MAXIMUM LIKELIHOOD ESTIMATES FOR THE PARAMETERS OF AN INFLATED POISSON-LINDLEY DISTRIBUTION

ARUN JHAJHARIA, PRATIBHA CHOUDHARY

Abstract : In the present paper, maximum likelihood estimates for the parameters of an inflated Poisson-Lindley distribution have been derived. The elements of the information matrix are given for the determination of asymptotic variances and covariance of the estimates. The suitability of the distribution is tested using the data of number of insects per leaf (Beall, 1940) and number of accidents per woman (Shankaran, 1970). The fit is also compared with inflated Poisson distribution.

Keywords : Inflated Poisson-Lindley distribution, Maximum likelihood, Asymptotic variance-covariance, Parameter.

Introduction : In recent years, the applied distributions involving probability have attracted increasing attention of researchers, scientists who are associated in the experiments with count data having excess of zeroes. For example to improve electronics manufacturing quality, in medical research of HIV patients with high risk behaviours and in agricultural study of the number of aphids on mustard and safflower crops, the number of leaf affected due to virus of potato, the number of insects per leaf, the number of larvae on castor and paddy crops, the number of little leaf of brinjal due to bacteria etc.

As a matter of fact, probability distributions are a type of abstract of simplified representation of the essentially important aspects of the real phenomena. A major motivating force was the empirical observations that many distributions obtained in the course of experimental investigation often had an excess of zeroes as compared with a Poisson distribution with the same mean. This phenomenon is to be expected when some kind of clustering is present, and, indeed, many of the distributions possess the property that the proportion in the zero class is greater than exp [-(expected value)], which is the value which would be predicted on the basis of a Poisson distribution. The simplest way of increasing the proportion of zeroes is just to add an arbitrary proportion of zeroes decreasing the remaining proportions in an appropriate constant ratio (Johnson, Kotz and Kemp 1992).

The aim of the present paper is to estimate the parameters of an inflated Poisson-Lindley distribution by maximum likelihood method. For completeness, Section 2 deals with inflated Poisson-Lindley distribution. Section 3 devotes the estimation procedure. Illustrative examples are included at the end to investigate the suitability of the distribution.

Inflated Poisson-Lindley Distribution : The derivation of Poison-Lindley distribution is given below, which is based on two strong assumptions:

The objects which are counted occur in groups, the number of groups follows a Poisson distribution with parameter λ i.e.,

$$P\,[Y=k] = \frac{e^{-\lambda}\lambda^{k}}{k\,!} \qquad \lambda>0$$

k=0, 1, 2, 3...

The number of individuals per group , has the continuous Lindley-distribution with parameter θ,

$$g_x\,[\lambda,\theta] = e^{-\lambda\theta}(\lambda+1)\beta(\theta)$$

$$\text{where } \beta(\theta) = \frac{\theta^2}{\theta+1} \qquad \lambda,\theta>0$$

Using the method of compounding, the distribution becomes compound Poisson-Lindley distribution.

$$P_X(X=k) = \int_0^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} e^{-\lambda\theta}(\lambda+1)\beta(\theta)d\lambda$$

$$= \frac{\beta(\theta)}{\beta(\theta-1)}\frac{\mu_k'(\theta-1)}{k!}$$

k = 0, 1, 2, 3...

where $\mu_k'(\theta-1)$ is the raw moment of order k of Lindley distribution with $(\theta-1)$ as the parameter. After simplification the distribution becomes:

$$P_k(\theta) = \frac{\theta^2(\theta+2+k)}{(\theta+1)^{k+3}} \qquad k=0, 1, 2, 3...$$

Let X is the random variable denoting the number of insects /leaf and the number of accidents per woman and ω be the proportion of leaves and women which are exposed to the risk of bearing insects & accidents and (l –ω) not exposed to the risk, then the inflated Poisson-Lindley distribution is as follows

$$P[X=0] = 1-\omega+\omega\frac{\theta^2(\theta+2)}{(\theta+1)^3} \qquad \text{for } k=0$$

