Estimation of Parameters of Inflated Poisson-Lindley Distribution for Adult Out-migration

Brijesh P. Singh, Niraj K. Singh and Shweta Dixit

Faculty of Commerce & DST-Centre for Interdisciplinary Mathematical Sciences
Banaras Hindu University, Varanasi, India
Email- :brijesh@bhu.ac.in

ABSTRACT

The aim of this paper is to study the distribution of adult migrants from the household. The inflated Poisson-Lindley distribution is found to be a suitable model to describe this distribution. The parameters of this distribution have been estimated by two methods and suitability of model is checked by using real data sets. Findings may help planners and policy makers for designing more effective and reasonable policies.

Keywords: Migration, Poisson-Lindley distribution, Inflation of probability distribution.

INTRODUCTION

In India, the caste system controls the various demographic behavior of the person and also an important aspect of social stratification and plays an important role in decision taken for migration. The rigidity of caste system and attachment to caste tend to slow down spatial mobility but, at the same time, it pushes the process of migration through the presence of social basis of economic inequality and conflict also. Several attempts have been made to study the pattern of rural out-migration through the use of probability models (Iwunor 1995, Singh *et al.* 1980, 1985, Sharma 1987, 1985, Singh & Yadava 1981, Yadava & Singh 1983). In 1981, Singh and Yadava (1981) proposed negative binomial distribution to describe the trend of rural out-migration for male adult migrants aged 15 years and above. The idea of cluster was incorporated by Yadava and Singh (1983) and found that Thomas distribution is well suited to describe the number of adult migrants from a household. Yadava and Yadava (1988) extended the idea of cluster and assumed that the occurrence of migration in cluster varies from household to household and the number of adult migrants to a cluster follows truncated displaced geometric distribution. A probability distribution under such assumptions fitted well the distribution of male adult migrants aged 15 years and above.

Since probability models provide concise and clear representations of extensive data sets in a better way. Singh and Yadava (1981) have introduced the negative binomial distribution to study the pattern of rural-out migration at household level. Sharma (1985) used the method of moments to estimate the parameters of

inflated geometric distribution and obtained the asymptotic expressions for variance and covariance of the estimators using multivariate central limit theorem. Iwunor (1995) proposed an alternative estimation technique based on likelihood function and obtained the variance and covariance of the estimators. Though he used the likelihood function using multinomial combination, but finally estimated the parameters by mean-zero frequency method. Hossain (2000) and Aryal (2011) used maximum likelihood method to estimate the parameters of the model considered.

Some mixture of Poisson Lindley distributions were studied by Borah and Deka (2001), where certain properties of Poisson-Poisson-Lindley, Poisson-Lindley-Poisson distributions and inflated Poisson Lindley distribution were investigated. The discrete inflated distribution was first investigated by Singh (1963). He studied Inflated Poisson distribution to serve the probabilistic description of an experiment with a slight inflation at a point, say zero. Gerstenkorn (1979) established the recurrence relation for the moments for the Inflated negative binomial, Poisson and Geometric distribution. Unlike other past attempts in the field of modeling the pattern of migration, Singh *et al.* (1985) introduced the idea of including the inflation parameter in modeling trends in rural out-migration at the micro level. He applied the inflated geometric distribution as well as the inflated generalized Poisson distribution.

In this paper, first tried to model, migration pattern by Poisson-Lindley but this distribution is fail to explain the pattern of adult migration then an attempt has been

made to study Inflated Poisson-Lindley (IPL) distribution which serve the probabilistic description of an experiment with a slight inflation of probability at zero. For fitting of the IPL distribution, four data sets related to distribution of adult migrants from the households are considered for an empirical comparison and it is observed that this distribution gives better fit in all the cases. Most of the studies mentioned above have used moment technique or mean-zero frequency method (equating observed and theoretical zero'th cell frequencies and means) to estimate the parameters involved in the models. Further, moment estimates are usually consistent, but they are often less efficient. Considering these limitations into account, the maximum likelihood estimation technique can be used but this procedure involve a lot of computational complexity and required programming however the estimate of parameters can be easily obtained with the moment method.

