Available online at http://www.journalijdr.com

International Journal of DEVELOPMENT RESEARCH

ISSN: 2230-9926

International Journal of Development Research
Vol. 4, Issue, 2, pp. 234-239, February, 2014

Full Length Research Article

AN ESTIMATION OF SIZE-BIASED GENERALIZED LOGARITHMIC SERIES DISTRIBUTION

[1,*]Jemilohun, Vincent Gbenga, [2]Adebara, Lanre and [3]Adejuwon Samuel Oluwaseun

[1]College of Sciences, Department of Statistics, Afe Babalola University, Ado-Ekiti, Ekiti State, Nigeria
[2]Federal Polytechnic Ado – Ekiti, Ekiti State, Nigeria
[3]Academic Planning Unit, Afe Babalola University, Ado-Ekiti, Ekiti State, Nigeria

ARTICLE INFO

Article History:
Received 17th November, 2013
Received in revised form
11th December, 2013
Accepted 29th January, 2014
Published online 21st February, 2014

Key words:
Akaike Information Criterion
Generalized Logarithmic Series Distribution,
Mean Square Error,
Moments,
Pearson's Chi-square.

ABSTRACT

This paper sets out to introduce a Size-biased Generalized Logarithmic Series Distribution (SBGLSD). An attempt is also made at obtaining the estimates of the parameters of SBGLSD by employing the method of moments and a proposed new method using the non-zero frequency of a variable up to a finite value. Comparison is also made among different estimation methods by means of Pearson's Chi-square, Akaike Information Criterion (AIC) and Bayesian Information Criterion Techniques (BIC). It was observed that the proposed estimator gives better results in comparison to moment estimators. This suggested method has an advantage over the method of moments.

INTRODUCTION

Jain and Gupta (2000) defined the Generalized Logarithmic Series Distribution (GLSD) characterized by two parameters α and β. The probability function of the model is given by

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

$$X=1, 2, \qquad \ldots\ldots\ldots\ldots(1)$$

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

$$\text{Where } \theta = \frac{-1}{\text{Log}(1-\alpha)}$$

The model (1) reduces to simple logarithmic series distribution when β = 1. GLSD is a member of Gupta's (2002) modified power series distribution and of Consul and Shenton's (2004) Lagrangian probability distributions. The model (1) above is also a limiting

Corresponding author: Jemilohun, Vincent Gbenga
College of Sciences, Department of Statistics, Afe Babalola University, Ado-Ekiti, Ekiti State, Nigeria

form of zero-truncated form of Jain and Consul (2003) generalized negative binomial distribution. Patel (2001) defined GLSD and obtained the estimates of the parameters by the method of moments. Famoye (2000) showed that the GLSD is unimodal and the mode is at the point x = 1. Famoye (2002) obtained the moment estimators, Jani and Shah (2005) discussed the maximum likelihood and moment method of estimation for two parameter GLSD model. Mishra and Tiwary (2006) suggested an alternative 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 medicine and engineering. Famoye (2003) discussed the fitting of GLSD. Tripathi and Gupta (2007) studied another generalization.

The first four moments about origin of GLSD are given as

$\mu'_1 = \theta\,(1-\alpha\,\hat{a})^{-1}\,\alpha$ ……………………(2)
$\mu'_2 = \theta\,(1-\alpha\beta)^{-3}\,\alpha\,(1-\alpha)$ ……………………(3)
$\mu'_3 = \theta\,(1-\alpha\beta)^{-5}\,\alpha\,(1-\alpha)(1-\alpha)(1-2\alpha+2\alpha\beta-\alpha^2\beta)$ ……………………(4)
$\mu'_4 = \theta(1-\alpha\beta)^{-7}\alpha(1-\alpha)(1-6\alpha+6\alpha^2+2\alpha\beta\,(4-9\alpha+4\alpha^2))+\beta^2\,\alpha^2\,(6-6\alpha+\alpha^2)$ ……………………(5)

The recurrence relation among the central moments is given as

$$\mu_{r+1} = \frac{\alpha\,(1-\alpha)}{1-\alpha\beta}\ \frac{-d\ \mu r}{d\alpha} + r\,\mu_2\,\mu_{r-1} \qquad \text{……………………(6)}$$