P[x=k] = $\omega \dfrac{\theta^2(\theta+2+k)}{(\theta+1)^{k+3}}$ for k > o ...(2)

Estimation Of Parameters By Maximum Likelihood Method : Let us consider a sample consisting of N observations of the random variable X with probability mass function given by eq. (1) & (2). The likelihood function can be written as;

$$L \cong \left(1 - w + w\theta^2 \frac{(\theta+2)}{(\theta+1)^3}\right)^{n_0}$$

$$\prod_{k=1}^{R}\left(\frac{w\theta^2(\theta+2+k)}{(\theta+1)^{K+3}}\right)^{n_k} \quad ...(3)$$

Where n_k is the sample frequency of k, n is the number of non-zero sample observations n=N- n_0 and $\prod$ denotes the products over n non-zero observations. R is the largest numbers of observations. Taking logarithm of L, differentiating with respect to ω and θ in turn and setting the derivatives equal to zero gives the estimating equations.

We have,

$$LogL = n_0 Log\left(1 - \omega + \omega\frac{\theta^2(\theta+2)}{(\theta+1)^3}\right)$$

$$+\sum_{k+1}^{R} n_k Log\left(\omega\frac{\theta^2(\theta+2+k)}{(\theta+1)^{k+3}}\right)$$

$$LogL = n_0 Log\left(\frac{(\theta+1)^3 - \omega(1+3\theta+\theta^2)}{(\theta+1)^3}\right) + \sum_{k+1}^{R} n_k Log\omega$$

$$+\sum_{K=1}^{R} 2n_k Log\theta + \sum n_k Log(\theta+2+k) - \sum_{k=1}^{R}(k+3)n_k Log(\theta+1)$$

$$\frac{\partial \log L}{\partial\theta} = \frac{n_0}{(\theta+1)^3 - \omega(1+3\theta+\theta^2)} \times \frac{\omega\theta(\theta+4)}{(\theta+1)}$$

$$+\sum_{k=1}^{R}\frac{2n_k}{\theta} + \sum_{k=1}^{R}\frac{n_k}{(\theta+2+k)} - \sum_{k=1}^{R}\frac{(k+3)n_k}{(\theta+1)} = 0$$

$$\frac{\partial LogL}{\partial\omega} = \frac{n_0\left[(\theta+1)^3 \times -(1+3\theta+\theta^2)\right]}{(\theta+1)^3 - \omega(1+3\theta+\theta^2)(\theta+1)^3} + \sum_{k+1}^{R}\frac{n_k}{\omega} = 0$$

$$\omega = \frac{N - n_0(\theta+1)^3}{N(1+3\theta+\theta^2)}$$

$$\hat{\omega} = \frac{N-n_0}{N}\cdot\frac{(\theta+1)^3}{(1+3\theta+\theta^2)}$$

Maximum likelihood estimates, when they exist may be determined by solving the system of equations (3.1) and (3.2) by employing a trial and error technique by which convergence is greatly accelerated (Sampford 1955).

Let us assume a specified value of ω. From equation (3.2), θ can be obtained. We can check whether the equation (3.3) converges to zero or not, if it does not converge into zero, then a slight change in the original value of ω may be taken and repeated all the steps accordingly. This trial and error process is repeated until, a true set of estimation of the parameters are not obtained. The asymptotic variance-covariance matrix is obtained by inverting the information matrix whose elements are negatives of expected values of the second order derivatives of logarithms of likelihood function. These derivatives are given below:

$$\frac{\partial^2 \log L}{\partial\omega^2} = -\frac{n_0(1+3\theta+\theta^2)^2}{\left[(\theta+1)^3 - \omega(1+3\theta+\theta^2)\right]^2} - \frac{N-n_0}{\omega^2}$$

$$\frac{\partial LogL}{\partial\theta^2} = n_0\omega\frac{\left[(2\theta+4)\left\{(\theta+1)^4 - \omega(1+3\theta+\theta^2)(\theta+1)\right\} - \theta(\theta+4)\left\{4(\theta+1)^3 - \omega(3\theta^2+8\theta+4)\right\}\right]}{\left[(\theta+1)^4 - \omega(1+3\theta+\theta^2)(\theta+1)^2\right]}$$

