - \sqrt{(1-\alpha/\overline{x})}\right](2-\alpha) - \alpha}{(1-\alpha)}. \tag{23}$$

Solving above equation for α, we get the estimate for α and substituting that value in equation (21), we get the estimate of β.

3.2. Proposed New Estimator for SBGLSD

In this method, only one parameter α is estimated with the help of the first moment of the SBGLSD and the other parameter β is estimated based on non-zero frequency classes. Thus, this method may be much easy and quick in practice.

The condition in the SBGLSD that

$$P[X = x] = 0 \text{ for } x \ge t \text{ if } \beta \ t\text{-}t\text{-}1 \le 0 \tag{24}$$

gives a relationship between the parameter β and the number of the classes of non-zero frequencies of the GLSD. Hence in those cases when the number of the classes of non-zero frequency is finite, β may be readily estimated using the equation (24)

Let us suppose that in a sample of size n, the first (t-1) classes have non-zero frequencies and the rest of the classes have zero frequencies, then

$$P[X = x] \ne 0 \text{ if } x < t \quad =0 \text{ if } x \ge t \tag{25}$$

From the equation (24), we have estimate of β, say β_0, as

$$\beta_0 = \frac{t+1}{t}. \tag{26}$$

Thus the value of β_0, is obtained directly from the non-zero frequency classes and may be treated as predetermined as n in the case of binomial distribution.

Now substituting the estimate of β in the expression (21) for the mean of the SBGLSD and replacing μ_1' by the sample mean $\overline{x}$, we get

$$\overline{x} = \frac{(1-\alpha)}{(1-\alpha\beta_0)^2}. \tag{27}$$

Solving this for α we get the estimate of α.

3.3. Efficiency of Proposed Estimator

In order to check the usefulness of new proposed method, the efficiency of the parameter α is studied. For this purpose, an extensive computer simulation is done by taking n=15, 20, 30, 50,100, α =0.2, 0.5, 1.0, 2.0 and β =1.10, 1.12, 1.16, 1.25, 1.5. For each combination of n and α we generate a sample of size n from SBGLSD and estimate α by different methods. We report the average values of $\left(\dfrac{\hat{\alpha}}{\alpha}\right)$ and the corresponding average MSE's. All the reported results are based on 10,000 replications. The results are presented in Table **1.1**. Here we report the average values of $\left(\dfrac{\hat{\alpha}}{\alpha}\right)$ for each method and the corresponding MSE's are reported within brackets. From the table it is immediate that the average biases and the average MSE's decrease as sample size increases. It indicates that all the methods provide asymptotically unbiased and the consistent estimators. It is also observed that the average biases and the average MSE's of $\left(\dfrac{\hat{\alpha}}{\alpha}\right)$ depend on α. On comparing the performances of all the methods it is clear that as far as the minimum bias is concerned, the proposed estimator works the best in almost all the cases.

3.4. Goodness of Fit

An attempt is made to fit the SBGLSD to observed data estimating the parameters α and β by suggested alternative method. To know how much good or bad the fits are due to this method in comparison to those due to method of moments, we have used the data sets of Guire *et al.* [17] and Student [18]. The expected frequencies according to both the methods along with the estimates of both the parameters and the values of chi-square, AIC and BIC are given in Tables **1.2** and **1.3**.

CONCLUSIONS

It is encouraging to observe from the above tables that the proposed estimator is giving best results in comparison to moment estimators. Furthermore, the suggested method has

Table-1.1. Average Relative Estimates and Average Relative Mean Squared Errors of α

n	b	Method	α =0.2	α =0.5	α =1.0	α=2.0
15	1.10	Proposed Estimator	1.046(0.216)	1.244(0.113)	1.115(0.305)	1.211(0.718)
		Moment Estimator	1.432(0.758)	1.412(0.517)	1.351(0.501)	1.366(1.201)
20	1.12	Proposed Estimator	1.041(0.201)	1.204(0.109)	1.104(0.218)	1.201(0.554)
		Moment Estimator	1.416(0.630)	1.401(0.499)	1.301(0.411)	1.297(0.254)
30	1.16	Proposed Estimator	1.011(0.145)	1.125(0.102)	1.109(0.251)	1.187(0.441)
		Moment Estimator	1.368(0.514)	1.356(0.325)	1.201(0.226)	1.202(0.154)
50	1.25	Proposed Estimator	1.024(0.036)	1.101(0.023)	1.015(0.125)	1.101(0.0254)
		Moment Estimator	1.221(0.299)	1.255(0.217)	1.154(0.119)	1.165(0.125)
100	1.5	Proposed Estimator	1.07(0.017)	1.021(0.020)	1.001(0.012)	1.021(0.021)
		Moment Estimator	1.135(0.054)	1.132(0.012)	1.012(0.031)	1.125(0.031)