CONSTRUCTION OF MODEL

When we think about the number of adult migrants from household in a society, some households have varying number of adult migrants and some household have no adult migrants. At the same time we can also think that some household have intention of migration but reported no migration. Thus number of household with no migration becomes inflated. Keeping this fact into consideration under some assumptions an attempt has been made to develop a probability model for the number of rural male out- migrants aged 15 years and above from a household:

(i) At any point of time, let α be the probability of household in which adult migration is present and $(1-\alpha)$ be the probability of household having no migration.

(ii) If y represents the number of adult migrants from a household and follows Poisson distribution with parameter λ and it is assumed that this parameter varies according to Lindley distribution because there are huge disparity in terms of social standard of the household present in the society which affects the intensity of migration. Thus finally we may model the number of adult migrants which is a random variable and follows as Inflated Poisson-Lindley distribution. The Poisson-Lindley distribution is given by;

$$P(X = x) = \frac{\theta^2 (x + \theta + 2)}{(\theta + 1)^{x+3}}; x = 0, 1, 2 \ldots \& \theta > 0 \tag{1}$$

Thus Inflated Poisson-Lindley distribution will be;

$$P(Y = 0) = 1 - \alpha + \alpha \left[\frac{\theta^2 (\theta + 2)}{(\theta + 1)^3} \right]; y = 0 \quad 0 < \alpha < 1, \ \theta > 0$$

$$P(Y = y) = \frac{\alpha \theta^2 (y + \theta + 2)}{(\theta + 1)^{(y+3)}} \ ; \ y = 1, 2, 3, \ldots \tag{2}$$

ESTIMATION

The method of moment estimation procedure and zero cell frequency method have been considered to estimate the parameters involved in the models. In fact these estimation procedures are easier and quicker than other estimation procedure and also there is no need of soft computation in the estimation procedure. IPL distribution has two parameters α and θ to be estimated.

Estimation using Zero Cell frequency:

We know that,

$$\mu_1' = \alpha \frac{(\theta + 2)}{\theta(\theta + 1)} = m \ (say) \tag{3}$$

$$p(0) = 1 - \alpha + \alpha \frac{\theta^2 (\theta + 2)}{(\theta + 1)^3} \tag{4}$$

From (3)

$$\alpha = \frac{m\theta(\theta + 1)}{(\theta + 2)}$$

Putting this into (4)

$$p(0) = 1 - \frac{m\theta(\theta + 1)}{(\theta + 2)} + \frac{m\theta(\theta + 1)}{(\theta + 2)} \frac{\theta^2 (\theta + 2)}{(\theta + 3)^3}$$

$$= 1 - \frac{m\theta(\theta + 1)}{(\theta + 2)} + \frac{m\theta^3}{(\theta + 1)^2}$$

$$= \frac{(\theta + 2)(\theta + 1)^2 - m\theta(\theta + 1)^3 + m\theta^3(\theta + 2)}{(\theta + 2)(\theta + 1)^2}$$

$$= \frac{(\theta + 2)(\theta^2 + 2\theta + 1) - m\theta(\theta^3 + 3\theta^2 + 3\theta + 1) + m\theta^4 + 2m\theta^3}{(\theta + 2)(\theta^2 + 2\theta + 1)}$$

$$= \frac{(\theta^3 + 2\theta^2 + \theta + 2\theta^2 + 4\theta + 2) - (m\theta^4 + 3m\theta^3 + 3m\theta^2 + m\theta) + m\theta^4 + 2m\theta^3}{(\theta^3 + 2\theta^2 + \theta + 2\theta^2 + 4\theta + 2)}$$

$$= \frac{(\theta^3 + 4\theta^2 + 5\theta + 2) - m\theta(\theta^2 + 3\theta + 1)}{(\theta^3 + 4\theta^2 + 5\theta + 2)}$$

$$= 1 - \frac{m\theta(\theta^2 + 3\theta + 1)}{(\theta^3 + 4\theta^2 + 5\theta + 2)}$$

