ON FITTING OF THE GENERALIZED LOGARITHMIC SERIES DISTRIBUTION

BY

P. N. JANI AND S.M. SHAH

Sardar Patel University, Vallabh Vidyanagar

(Received: May, 1978)

1. INTRODUCTION

A researcher working in a practical field has always a problem that which distribution, out of several distributions, will be most suitable fitted to his observed data. Fisher *et. al.* [2], Williams [8], [9], Bliss [1] and many others have fitted negative binomial distribution and logarithmic series distribution (LSD) to ecological and biological data. LSD is widely used distribution for fitting zero-truncated discrete observed data. Recently Jain and Gupta [3] and Jani [4] have generalized the LSD and obtained a new distribution called the generalized logarithmic series eistribution (GLSD). For a random variable X, the GLSD is defined by its probability function

$$P(X=i)=p_i(\beta, \theta) = \frac{\Gamma(i\beta)}{i!\,\Gamma(i\beta-i+1)}\alpha(\theta(1-\theta)^{\beta-1})^i, \qquad ...(1.1)$$

where $\alpha=(-\log(1-\theta))^{-1}$, $0<\theta<1$, $\beta\theta<1$ and $i=1,2,\ldots\ldots$
and

$$p_i(\beta, \theta)=0, \text{ whenever } i\beta-i+1\leq 0 \qquad ...(1.2)$$

The mean, the variance and the recurrence relation for higher moments given by Jani (1977) were

$$\mu=\alpha\theta/(1-\beta\theta) \qquad ...(1.3)$$

$$\mu_2=\alpha\theta(1-\theta-\alpha\theta+\alpha\beta\theta^2)/(1-\beta\theta)^3 \qquad ...(1.4)$$

$$\mu_{r+1}=\frac{\theta(1-\theta)}{1-\beta\theta}\frac{d\mu_r}{d\theta}+r.\,\mu_2\mu_{r-1} \qquad ...(1.5)$$

Since the GLSD (1.1) is a generalation of the LSD it is of interest to determine whether it may provide satisfactory fits to some observed distributions which are not well fitted, and the superior fits

to the data which are wellfitted, by the LSD. The paper contains the properties of the GLSD in the Section 2. In the Section 3, the problem of estimation of parameters by the methods of maximum likelyhood (M.L.) and moments has been considered. The asymptotic variances—covariances of the estimators obtained by both the methods have been derived. The M.L. equations are very complicated to solve hence, in Section 4, the asymptotic efficiency of the method of moments relative to the M.L. method has been computed for different values of the parameters θ and β. It is shown that the method of moments is an efficient as the M.L. method. I section 5, three zero-truncated biological observed data with a good fit, a poor fit and a worse fit by the LSD have been considered. All resulted in better fits by the GLSD as measured by the test of probabilities of χ^2.

2. Properties of the GLSD

Since the additional parameter β in GLSD (1.1) characterizes the distribution, one has to be very careful while selecting a value of β. We shall study a few properties of the GLSD which will help us in fitting.

(*i*) For $\beta=1$, the GLSD (1.1) reduces to the usual LSD with probability function

$$P(X=i)=p_i(\theta)=\alpha\theta^i/i.$$

(*ii*) The probability function of the GLSD of (1.1) can be written as

$$P(X=i)=p_i(\beta, \theta)=(i\beta\theta-\theta)\ (i\beta\theta-2\theta)\ldots(i\beta\theta-i\beta\theta-i\theta+\theta)\ (1-\theta)^{i\beta-1}/(i!.\ \log(1-\theta)^{-1/\theta}).$$

Taking limit $\beta\to\infty$ and $\theta\to 0$, such that $\beta\theta$ remains constant, equal to say δ, we get

$$P(X=i)=p_i(\delta)=(i\delta)^{i-1}.\ e^{-i\delta}/i!,\ i=1, 2, \ldots,$$

which is the probability function of the Borel distribution with one parameter δ and is a particular case of the generalized Poisson distribution.

(*iii*) Since the parameter β is connected with the mean and the variance of the GLSD (1.1), a value of β must depend on the observed mean and the variance. From (1.3) and (1.4) we deduce that the mean is equal, greater or smaller than the standard deviation (s.d) according to β is equal, smaller or greater than the term $f(\theta)=(1-(1-\theta)/2\alpha\theta)/\theta$ respectively. The Table—1 shows the values of $f(\theta)$ for different values of θ.

(*iv*) (1.4) shows that if $\beta<(1-(1-\theta)/\alpha\theta)/\theta=2f(\theta)-1/\theta$, then $\mu_2<0$.

