

On the Distribution of Risk of Migration and Its Estimation

Brijesh P. Singh¹, Niraj K. Singh²

¹Faculty of Commerce & DST-Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University, Varanasi, India

²Department of Statistics, Amity University, Noida, India

Email address:

brijesh@bhu.ac.in (B. P. Singh), nksingh5@gmail.com (N. K. Singh)

To cite this article:

Brijesh P. Singh, Niraj K. Singh. On the Distribution of Risk of Migration and Its Estimation. *International Journal of Statistical Distributions and Applications*. Vol. 2, No. 4, 2016, pp. 67-71. doi: 10.11648/j.ijstd.20160204.14

Received: September 14, 2016; **Accepted:** November 21, 2016; **Published:** December 30, 2016

Abstract: In India, caste system through economic condition has strong roots in society and it affects the environment in which migration decision takes place. In the present study an attempt has been made to study the trends in rural adult out migration at the household level for different regions to understand the pattern of risk of adult out migration. Some probability models have been proposed to describe the phenomenon and it has been applied to the observed distribution of migrants from the households. Under certain assumptions, it was found that inflated geometric and beta geometric distributions explain satisfactorily the pattern of migration. Also in this study an attempt has been made to know the distribution of risk of migration which cannot be observed directly.

Keywords: Beta Distribution, Inflated Distribution, Adult Migration

1. Introduction

In developing countries, like India over two third populations lived in the rural areas where health and education services, good quality water, electricity supply and employment opportunities are either unavailable or grossly insufficient. These “push and pull” factors caused in a rapid rate of rural to urban migration. The effect of migration on socio-economic activities is thoroughly discussed by Lee (1966) and Oberai et al. (1989). In India, it is well known fact that the caste system is an important aspect of social stratification and it also plays an important role in decision taken for migration. Caste system imposes certain restrictions on persons in deciding the social interaction, choice of occupation and even spatial and vertical mobility. The educational and occupational opportunities and the pattern of land ownership are very closely related with caste groups in India. The rigidity of caste system tends 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.

The ascending order of positions in the caste-hierarchy determines motives of migration; the upper caste groups are structurally different than motives of the lower caste groups and the propensity to migrants is high in the lower caste groups (Kothari, 1980). The rate of migration varies due to

caste and the economic status of the household. 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. 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. Sharma (1985) introduced the idea of including the inflation parameter in modelling trends in rural outmigration at the micro level. He applied the inflated geometric distribution as well as inflated generalized Poisson distribution. Some other studies have been done to formulate probability models in the field of migration (Yadava, 1977; Sharma, 1987; Iwunor, 1995 and Aryal, 2011). Recently Singh et al. (2014) developed a model for adult migration for the fixed household size to know the effect of size of household on the adult migration. Singh et al. (2015) introduced inflated Poisson-Lindley distribution for the pattern of adult migration in the household. In this study an attempt has been made to know the pattern of risk of adult migration by applying suitable models because the risk of

adult migration is not observable directly. Only the number of adult migration in the household is observed, using this data and probability model this study has been performed.

2. Model

When we think about the number of migrants from household in a society, some households have varying number of migrants and some household have no 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.

2.1. Model-I

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 having intention of migration of the household member and $(1-\alpha)$ be the probability of household having no intention of migration.
- (ii) If p is be the probability of migration from a household which follows the geometric distribution.

If X represents the number of migrants from a household, then X follows the inflated geometric distribution with probability density function as

$$\left. \begin{aligned} P(X=0) &= 1-\alpha + \alpha p \\ P(X=k) &= \alpha q^k p \text{ for } k=1, 2, 3, \dots \end{aligned} \right\} \quad (1)$$

where $p+q=1$.

As mentioned above, Sharma (1985) used the method of moments to estimate the parameters α and p of model (1) 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.

2.2. Model-II

In the model-I, it has been assumed that number of migrants aged more than 15 years in the household having fixed probability ' p ' for all but p is effected by a number of factors and therefore assumption of p being a constant for all households in the population can easily be questioned. Moreover, neither the distribution of p can be directly observed nor data on p can be obtained. In such a situation, it seems more logical to consider that p is a random variable following some distribution. In other words, it is assumed that probability p varies from household to household and follows a probability distribution $g(p)$. Here we are assuming that $g(p)$ is distribution of first kind with parameters (a, b) . The p.d.f.

of p i.e. $g(p)$ is given below:

$$g(p) = \frac{1}{\beta(a,b)} p^{(a-1)} (1-p)^{(b-1)}; \quad 0 \leq p \leq 1; \quad a, b > 0 \quad (2)$$

Since ' p ' the risk of migration varies from 0 to 1 and beta distribution of first kind is flexible enough and thus is capable of accommodating wide range of variability, this model seems to be more realistic model for p . Thus for a given household, the distribution of number of migrant X in the household follow a geometric distribution and the probability of migration ' p ' follows beta distribution of first kind with parameters a and b , the joint distribution of X and p is given by

$$\begin{aligned} P[X=x \cap P=p] &= P[X=x/p] \times g(p) \\ &= p(1-p)^x \frac{1}{\beta(a,b)} p^{(a-1)} (1-p)^{(b-1)} \end{aligned} \quad (3)$$

and marginal distribution of X is as follows

$$P[X=x] = \frac{\beta(a+1, x+b)}{\beta(a,b)}; \quad a, b > 0 \quad (4)$$

Where $x = 0, 1, 2, \dots, \infty$

The above distribution (4) is known as beta-geometric distribution and it is natural extension of geometric model under the consideration for random nature of ' p ' in the population. The parameters a and b are its shape parameters. It may be worthwhile to mention here that the parameters a and b can be estimated (say \hat{a} and \hat{b}) by analyzing data through (4) and hence an estimate of the distribution of p in the population can also be obtained. This model helps us to know the pattern of the risk of migration (p) in the various regions, which is generally unknown.