$$... - \sum_{k=1}^{R}\frac{2n_k}{\theta^2} - \sum_{k=1}^{R}\frac{n_k}{(\theta+2+k)^2} - \sum_{k=1}^{R}\frac{(k+3)n_k}{(\theta+1)^2} = 0$$

$$\frac{\partial^2 \log L}{\partial\omega\partial\theta} = \frac{\partial^2 \log L}{\partial\theta\partial\omega} = \frac{\left[n_0(3+2\theta)\right]\left\{(\theta+1)^3 - \omega(1+3\theta+\theta^2)^2\right\} - n_0(1+3\theta+\theta^2)\left[3(\theta+1)^2 - \omega(3+2\theta)\right]}{\left[(\theta+1)^3 - \omega(1+3\theta+\theta^2)\right]^2}$$

$$E\left[-\frac{\partial^2 \log L}{\partial\omega^2}\right] = \frac{E(n_0)(1+3\theta+\theta^2)^2}{\left[(\theta+1)^3 - \omega(1+3\theta+\theta^2)\right]^2} + \frac{E(N-n_0)}{\omega^2}$$

$$E\left(\frac{-\partial^2 \log L}{\partial \theta^2}\right) = E\left(n_0 \omega\right)$$

$$\frac{\left[(2\theta+4)\left\{(\theta+1)^4 - \omega\left(1+3\theta+\theta^2\right)(\theta+1)\right\} - 4\theta(\theta+4)(\theta+1)^3 + \omega\theta(\theta+4)\left(3\theta^2+8\theta+4\right)\right]}{\left[(\theta+1)^4 - \omega\left(1+3\theta+\theta^2\right)(\theta+1)\right]^2}$$

$$\ldots\ldots\ldots\ldots \quad \ldots\ldots\ldots\ldots \quad \ldots - 2\sum_{k=1}^{R}\frac{n_k}{\theta^2} - \sum_{k=1}^{R}\frac{n_k}{(\theta+2+k)^2} - \sum_{k=1}^{R}\frac{(k+3)n_k}{(\theta+1)^2}$$

$$E\left(\frac{-\partial^2 \log L}{\partial\omega\partial\theta}\right) = E\left(\frac{-\partial^2 \log L}{\partial\theta\partial\omega}\right) = -E(n_0)\frac{\left[\theta(\theta+4)(\theta+1)^4(\theta+1)^2\right]}{(\theta+1)^3\left[(\theta+1)^3 - \omega\left(1+3\theta+\theta^2\right)\right]^2}$$

$$E\left[-\frac{\partial^2 \log L}{\partial\omega\partial\theta}\right] = E\left[-\frac{\partial^2 \log L}{\partial\theta\partial\omega}\right] = -\frac{E(n_0)(\theta+1)^2\theta[\theta+4]}{\left[(\theta+1)^3 - \omega\left(1+3\theta+\theta^2\right)\right]}$$

Thus, we can have the variance–covariance matrix of the parameter ω and θ as follows

$$v\begin{pmatrix}\hat{\omega}\\ \hat{\theta}\end{pmatrix} = \begin{bmatrix} v(\hat{\omega}) & \operatorname{cov}(\hat{\omega}\hat{\theta}) \\ \operatorname{cov}(\hat{\theta}\hat{\omega}) & v(\hat{\theta})\end{bmatrix}$$

Illustrative Examples :

The suitability of the Poisson-Lindley distribution is tested using the data of Beall (1940) and Shankaran (1970). Table 4.1 reveals the distribution of observed and expected number of leaves according to number of insects. It is important to $\frac{w(\theta+2)}{\theta(\theta+1)}$ note that gives the average number of insects per leaf. Once the estimates of ω and θ are obtained, expected frequencies can easily be computed. The estimated values of ω and θ are found to be 0.85 and 1.55 respectively which indicates that 15% of the leaves are not exposed to the risk of attracting insects on them.