$$\Rightarrow \frac{m\theta(\theta^2 + 3\theta + 1)}{(\theta^3 + 4\theta^2 + 5\theta + 2)} = 1 - p(0)$$

$$\Rightarrow \frac{(\theta^3 + 3\theta^2 + \theta)}{(\theta^3 + 4\theta^2 + 5\theta + 2)} = \frac{1 - p(0)}{m} = k$$

$$\Rightarrow k\theta^3 + 4k\theta^2 + 5k\theta + 2k - \theta^3 - 3\theta^2 - \theta = 0$$

$$\Rightarrow \theta^3(k - 1) + \theta^2(4k - 3) + \theta(5k - 1) + 2k = 0$$

$$\Rightarrow \theta^3(k - 1) + \theta^2(4k - 3) + \theta(5k - 1) + 2k = 0 \tag{5}$$

Solving it we get the estimates of θ and α by using equation (3).

Estimation using first two moments:

We know that mean and variance of Poisson-Lindley distribution is;

$$mean = \mu_1' = \frac{(\theta+2)}{\theta(\theta+1)}$$

$$Variance = \mu_2' - \mu_1'^2 = \frac{(\theta+2)}{\theta(\theta+1)} + \frac{2(\theta+3)}{\theta^2(\theta+1)} - \left(\frac{(\theta+2)}{\theta(\theta+1)}\right)^2$$

Let x is the Poisson-Lindley distribution and y is inflated Poisson-Lindley distribution then we know that;

$E(y) = \alpha E(x)$ and $E(y^2) = \alpha E(x^2)$.

Thus $Var(y) = \alpha Var(x) + \alpha(1-\alpha)[E(x)]^2$

Therefore

$$E(y) = \frac{\alpha(\theta+2)}{\theta(\theta+1)} = m \ (say) \tag{6}$$

$$Var(y) = \alpha\left[\left\{\frac{(\theta+2)}{\theta(\theta+1)} + \frac{2(\theta+3)}{\theta^2(\theta+1)}\right\} - \left\{\frac{(\theta+2)}{\theta(\theta+1)}\right\}^2\right] - \alpha(1-\alpha)\left[\frac{(\theta+2)}{\theta(\theta+1)}\right]^2 \tag{7}$$

$$= \alpha\left[\frac{(\theta+2)}{\theta(\theta+1)}\right] + \frac{2\alpha(\theta+3)}{\theta^2(\theta+1)} - \alpha\left[\frac{(\theta+2)}{\theta(\theta+1)}\right]^2 + \alpha\left[\frac{(\theta+2)}{\theta(\theta+1)}\right]^2 - \alpha^2\left[\frac{(\theta+2)}{\theta(\theta+1)}\right]^2$$

$$= m + \frac{2\alpha(\theta+3)}{\theta^2(\theta+1)} - m^2$$

$$\Rightarrow Var - m + m^2 = \frac{2\alpha(\theta+3)}{\theta^2(\theta+1)} = k$$

$$\Rightarrow k\theta^2(\theta+1) = 2\alpha(\theta+3)$$

$$\Rightarrow \alpha = \frac{k\theta^2(\theta+1)}{2(\theta+3)}$$

Putting the value of α in equation (6), we get;

$$m = \frac{k\theta(\theta+2)}{2(\theta+3)}$$

$$\Rightarrow k\theta(\theta+2) = 2m\theta + 6m$$

$$\Rightarrow k\theta^2 + 2k\theta - 2m\theta - 6m = 0$$

$$\Rightarrow k\theta^2 + 2\theta(k-m) - 6m = 0 \tag{8}$$

Solving it we get the estimates of θ by using above equations.

The second method is easier and quicker to get the estimate since it is a quadratic equation however in the first method was cubic equation.