(*v*) The identity

$$\sum_{i=1}^{\infty} p_i=1$$

is, in general, exact only when the series is infinite. However, the GLSD (1.1), due to (1.2), for $\beta<1$, the series is finite and for some values of β and θ, $\Sigma\, p_i>1$. For example, with $\beta\leqslant 0.5$, $0<\theta<1$, there will be only one non-zero term *i.e.* p_1 and in all such cases, except with very small θ, $p_1>1$. Similarly, for $\beta=0.6$, there will be two non-zero terms, for $\beta=0.7$, three non-zero terms and soon. The Table-2 shows the adproximate values of

$$\sum_{i=1}^{n} P_i.$$

(*vi*) For $\beta\theta\geqslant 1$, $\sum_{i=1}^{\infty} p_i<1$.

For example, with $\beta=2$ and

$\theta=0.9$, $p_1=0.3909$,

$p_2=0.00528$, $p_3=0.00105$,

$p_6=0.00002$.

(*vii*) For $\beta\geqslant 1$, as $\beta\theta$ increases the distribution will have longer tail. Perhaps, for this reason the observed data with short tails are better fitted by the GLSD with $\beta<1$ rather than the LSD where $\beta=1$.

TABLE 1

θ :	0.1	0.2	0.3	0.4	0.5
$f(\theta)$:	5.25878	2.76856	1.9626	1.954220	1.30685
θ :	0.6	0.7	0.8	0.9	
$f(\theta)$:	1.15762	1.06001	0.99852	0.96898	

TABLE 2

Approximate values for series $\sum_{i=1}^{n} p_i$

	$\theta=0.1$	$\theta=0.5$	$\theta=0.9$
$\beta=0.1$ ($n=1$)	1.04353	1.34608	3.10475
$\beta=0.3$ ($n=1$)	1.02177	1.17183	1.95897
$\beta=0.5$ ($n=1$)	1.00046	1.02014	1.23612
$\beta=0.7$ ($n=3$)	1.00002	1.00359	1.10607
$\beta=0.9$ ($n=9$)	1.00000	1.00000	1.00783

3. Estimation of Parameters

The M.L. method is the most efficient method for estimating the parameters but, sometimes, it involves so complicated forms of M.L. equations that they are difficult to solve for M.L. estimators. In this case, some other efficient estimators are to be find out. In this section we will study two methods for estimating the parameters namely the M.L. method and the method of moments.

3·1 The M.L. method

Consider a random sample of size N from the population (1.1) and let N_i be the observed frequency in the sample corresponding to $X=i$. Then, the likelihood function L is given by

$$L=\prod_{i=1}^{\infty} p_i^{N_i},$$

Taking natural logarithm of L, differentiating w.r.t. θ and β and equating with zero and writing $\Sigma N_i=N$ and $\Sigma i.\ N_i/N=\bar{X}$, we obtain the M.L. equations

$$\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} i.\ N_i/(i\hat{\beta}-j)=N\bar{X}/\hat{\alpha} \qquad ...(3.1)$$

$$1/\hat{\theta}-\hat{\alpha}/\bar{X}=\hat{\beta} \qquad ...(3.2)$$

which can be solved for $\hat{\theta}$ and $\hat{\beta}$ the M.L. estimators of θ and β respectively, by using an iterative technique such as the method of scoring [6]. M.L. equations (3.1) and (3.2) do not yield explicit expressions for the corresponding M.L. estimators $\hat{\theta}$ and $\hat{\beta}$ and they are very complicated to solve for $\hat{\theta}$ and $\hat{\beta}$.

The Fisher information matrix R of the M.L. estimators can be found to be

$$R=N[r_{pq}], \qquad ...(3.3)$$

where the elements r_{pq}; $p, q=1, 2$ are given by

$$r_{11}=(\mu/\theta^2-\alpha^2/(1-\theta))/(1-\theta)$$

$$r_{22}=\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} i^2.\ p_i/(i\beta-j)^2 \qquad ...(3.4)$$

$$r_{12}=r_{21}=\mu/(1-\theta).$$

R^{-1} will be the asymptotic variance-covariance matrix of M.L. estimators $\hat{\theta}$ and $\hat{\beta}$.

3·2 The method of moments

Since $\mu_2<0$ for $\beta\theta<1-(1-\theta)/\alpha\theta$, in the following we will obtain the moment estimators of the parameters of the GLSD for the restricted sample space $1-(1-\theta)/\alpha\theta\leqslant\beta\theta<1$.