Which gives the first four central moments is

$\mu_2 = \theta\,(1-\alpha\beta)^{-3}\,\alpha\,(1-\alpha-\theta\alpha)(1-\alpha\beta)$ ……………(7)
$\mu_3 = \theta\,(1-\alpha\beta)^{-5}\,\alpha\,(1-\alpha)\,[1-2\alpha+\alpha\beta\,(2-\alpha)-3\theta\alpha\,(1-\alpha\beta)+2\theta^2\alpha\,(1-\alpha\beta)^2]$ …………… (8)
$\mu_4 = \theta\,(1-\alpha\beta)^{-7}\,[\alpha\,(1-\alpha)(1-6\alpha+6\alpha^2+2\alpha\beta)\,(4-9\alpha+4\alpha^2)+\alpha^2\,\beta^2(6-6\alpha+\alpha^2)]-4\theta\alpha^2\,(1-\alpha)(1-\alpha\beta)$ ……………..(9)
$(1-2\alpha+2\alpha\beta-\alpha^2\,\beta)+6\theta^2\alpha^3\,(1-\alpha)\,(1-\alpha\beta)^2-3\theta^3\,\alpha^4\,(1-\alpha\beta)$.

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 parameters of SBGLSD are obtained by employing the method of moments and proposed new methods of estimation. It is very difficult to compare the theoretical performance of different estimator proposed in this paper. Therefore we perform extensive simulations to compare the performances of 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.

Size-Biased Generalized Logarithmic Series Distribution (SBGLSD)

Size-biased distributions are a special case of the more general form known as weighted distributions. Fishers (1934) introduced these distributions to model ascertainment bias and were later formalized on a unifying theory by Rao (1965). These distributions arise in practice when observation 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;θ), then the corresponding size-biased distribution is of the form

$$f^*(x;\theta) = \frac{xf(x;\theta)}{E\,(x)} \qquad \text{………………. (10)}$$

Where $E(x) = \int xf(x;\theta)\,dx$ for continuous case and $E(x) = \sum xp(X=x)$ for discrete case.

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

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

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

On simplification, the above equation is reduced to

$$\sum_{x=1}^{\infty} (1-\alpha\beta) \begin{bmatrix} \beta x & - & 1 \\ x & - & 1 \end{bmatrix} \alpha^{x-1}(1-\alpha)^{\beta x-x} = 1 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

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

$$P_1[X = x] = (1 - \alpha\beta) \begin{bmatrix} \beta x & - & 1 \\ x & - & 1 \end{bmatrix} \alpha^{x-1}(1-\alpha)^{\beta x-x};\ x = 1,2,\ldots$$
$$= 0 \text{ for } x \geq t \text{ if } \beta\, t\text{-}t\text{-}1 \leq 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

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

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

Moments

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

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

$$= \sum_{x=1}^{\infty} x^r (1-\alpha\beta) \begin{bmatrix} \beta x & - & 1 \\ x & - & 1 \end{bmatrix} \alpha^{x-1} (1-\alpha)^{\beta x-x} \qquad \ldots\ldots\ldots\ldots . \quad (16)$$

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

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

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

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

Where μ'_{r+1} is the $(r+1)^{th}$ moments about origin of GLSD (1)
The moments of SDLSD can be obtained by using equations (3) and (4) in (7) as

$$\mu'_1(s) = \text{mean} = \frac{(1-\alpha)}{(1-\alpha\beta)^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

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

$$\mu_2(s) = \text{variance} = \frac{(1-\alpha)(2\alpha\beta-\alpha-\alpha^2\beta)}{(1-\alpha\beta)^4} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (20)$$

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

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

Method of Moments

Replacing sample moments with population moments we got

$$\bar{x} = \frac{(1-\alpha)}{(1-\alpha\beta)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(21)}$$

$$S^2 = \frac{(1-\alpha)(2\alpha\beta - \alpha - \alpha^2\beta)}{(1-\alpha\beta)^4} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(22)}$$

From above two equations, we get

$$\frac{S^2}{x^2} = \frac{[1 - \sqrt{(1-\alpha/\bar{x})}\ ](2-\alpha) - \alpha}{(1-\alpha)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(23)}$$

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

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 for x ≥ t if β t-t-1 ≤ 0…(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 equation (24)

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

$P[X = x] \neq 0$ if $x<t = 0$ if $x\geq$ …………………… (25)

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

$$\beta_0 = \frac{t+1}{t} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(26)}$$

Thus the values 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 x, we get

$$\bar{X} = \frac{(1-\alpha)}{(1-\alpha\beta_0)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(27)}$$

Solving this for α, we get the estimate of α.

