Jour. Ind. Soc. Ag. Statistics
Vol. XXXVIII, No. 3 (1986), pp. 345-351

THE GENERALIZED LOGARITHMIC SERIES DISTRIBUTION WITH ZEROES

P. N. JANI
Sardar Patel University, Vallabh Vidyanagar

(Received : December, 1979)

SUMMARY

The generalized logarithmic series distribution with zeroes is defined and for estimating the parameters the method of maximum likelihood and the method using the zero-cell frequency and the first two sample moments are discussed. The asymptotic variances and covariances of the estimators obtained by both the methods and the asymptotic efficiency of the later method relative to the former are derived. To illustrate the practical application an example of fitting has been considered.

Keywords : Likelihood function; Moments; asymptotic variance and covariance; Efficiency.

Introduction

The generalized logarithmic series distribution (GLSD) is the recent introduction in discrete distributions (Jain and Gupta [2], Jani [3]). Though the GLSD provides a better fit than the usual logarithmic series distribution (Jani and Shah, [4]), it can be used only for zero-truncated data. To fit complete data with zeroes we define in the following the generalized logarithmic series distribution with zeroes :

For a random variable X, the GLSD with zeroes is defined by the probability function

$$P(X=i)=p_i=\begin{cases}(1-\alpha), & \text{for } i=0\\[4pt] \dfrac{\Gamma(i\,\beta)}{i!\,\Gamma(i\beta-i+1)}\,\alpha\,\phi\,\theta^{i}(1-\theta)^{i\beta-i}, & \text{for } i>0,\end{cases} \tag{1.1}$$

where $\phi = [-\log_e(1-\theta)]^{-1},\ 0 < \alpha < 1,\ 0 < \theta < 1,\ \beta\theta < 1$

and $\beta \geqslant 1$ For $\alpha = 1$, (1.1) reduces to the usual GLSD. If the GLSD provides a good fit to a zero-truncated data set then, the GLSD with zeroes must provide a satisfactory fit to the complete data set with zeroes.

Since the rth moment μ_r', about the origin, of the GLSD with zeroes is

$$\mu' = \alpha \text{ (the } r\text{th moment, about the origin, of the GLSD)}, \tag{1.2}$$

we have the first four moments as follows :

$$\mu_1' = \alpha\phi\,\theta\delta^{-1} \tag{1.3}$$

$$\mu_2' = \alpha\phi\theta(1-\theta)\,\delta^{-3} \tag{1.4}$$

$$\mu_3' = \alpha\phi\theta(1-\theta)\,\delta^{-5}[3(1-\theta) - \delta(2-\theta)] \tag{1.5}$$

$$\mu_4' = \alpha\phi\theta(1-\theta)\,\delta^{-7}[15(1-\theta)^2 - 10\,\delta(1-\theta)(2-\theta) + \delta^2(6 - 6\theta + \theta^2)] \tag{1.6}$$

where

$$\delta = 1 - \beta\theta. \tag{1.7}$$

2. Estimation of the Parameters by the Method of Maximum Likelihood

Consider a random sample of size N from the population (1.1) and let N_i be the observed frequency corresponding to $X = i$. Then taking the natural logarithm of the likelihood function $L = \prod_{i=0}^{\infty} p_i^{N_i}$ and differentiating w.r.t. θ, α and β we obtain the maximum likelihood equations for estimating the parameters θ, α and β of (1.1) as

$$\hat{\alpha} = N^{-1}(N - N_0) \tag{2.1}$$

$$\hat{\alpha}\,\hat{\phi}\,\hat{\theta}\,(1 - \hat{\beta}\,\hat{\theta})^{-1} = \overline{X} \tag{2.2}$$

$$\hat{\phi} \sum_{i=2}^{\infty} \sum_{j=1}^{i-1} i \cdot N_i\,(i\,\hat{\beta} - j)^{-1} = N\overline{X}, \tag{2.3}$$

where $N = \sum_{i=0}^{\infty} N_i$ and $\overline{X} = N^{-1} \cdot \sum_{i=0}^{\infty} i \cdot N_i$.

The equations (2.2) and (2.3) can be solved for the maximum likelihood estimators (m.l.e.'s) $\hat{\theta}$ and $\hat{\beta}$ by using the method of iterations or the Newton-Raphson method, although both methods may fail because of non-convergence.