From (1.3) and (1.4) we have

$$\alpha^2\theta^2-Q(1-\theta)=0. \qquad ...(3.5)$$

where

$$Q=\mu^3/(\mu^2+\mu_2) \qquad ...(3.6)$$

The equation (3.5) can be solved for θ by using the method of iterations. To obtain the initial value of θ, expanding $\alpha=(-\log$

$(1-\theta))^{-1}$ into power series expansion and neglecting θ^3 and terms higher than that we get (3.5) as

$$\theta^2-12\,(Q-1)\,(1-\theta)=0$$

which gives, for $\theta>0$.

$$\theta=2(\sqrt{(Q'(Q'+1))}-Q'), \qquad \ldots(3.7)$$

where $Q'=3(Q-1)$ and Q is as given by (3.6).

To get θ^*, the moment estimator of θ, μ and μ_2 are to be replaced by their respective estimates samplemean $\bar{X}$ and sample variance S^2 of the observed data.

Using (1.3) and replacing μ by $\bar{X}$, we get

$$\beta^*=1/\theta^*-\alpha^*/\bar{X} \qquad \ldots(3.8)$$

Using the differential method [5] we obtain the asymptotic variance—covariance matrix M of the moment estimators θ^* and β^*, to the order N^{-1}, as

$$M=N^{-1}\,[m_{pq}]. \qquad \ldots(3.9)$$

The elements m_{pq} ; $p, q=1, 2$ are given by

$$\begin{aligned} m_{11}&=(A^2\mu_2+C^2\left(\mu_4-\mu_2^2\right)+2AC\mu_3)/B^2 \\ m_{12}&=m_{21}=D.m_{11}+\alpha(A\mu_2+C\mu_3)/(B\mu^2) \qquad \ldots3(.10) \\ m_{22}&=2D.m_{12}-D^2.m_{11}+\alpha^2.\mu_2/\mu_4 \end{aligned}$$

where

$$\begin{aligned} A&=\mu\,(1-\theta)\,(3\mu\,(1-\theta)-2\alpha^2\theta^2) \\ B&=2\alpha^2\theta\,(\mu^2+\mu_2)\,(1-\theta-\alpha\theta)+\mu^3\,(1-\theta) \\ C&=-\alpha^2\theta^2\,(1-\theta) \qquad \ldots(3.11) \\ D&=\alpha^2/(\mu\,(1-\theta))-1/\theta^2. \end{aligned}$$

4. Comparision of Asymptotic Efficiencies

The joint asymptotic efficiency E of the moment estimators (θ^*, β^*) relative to the M. L. estimators $(\hat{\theta}, \hat{\beta})$, discussed in Section 3, is given by

$$E=1/(\,|\,R\,|\,.\,|\,M\,|\,), \qquad \ldots(4.1)$$

where $|\,X\,|$ is the determinant of a matrix X.

Considering the restricted sample space $1-(1-\theta)/\alpha\theta \leqslant \beta\theta < 1$, the asymptotic efficiency E has been computed for $\beta=0.93$, 1.00, 1.03 and 2.03 and $\theta=0.1$ (0.2) 0.7 and tabulated in Table 3. The table shows that the method of moment is equally efficient as the M.L. method and hence one can safely use this simple method in place of M.L. method for fitting the GLSD.

TABLE 3

The asymptotic efficiencies (in %) of the method of moments relative to the M.L. method

β \ θ	*0.1*	*0.3*	*0.5*	*0.7*
0.93	100	100	98.4	90.9
1.00	99.3	97.2	94.8	91.5
1.03	98.9	95.4	90.6	81.9
2.03	88.6	70.0		

5. Fitting to the GLSD

We have fitted the GLSD (1.1) to many zero-truncated biological data and observed that in most of the cases the GLSD provides a better fit than the usual LSD. Here we present data obtained from three different samples where the fits by the ISD are good, poor and worst. In all the cases the GLSD gives better fit than the LSD. Since the M.L. equations are complicated and the method of moments is equally efficient we have used the moment estimators for fitting the GLSD. The comparision between two fits, the fit by the LSD and the fit by the GLSD, is done on the basis of the values of the probability integrals of the χ^2 values $\left\{ P\left(\chi^2_v \right) \right\}$ at respective degrees of freedom (D. F.) v.

The data given in Table 4 are the zero-truncated data of *P.* Garman on Counts of the number of european red mites on apple leaves where there is a good fit by the LSD. But as measured by the probabilities of two χ^2_s, $P\left(\chi^2_2 \right)$ due to the LSD fit is 0.40 while $P\left(\chi^2_1 \right)$ due to the GLSD fit is 0.69, the fit by the GLSD is superior than the fit by the LSD.

Table 5 shows the zero-truncated data of haemaycytometer yeast cell counts per square observed by Student [7] where the fit by the LSD is poor. Our fit by the GLSD is better than the LSD since $P\left(\chi_1^2\right)$ due to the GLSD fit is 0.15 while $P\left(\chi_2^2\right)$ due to the LSD fit is only 0.06.