APPLICATION OF THE MODEL

The models have been applied to the primary data taken from a survey entitled "Migration and Related Characteristics-a Case Study of North-Eastern Bihar" conducted during October 2009 to June 2010. This analysis is based on the information collected from 664 households. Further, the models proposed have also been applied to other three data sets for different time and space to check its suitability. The Varanasi data was collected under a sample survey "Rural development and population growth (RDPG) survey" conducted in 1978 in Varanasi district and used by Sharma (1985) and Iwunor (1995). The Nepal data is taken from a sample survey of the Rupandehi and Palpa districts in Nepal and used by Aryal (2011). The Bangladesh data was collected under a sample survey "Impact of Migration on Fertility in Bangladesh: A study of Comilla district" conducted in 1997 and used by Hossain (2000).

CONCLUSION

Table 1 shows that the proposed model fit the data excellently. Estimate of parameters and the value of χ^2 with degree of freedom are also given in the tables. The value of χ^2 shown in the tables clearly indicate that the models describe the distribution of number of adult migrants from households. Further to check the suitability, when the model has been applied to the data set collected from Varanasi, Nepal (Aryal 2011) and Comilla district of Bangladesh (Hossain 2002), it is given in table 2, 3 and 4 respectively. We observe that on the basis of χ^2 and *p*-value the agreement between observed and expected frequencies for proposed model is fairly good. The value of α that is probability of household having adult migrants is maximum for Comilla (92%) followed by flooded area of Kosi river, Bihar (80%). It is least in the older data of Varanasi (31 %) and in Nepal the probability of household having adult migrants is about 67 percent. The estimate of other parameters is 1.51, 2.43, 2.46 and 3.03 for flooded area of Kosi river, Bihar, Varanasi, Nepal and Comilla, Bangladesh which has no physical interpretation but we can use this estimate to know the variation in the average number of adult migrants in the households.

REFERENCES

Aryal, T.R. 2011. Inflated Geometric Distribution to Study the Distribution of Rural Out-migrants. *Journal of the Institute of Engineering* **8(1)**: 266-268.

Borah, M., Deka, N.A. 2001. A Study on the Inflated Poisson Lindley Distribution. *Jour. lnd. Soc. Ag. Statistics* **54(3):** 317-323.

Borah, M., Deka, N.A. 2001. Poisson Lindley and some of its mixture distributions. *Pure and Applied Mathematika Sciences* **53(1-2)**: 1-8.

Gerstenkom, T. 1979. Moment recurrence relation for the Inflated Negative Binomial, Poisson and Geometric distribution. *Demonstratio Mathematica* **11(2)**: 389-411.

Hossain, M.Z. 2000. Some demographic models and their application with special reference to Bangladesh. unpublished PhD thesis, Banaras Hindu University, India.

Iwunor, C.C.O. 1995. Estimating of parameters of the Inflated Gometric distribution for rural out-migration. *Genus* **51**: 3-4.

Sharma, H.L. 1985. A Probability Distribution for Rural Out-migration at Micro-level. *Rural Demography* **12(1-2):** 63-69.

Sharma, H.L. 1987. A probability distribution for out-migration. *Janasamkhya -A Journal of Demography* **5(2)**: 95-101.

Sankaran, M. 1970. The discrete Poisson-Lindley distribution. *Biometrics* **26(1):** 145-149.

Singh Brijesh, P., Singh, N.K., Roy, T.K. 2014. On the pattern of migration in the household: an explanation through Binomial Law. *International Journal of Statistics and Systems* **9(2):** 203-214.

Singh, S.N. 1963. A note on Inflated Poisson distribution. *J. Ind. Statist. Assoc.* **1(3):** 140-144.

Singh, S.N., Sharma, H.L. 1984. A study of pattern of out-migration: a regression analysis. *Journal of Rural Development* **3(3):** 166-169.

Singh, S.N., Yadava, K.N.S. 1981. Trends in rural out-migration at household level. *Rural Demography* **8(1):** 53-61.

Singh, S.N., Yadava, R.C., Sharma, L., Sharma, H.L. 1980. Out-migration from Varanasi (Rural): some result on differentials. *Demography India* **9**: 28-31.

Singh, S.N., Yadava, R.C., Sharma, H.L. 1985. A model for rural out-migration at household level. *Janasamkhya -A Journal of Demography* **3(1):** 1-7.