Table: 4.1 Distribution of observed and expected number of leaves according to the number of insects (Beall, 1940)

No. of Insects	Observed Frequency	Expected frequency: Poisson Lindley	Expected frequency: Inflated Poisson
0	33	32.89	32.86
1	12	12.31	11.01
2	6	5.89	7.36
3	3	4.91	4.77
4	1		
5	1		
Total	56	56.00	56.00

Parameter estimates $\hat{\theta}$=1.55 $\hat{\lambda}$=1.34

$\hat{\omega}$ = 0.85 $\hat{\omega}$ = 0.56

$\chi^2 = 0.01$ $\chi^2 = 0.35$

d.f.= 1 d.f.=1

For applying a chi-square test, some last cells are grouped together. The value of chi-square does not seem to be significant in the case of Poisson-Lindley and Poisson distribution. But the value of chi-square in the case of inflated Poisson is found to be larger rather than that of inflated Poisson-Lindley. Thus we conclude that inflated Poisson-Lindley distribution provides excellent fitting for the data having excess of zeroes. Hence this distribution is recommended for the kinds of data where we do have the excess of zeroes specially for the number of insects/leaf.

Table 4.2 provides the distribution of observed and expected number of accidents per woman. It is to be noted that $\frac{w(\theta+2)}{\theta(\theta+1)}$ gives the average number of accidents. Once the estimates of ω and θ are obtained, expected frequencies can easily be computed. The estimated values of ω and θ are found to be 0.88 and 2.45 respectively which indicates that 12% of the women are not exposed to the risk of accidents .

Table: 4.2 Distribution of observed and expected number of accidents per woman (Shankaran, 1970)

No. of accidents	Observed Frequency	Expected frequency	
		Poisson-Lindley	Inflated Poisson
0	447	446.68	447.00
1	132	133.39	124.70
2	42	45.04	54.90
3	21	14.85	16.11
4	3	7.04	4.29
5	2		
Total	647	647.00	647.00

$\hat{\theta}$ =2.45 $\hat{\lambda}=$ .8805

Parameter estimates $\hat{\omega}$ = 0.88 $\hat{\omega}$ = .528

$\chi^2 = 3.358$ $\chi^2 = 5.06$

d.f.= 2 d.f.=2

For applying a chi-square test, some last cells are grouped together. The value of chi-square does not occur to be significant in the case of Poisson-Lindley and Poisson distribution. But the value of chi-square in the case of inflated Poisson-Lindley is found to be smaller rather than that of inflated Poisson. Thus we conclude that inflated Poisson-Lindley distribution provides excellent fitting for the data having excess of zeroes. Hence this distribution is recommended for the types of data where the excess of zeroes specially for the number of accidents per woman is happened.

References :

1. Beall, G. The Fit and Significance of Contagious Distributions When Applied to Observations on Larval Insects. Ecology, 21, 1940, pp. 460-474.
2. Johnson ,N.L. Kotz, S. and Kemp, A.W., Univariate Discrete Distributions: The Distributions in statistics, 2^{nd} Edn, John Wiley & Sons Inc. 1992.
3. Sampford, M.R., The Truncated Negative-binomial Distribution, Biometrika, 42, 1955. pp 58-69.
4. Sankaran,M., The discrete Poisson Lindley distribution, Biometrics, 26, 1970. pp 145-146

* * *

D-1-214, Vallabh Garden, Rajmata Sudarsana Nagar, Bikaner- 334001, Rajasthan
Department of Farm Engineering, Institute of Agricultural Sciences,Banaras Hindu University, Varanasi221005
Email: arunstatistics@gmail.com, Mobile No. 08765120512

New Housing Board Colony M/100, Prem Nagar, Balaghat- 481001 (M.P.)
Department of Genetics and Plant Breeding, College of Agriculture, Rewa Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur- (M. P.), India
Email: prati.gudiya1989@gmail.com , Mobile No. 07898880326