Table 6 consists of data on *P.* nubilalis (European corn borer) of McGuire *et. al* (1957) and two fits, one by the LSD and another by the GLSD. Here $P\left(\chi_2^2\right)$ due to the LSD fit is zero while $P\left(\chi_1^2\right)$ due to the GLSD fit is 0.07. This shows that the fit by the LSD is worst while there is a good fit by the GLSD.

TABLE 4

Counts of the european red mites on apple leaves. (The zero-truncated data of P. Garman)

No. of mites per leaf	*Leaves observed*	*Expected frequency*	
		LSD	*GLSD*
1	38	43.46	39.10
2	17	16.24	17.40
3	10	8.09	9.73
4	9	4.53	5.83
5	3	2.71	3.55
6	2	1.69	2.17
7	1	1.08	1.27
⩾ 8	0	2.20	0.95
Total	80	80.00	80.00
Mean	2.1500		
s.d.	1.4504		
χ^2		1.81	0.16
D.F. (ν)		2	1
$P\left(\chi_\nu^2\right)$		0.40	0.69
Estimates θ :		0.7473	0.8898
β :			0.9129

TABLE 5

Zero-truncated data of Haemaycytometerye yeast cell counts per square observed by 'Student' [7]

No. of cells per square	*Observed No. of squares*	*Expected frequency*	
		LSD	*GLSD*
1	128	133.66	128.19
2	37	34.11	38.84
3	18	11.61	13.73
4	3	4.42	4.75
5	1	1.81	1.45
$\geqslant$ 6	0	1.39	0.04
Total	187	187.00	187.00
Mean	1.4599		
s.d.	0.7776		
χ^2		5.73	2.22
D.F. (ν)		2	2
$P\left(\chi^2_\nu\right)$		0.06	0.15
Estimates θ :		0.5104	0.7135
β :			0.8536

TABLE 6

Zero-truncated data on P. nubilalis (European corn borer) of Mc. Guire *et. al.* (1957)

No. of bores per plant	*Observed frequency*	*Expected frequency*	
		LSD	*GLSD*
1	83	99.16	86.91
2	36	24.18	30.38
3	14	7.87	12.34
4	2	2.88	4.88
$\geqslant$ 5	1	1.91	1.49
Total	136	136.00	136.00
Mean	1.5441		
s.d.	0.7969		
χ^2		13.85	3.22
D.F. (ν)		2	1
$P\left(\chi^2_\nu\right)$		0.00	0.07
Estimates θ :		0.4878	0.7895
β :			0.8510

SUMMARY

The GLSD is a generalization of the LSD and contains an additional parameter which characterizes the distribution. A few important properties, depending on the values of the additional parameter, have been discussed. Two methods, the maximum likelihood and moments, of estimating the parameters of the GLSD has been discussed. The variances and the covariances of the estimators in both the cases have been obtained. The asymptotic efficiencies of the method of moments relative to the maximum likelihood method have been derived and computed for a set of values of parameters. It is observed that the method of moments is as efficient as the maximum likelihood method. For fitting purpose, data obtained from three different samples, good fit, a poor fit and a worse fit by the LSD have been considered. All have been resulted in better fits by the GLSD.

REFERENCES

[1] Bliss, C.I. (1963) : Fitting the negative binomial distribution to biological data. Biometrics, 9, 176-200.

[2] Fisher, R.A., Corbet, A.S. and Williams, C.B. (1943) : The relation between the number of species and the number of individuals in a random sample of an animal population. Journal of animal ecology, 12, 42-57.

[3] Jain, G.C. and Gupta, R.P. (1973) : A logarithmic type distribution. Trabjos Estadist, 24, 99-105.

[4] Jani, P.N. (1977) : Minimum variance unbiased estimation for some left—truncated modified power series distributions, Sankhya, B, Vol. 39, 258-278.

[5] Kendall, M.G. and Stuart, A. (1969) : 'The advanced theory of Statistics' Vol. 1, Charles Griffin and Co., London.

[6] Rao, C.R. (1974) : 'Linear Statistical Inference and its Applications', Willey Estern Pvt. Ltd., New Delhi.

[7] 'Student' (1907) : 'On the error of counting with haemaycytometer'. Biometrika, 5, 351-360.

[8] Williams, C.B. (1944) : Some applications of the logarithmic series and the index of diversity to ecological problems. Jr. of Ecology, 32, 1—44.

[9] Williams, C.B. (1947) : 'The logarithmic series and its applications to biological problems. Jr. of Econogy, 34, 253-272.