Yadava, K.N.S. 1977. Some aspects of out-migration and its effects on fertility. unpublished Ph.D. thesis in statistics, *Banaras Hindu University*, India.

Yadava, K.N.S. 1989. *Rural-urban migration in India.* Independent Publishing Company, Delhi.

Yadava, K.N.S., Singh, S.R.J. 1983. A model for the number of rural out-migrants at household level. *Rural Demography* **10**: 23-33.

Yadava, K.N.S., Yadava, G.S. 1988. On some probability models and their applications to the distribution of the number of migrants from a household *Janasamkhya-A Journal of Demography* **6(2)**: 137-158.

Table 1. Observed and expected frequency of the number of households according to the migrants in flooded area of Kosi river

Number of migrants	Observed number of households	Expected number of households	
		Zero cell frequency Method	Moment Method
0	401	401.00	403.18
1	147	139.22	136.83
2	57	67.05	66.59
3	29	31.24	31.34
4	16	14.21	14.40
5	8	6.35	6.50
6	5	2.80	2.89
7	1	2.14	2.25
Total	**664**	**664.00**	**664.00**
Mean= 0.7365 Variance= 1.4471		χ^2= 2.98 (after pooling)	χ^2= 2.99 (after pooling)
		p-value = 0.395 (d.f.=3)	p-value = 0.560 (d.f.=4)
Estimated value of parameters		θ = 1.5344	θ = 1.5103
		α = 0.8103	α = 0.7954

Table 2. Observed and expected frequency of the number of households according to the migrants in Varanasi District

Number of migrants	Observed number of households	Expected number of households	
		Zero cell frequency Method	Moment Method
0	1032	1032.00	1036.55
1	95	88.18	81.97
2	19	28.16	28.29
3	10	8.79	9.53
4	2	2.70	3.15
5	2	0.82	1.03
6	0	0.24	0.33
7	1	0.10	0.15
Total	**1161**	**1161.00**	**1161.00**
Mean= 0.1619 Variance= 0.3114		χ^2= 3.94 (after pooling)	χ^2= 5.19 (after pooling)
		p-value = 0.047 (d.f.=1)	p-value = 0.023 (d.f.=1)
Estimated value of parameters		θ = 2.6822	θ = 2.4312
		α = 0.3416	α = 0.3048

Data source: Varanasi (1978)

Table 3. Observed and expected frequency of the number of households according to migrants in Nepal

Number of migrants	Observed number of households	Expected number of households	
		Zero cell frequency Method	Moment Method
0	623	623.00	624.23
1	126	125.20	123.54
2	42	42.26	42.28
3	13	13.93	14.12
4	4	4.51	4.63
5	2	1.44	1.50
6	1	0.66	0.70
Total	**811**	**811.00**	**811**
Mean= 0.3465 Variance= 0.5717		χ^2= 0.09 (after pooling)	χ^2= 0.15 (after pooling)
		p-value = 0.955 (d.f.=2)	p-value = 0.929 (d.f.=2)
Estimated value of parameters		θ = 2.5011	θ = 2.4574
		α = 0.6741	α = 0.6605

Data source: Aryal (2002).

Table 4. Observed and expected frequency of the number of households according to migrants in Comilla district of Bangladesh

Number of migrants	Observed number of households	Expected number of households	
		Zero cell frequency Method	Moment Method
0	1941	1941.00	1949.97
1	542	528.19	515.18
2	124	148.08	149.20
3	48	40.70	42.34
4	13	11.02	11.83
5	4	2.95	3.26
6	1	1.06	1.22
Total	**2673**	**2673.00**	**2673.00**
mean=0.3786 variance=0.5353		χ^2= 6.18 (after pooling)	χ^2= 6.63 (after pooling)
		p-value = 0.046 (d.f.=2)	p-value = 0.036 (d.f.=2)
Estimated value of parameters		θ = 3.1472	θ = 3.0258
		α = 0.9600	α = 0.9176

Data source: Hossain (2000).