The Fisher information matrix U of the m.l.e.'s $(\hat{\theta}, \hat{\alpha}, \hat{\beta})$ can be found to be

$$U = N[u_{ij}], \tag{2.4}$$

where the elements u_{ij}, $i, j = 1, 2, 3$ are given by

$$\begin{aligned}
u_{11} &= \alpha\, \phi(\theta\, \delta)^{-1} (1 - \theta - \phi\theta\, \delta)\, (1 - \theta)^{-2} \\
u_{22} &= \alpha^{-1}(1 - \alpha)^{-1} \\
u_{33} &= \sum_{i=2}^{\infty} \sum_{j=1}^{i-1} i^2\, p_i(i\beta - j)^{-2} \\
u_{12} &= u_{21} = 0 \\
u_{23} &= u_{32} = 0 \\
u_{13} &= u_{31} = \alpha\, \phi\, \delta^{-1}(1 - \theta)^{-1}.
\end{aligned} \tag{2.5}$$

Hence, the asymptotic variances and covariances of the m.l.e.'s can easily by obtained as

$$\begin{aligned}
V(\hat{\theta}) &= N^{-1}\theta\, \delta^2(1 - \theta)^3\, (\alpha\phi)^{-1}\, u_{33}Q \\
V(\hat{\alpha}) &= N^{-1}\, \alpha(1 - \alpha) \\
V(\hat{\beta}) &= N^{-1}\, (1 - \theta - \phi\theta\delta)\, \delta Q \\
\mathrm{Cov}\, (\theta, \hat{\beta}) &= -N^{-1}\, \delta\, \theta^2(1 - \theta)\, Q \\
\mathrm{Cov}\, (\hat{\alpha}, \hat{\theta}) &= \mathrm{Cov}\, (\hat{\alpha}, \hat{\beta}) = 0,
\end{aligned} \tag{2.6}$$

where

$$Q = [\delta(1 - \theta - \phi\theta\delta)\, u_{33} - \alpha\phi\theta^3]^{-1} \tag{2.7}$$

3. Method Using the Zero-Cell Frequency and the First Two Sample Moments

As we have seen in Section 2, the method of maximum likelihood involves the equation for estimating the parameters in double series form which is not solved explicitly for parameters. It has been observed that in most of the cases, in practice, the convergence is not achieved. Hence, we seek for another, but efficient, method for estimating the parameters. In the following we discuss a method which makes use of the zero-cell frequency and the first two sample moments.

Using (1.1), (1.4) and (1.5) and replacing p_0, μ_1' and μ_2' by their respective sample estimators $N^{-1}N_0$, $\bar{X}$ and $M_2' = N^{-1} \sum_{i=0}^{\infty} i^2 \cdot p_i$ we get the following equations for estimators $\tilde{\theta}$, $\tilde{\alpha}$, and $\tilde{\beta}$:

$$\tilde{\alpha} = N^{-1}(N - N_0) \tag{3.1}$$

$$\tilde{\theta} + C(\tilde{\phi}\ \tilde{\theta})^2 - 1 = 0 \tag{3.2}$$

$$\tilde{\beta} = (\bar{X}\tilde{\theta})^{-1} (\bar{X} - \tilde{\alpha}\ \tilde{\phi}\ \tilde{\theta}) \tag{3.3}$$

where

$$C = N^{-2}(N - N_0)^2 M_2' \bar{X}^{-3} \tag{3.4}$$

To obtain $\tilde{\theta}$ from (3.2), find two values of θ for which the left hand side of (3.2) gives negative and positive values. Then by linear interpolation we get $\tilde{\theta}$. The table of the values of $(\phi\theta)$, computed for $\theta = 0.01\ (0.01)\ 0.99$ correct upto six decimal places shown in Table 2 in the Appendix, will be found useful for this purpose.

Using the differential-method (Kendall and Stuart [5]) we obtain the asymptotic variance-covariance matrix W of the estimators $\tilde{\theta}$, $\tilde{\alpha}$ and $\tilde{\beta}$, to the order N^{-1}, as

$$W = N^{-1} [w_{ij}], \tag{3.5}$$

where the elements w_{ij}, $i, j = 1, 2, 3$ are given by

$$\begin{aligned} w_{11} &= N \cdot V(\tilde{\theta}) = \theta(1-\theta)(\alpha\phi\delta)^{-1}(2-\theta-2\phi\theta)^{-2}[6(1-\theta)^2 \\ &\qquad - 4\delta(1-\theta)(2-\theta) + \delta^2(6 - 6\theta + \theta^2) - 4\theta\phi\delta(1-\theta)] \\ w_{22} &= N \cdot V(\tilde{\alpha}) = \alpha(1-\alpha) \\ w_{12} &= N \cdot \mathrm{Cov}(\tilde{\theta}, \tilde{\alpha}) = 0 \\ w_{23} &= N \cdot \mathrm{Cov}(\tilde{\alpha}, \tilde{\beta}) = 0 \\ w_{13} &= N \cdot \mathrm{Cov}(\tilde{\theta}, \tilde{\beta}) = A \cdot w_{11} + \delta(1-\theta)(\alpha\phi\theta)^{-1} \\ w_{33} &= N \cdot V(\tilde{\beta}) = A \cdot w_{13} \end{aligned} \tag{3.6}$$