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 taken 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 SBGLSB and estimate α by different methods. We report the average value of [$\frac{\bar{\propto}}{\propto}$] 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 [$\frac{\bar{\propto}}{\propto}$] 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 decreases 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 α depend on [$\frac{\bar{\propto}}{\propto}$]. On comparing the performances of all methods, it is clear that as far as the minimum bias is concerned, the proposed estimator works the best in almost all the cases.

Goodness of Fit

An attempt was made to fit the SBGLSD to observed data estimating the parameter α and β by suggested alternative method. To know how much good or bad the fits are due to this method of moments, we have used the data set of Guire et al (2007) the distribution of European corn borer larvae *pyransta nubilalis* in field corn and Strod (2005) on the error of counting with haemocytometer. The expected frequencies according to both the methods along with values of Chi-square, AIC and BIC are given in Tables 1.2 and 1.3.

Conclusion and Recommendation

It is encouraging to observe from the appendix that the proposed estimator gives better results in comparison to moment estimators. In addition, the suggested method has an advantage over the method of moments in certain situations. It can be applied in urgent situation since it is relatively very quick to obtain hence it may be preferred to other method when the results are needed urgently.

REFERENCES

Jain, G.C. and Gupta, R.C. A logarithmic type distribution. Trab-jos. Esta-distist., *25,* 2000, 99-105.

Gupta, R.C. Modified power series distribution and some of its applications. Sankhya. Ser B, *35*, 2002, 288-298.

Consul, P.C. and Shenton, L.R. Use of Lagrangian expansion for generating generalized probability distribution. *SIAM Journal of Applied Mathematics*, *23*, 2004, 239 – 248.

Jain, G.C. and consul, P.C. A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics*, *21*, 2003, 501-513.

Patel, I.D. A generalization of logarithmic series distribution. *Journal of Indian society of Agricultural statistics*, *19*, 2001, 129-132.

Famoye, F. A short note on generalized logarithmic series distribution. *Statistical Probability Lett*., *5,* 2000, 315-315.

Famoye, F. On certain methods of estimation for the generalized logarithmic series distribution. *Journal of Applied statistics*, *2*, 2002, 103-117.

Jain, P.N. and Shah, S.M. On fitting of the generalized logarithmic series distribution. *Journal of Indian society of Agricultural statistics*, *3*(*3*), 2005, 1-10.

Mishra, A. and Tiwary, D. On generalized logarithmic series distribution. *Journal of Indian society of Agricultural statistics*. *37*(*3*), 2006, 219-222.

Famoye, F. Goodness of fit tests for generalized logarithmic series distribution. *Journal of computer statistics Data Analysis*, *33*, 2003, 59-67.

Tripathi, R.C. and Gupta, R.C. Another generalization of the logarithmic series and the Geometric distributions. *Commun. Stat*. *17*, 2007, 1541-1547.

Fisher, R.A. The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugen*, *6*, 1934, 13 – 25.

Rao, C.R. On discrete distributions arising out of method of ascertainment. In classic. Contag Discrete Distributions. G.P. Patil, Ed. Statistical Publishing society, Calcutta, 1965, 320-332.

Guire, M., Judson, U., Brindley, T.A., and Bancroft, T.A. The distribution of European corn borer larvae *pyranta nubilalis* in field corn. *Biometrics 13*, 2007, 65-78.

APPENDICES

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.

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			
$\hat{a}$		0.567	0.754
$\hat{\beta}$		2.451	1.2

Table 1.3. Zero-Truncated Data of Haemocytometer Yeast cell Counts Per Square Observed by Student

No. of Bores Per Plant	Observed Frequency	Expected Frequency	
		Method of Moments	Proposed Method
1	128	126.43	127.41
2	37	34.45	34.4
3	18	20.56	19.41
4	3	4.34	3.56
5	1	1.22	1.22
Total	136	187	187
χ^2		0.99	0.21
AIC		198	176
BIC		223	211
Estimates			
$\hat{\alpha}$		0.451	0.652
β		2.367	1.2

****