Table 1.2. Zero-Truncated Data on *P. nubilalis* (European Corn Borer) of Guire *et al*. [17]

No. of Bores Per Plant	Observed Frequency	Expected Frequency	
		Method of Moments	Proposed Method
1	83	81.56	82.4
2	36	34.54	34.96
3	14	15.01	14.34
4	2	2.04	2.01
5	1	2.85	2.29
Total	136	136	136
χ^2		1.362	0.77
AIC		201	185
BIC		225	203
Estimates $\hat{\alpha}$ $\hat{\beta}$		0.567 2.451	0.754 1.2

Table 1.3. Zero-Truncated Data of Haemocytometer Yeast Cell Counts Per Square Observed by STUDENT [18]

No. of Cells Per Square	Observed Frequency	Expected Frequency	
		Method of Moments	Proposed Method
1	128	126.43	127.41
2	37	34.45	35.4
3	18	20.56	19.41
4	3	4.34	3.56
5	1	1.22	1.22
Total	187	187	187
χ^2		0.99	0.21
AIC		198	176
BIC		223	211
Estimates $\hat{\alpha}$ $\hat{\beta}$		0.451 2.367	0.652 1.2

an advantage over the method of moments in certain situations. It can be applied whenever, it is relatively very quick to be obtained and so it may be preferred to others when very quick results are required.

ACKNOWLEDGEMENT

REFERENCES

[1]   G.C. Jain, and R.C. Gupta, "A logarithmic type distribution", *Trabjos. Estadist.*, vol. 24, pp. 99-105, 1973.

[2]   R.C.Gupta, "Modified power series distribution and some of its applications", *Sankhya. Ser B,* vol. 35, pp. 288-298, 1974.

[3]   P.C. Consul, and L.R. Shenton, "Use of Lagrangian expansion for generating generalized probability distributions", *SIAM J. Appl. Math.,* vol. 23, pp.239-248, 1972.

[4]   G.C.Jain, and P.C.Consul, "A generalized negative binomial distribution", *SIAM J. Appl. Math.,* vol. 21, pp. 501-513, 1971.

[5]   I.D. Patel, "A generalization of logarithmic series distribution", *J. Indian Soc. Agric. Stat.*, vol. 19, pp. 129 -132, 1981.

[6]   F. Famoye, "A short note on the generalized logarithmic series distribution", *Stat. Probab. Lett.*, vol. 5, pp. 315-316, 1987.

[7]   F. Famoye, "Sampling from the generalized logarithmic series distribution", *Computing,* vol. 58, pp. 365-375, 1997.

[8]   F. Famoye, "On certain methods of estimation for the generalized logarithmic series distribution", *J. Appl. Stat. Sci.,* vol. 2 pp. 103-117, 1995.

[9]   P.N. Jani, and S.M. Shah, "On fitting of the generalized logarithmic series distribution", *J. Indian Soc. Agricul. Stat.,* vol. 30(3), pp. 1-10, 1979.

[10]   A. Mishra, and D. Tiwary, "On generalized logarithmic series distribution", *J. Indian. Soc. Agric. Stat.,* vol. 37(3), pp. 219-222, 1985.

[11]   F. Famoye, "Goodness of fit tests for generalized logarithmic series distribution", *J. Comput. Stat. Data Anal.,* vol. 33, pp. 59-67, 2000.

[12]   R.C. Tripathi, and R.C. Gupta, "Another generalization of the logarithmic series and the Geometric distributions", *Commun. Stat.,* vol. 17, pp. 1541-1547, 1988.

[13]   N.L.Johnson, S. Kotz and A.W.Kemp, "*Univariate discrete distributions*", 3rd ed, Wiley, 2005.

[14]   P.C. Consul and F. Famoye, *"On the Lagrangian probability distributions"*, Birkauser publications, 2006.

[15]   R.A.Fisher, "The effects of methods of ascertainment upon the estimation of frequencies", *Ann. Eugen.,* vol. 6, pp. 13-25, 1934.

[16]   C.R. Rao, "On discrete distributions arising out of methods of ascertainment" in *Classic. Contag Discrete Distributions*. G.P. Patil, Ed. Statistical Publishing Society, Calcutta, pp. 320-332,1965.

[17]   M. Guire, U. Judson, T.A. Brindley, and T.A Bancroft, "The distribution of European corn borer larvae pyrausta nubilalis in field corn", *Biometrics*, vol. 13, pp. 65-78, 1957.

[18]   Student, "On the error of counting with haemocytometer", *Biometrika.*, vol. 5, pp. 351-360, 1907.