where

$$A = (\delta\phi\theta + \theta - 1)(1-\theta)^{-1}\theta^{-2}. \tag{3.7}$$

The joint asymptotic efficiency E of the estimators $\tilde{\theta}$, $\tilde{\alpha}$ and $\tilde{\beta}$ relative to the m.l.e.'s $(\hat{\theta}, \hat{\alpha}, \hat{\beta})$ is given by

$$E = (\,|U| \cdot |W|\,)^{-1} \tag{3.8}$$

where $|X|$ is the determinant of a matrix X. Upon simplification of (3.8) we find that E is independent of α and it is the same as the efficiency of the moment estimators, relative to the maximum likelihood, in the case of the GLSD. These efficiencies have been computed for different values of the parameters θ and β and given in Jani and Shah [4].

4. Example

The examples given in Jani and Shah [4] were the zero-truncated data fitted by the GLSD. These examples with complete data with zeroes are well fitted by the GLSD with zeroes. However, we consider here Garman's complete data on counts of the number of European red mites on apple leaves (Bliss, [1]). Table 1 shows the fit by the GLSD with zeroes together

TABLE 1—P. GARMAN'S DATA SHOWING COUNTS OF THE NUMBER OF EUROPEAN RED MITES ON APPLE LEAVES (BLISS, [1])

No. of mites per leaf i	*Leaves observed* N_i	*Expected frequency*		
		NB	*GNB*	*GLSD with zeroes*
0	70	67.49	71.48	70.00
1	38	39.03	33.98	39.04
2	17	20.96	19.80	17.42
3	10	10.97	11.59	9.76
4	9	5.66	6.57	5.85
5	3	2.90	3.55	3.58
6	2	1.48	1.80	2.17
7	1	0.75	0.84	1.28
$\geqslant 8$	0	0.76	0.39	0.90
Total	150	150.00	150.00	150.00
Mean	1.14667			
Variance	2.27365			
χ^2		1.99	1.38	0.15
d.f. (ν)		2	1	1
$P(\chi^2_\nu)$		0.38	0.25	0.70

Estimates : $\tilde{\theta} = 0.89115$, $\tilde{\alpha} = 0.53333$, $\tilde{\beta} = 0.91243$.

with the fits by the negative binomial (NB) and the generalized negative binomial (GNB) distributions. The method of Section 3 is used in fitting the GLSD with zeroes.

REFERENCES

[1] Bliss, C. I. (1953) : Fitting the negative binomial distribution to biological data, *Biometrics*, **9** : 176-196.

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

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

[4] Jani, P. N. and Shah, S. M. (1979) : On fitting of the generalized logarithmic series distribution. To appear in the *Journal of the Indian Society of Agricultural Statistics* XXXI (3).

[5] Kendall, M. G. and Stuart, A. (1969) : *The Advanced Theory of Statistics*, Vol. I, Charles Griffin and Co., London.

APPENDIX

TABLE 2—THE VALUES OF $\phi\theta$, WHERE $\phi = [-\log_e (1-\theta)]^{-1}$

θ	$\phi\theta$	θ	$\phi\theta$	θ	$\phi\theta$	θ	$\phi\theta$
0.01	0.995025	0.26	0.863486	0.51	0.714937	0.76	0.532543
2	89952	7	57930	2	08477	7	23925
3	84933	8	52349	3	01965	8	15148
4	79864	9	46740	4	0.695402	9	06200
5	74792	0.30	41102	5	88785	0.80	0.497068
6	69697	1	35435	6	82111	1	87737
7	64573	2	29742	7	75380	2	78190
8	59440	3	24015	8	68587	3	68409
9	54290	4	18262	9	61733	4	58370
0.10	49118	5	12474	0.60	54814	5	48048
1	43931	6	06656	1	47827	6	37411
2	38725	7	00805	2	40771	7	26424
3	33492	8	0.794919	3	33642	8	15043
4	28240	9	89001	4	26437	9	07743
5	22969	0.40	83046	5	19153	0.90	0.390865
6	17679	1	77055	6	11785	1	77916
7	12360	2	71028	7	04332	2	64251
8	07025	3	64963	8	0.596788	3	49721
9	01666	4	58859	9	89148	4	34114
0.20	0.896282	5	52714	0.70	81408	5	17118
1	90880	6	46528	1	73564	6	0.298241
2	85451	7	40300	2	65608	7	76625
3	79995	8	34028	3	57536	8	50510
4	74518	9	27710	4	49339	9	14976
5	69015	0.50	21348	5	41011		