3. Estimation

In this section, some estimation procedure for the model under consideration will be discussed for the number of migration in the household. For this purpose the method of moment estimation procedure has been considered to estimate the parameters involved in the models. In fact this estimation procedure is easier and quicker than other estimation procedure and also there is no need of soft computation in the estimation procedure.

3.1. Model-I

3.1.1. Method of Moments

Inflated geometric distribution has two parameters α and p to be estimated. Let we have

$$E(X) = \frac{\alpha(1-p)}{p} = \bar{x} \quad (5)$$

and,

$$E(X^2) = \frac{\alpha(1-p)(2-p)}{p^2} \quad (6)$$

from (5) and (6) we get

$$\begin{aligned}
 E(X^2) &= \bar{x} \cdot \left[\frac{2-p}{p} \right] \Rightarrow \frac{2-p}{p} = \frac{E(X^2)}{\bar{x}} \\
 \Rightarrow \frac{2}{p} - 1 &= \frac{E(X^2)}{\bar{x}} \Rightarrow \frac{2}{p} = \frac{E(X^2) + \bar{x}}{\bar{x}} \\
 \therefore \hat{p} &= \frac{2\bar{x}}{(E(X^2) + \bar{x})} \\
 \text{or, } \hat{p} &= \frac{2\bar{x}}{\text{variance} + \bar{x}^2 + \bar{x}}
 \end{aligned} \quad (7)$$

Putting the estimate of p in equation (5) we get the estimate of α , i.e.

$$\hat{\alpha} = \frac{p\bar{x}}{1-p} = \frac{2\bar{x}^2}{\text{variance} + \bar{x}^2 - \bar{x}}$$

3.1.2. Second Method

We know that from equation (5)

$$\bar{x} = \frac{\alpha(1-p)}{p} \Rightarrow \alpha(1-p) = p\bar{x}$$

Also we know that

$$\begin{aligned}
 \frac{n_0}{N} &= 1 - \alpha + \alpha p = 1 - \alpha(1-p) = 1 - p\bar{x} \\
 \therefore p\bar{x} &= \frac{N - n_0}{N} \Rightarrow \hat{p} = \frac{N - n_0}{N\bar{x}} \\
 \therefore \hat{\alpha} &= \frac{N - n_0}{N(1-p)}
 \end{aligned}$$

Here n_0 is the first cell frequency. The second method is easier and quicker to get the estimate. In the above method we need $E(X)$, $E(X^2)$ or $\text{var}(X)$ but in this method we need only \bar{x} .

3.2. Model-II

The estimators of parameters a and b of model-II can be obtained as follows

$$E(X) = \frac{b}{(a-1)} \quad \text{for } a > 1 \quad (8)$$

$$E(X^2) = \frac{b[2b+a]}{(a-1)(a-2)} \quad \text{for } a > 2 \quad (9)$$

Let μ'_1 and μ'_2 denotes the first two raw moments about origin for data in hand. Replacing $E(X)$ and $E(X^2)$ by μ'_1 and μ'_2 in above equations we get two equations with two unknowns a and b as given below:

$$\mu'_1 = \frac{b}{(a-1)} \quad (10)$$

and

$$\mu'_2 = \frac{b(2b+a)}{(a-1)(a-2)} \quad \text{for } a > 2 \quad (11)$$

We know that the

$$\text{Variance} = \mu'_2 - \mu'^2_1 = \frac{b(2b+a)}{(a-1)(a-2)} - \left(\frac{b}{a-1} \right)^2 \quad \text{for } a > 2 \quad (12)$$

After simplification it becomes

$$\text{Variance} = \frac{ab(a+b-1)}{(a-1)^2(a-2)} \quad \text{for } a > 2 \quad (13)$$

From equation (10) we have

$$a-1 = \frac{b}{\mu'_1} \quad \text{or} \quad a = \frac{b + \mu'_1}{\mu'_1} \quad (14)$$

Now substituting the value of a in the equation (13) and get

$$\begin{aligned}
 \text{Variance} = V \text{ (say)} &= \frac{b \left(\frac{b + \mu'_1}{\mu'_1} \right) \left(\frac{b + \mu'_1}{\mu'_1} - 1 \right)}{\left(\frac{b + \mu'_1}{\mu'_1} - 1 \right)^2 \left(\frac{b + \mu'_1}{\mu'_1} - 2 \right)} \\
 &= \frac{b^2 (b + \mu'_1) (\mu'_1 + 1) \mu'^3_1}{b^2 (\mu'^2_1) (b - \mu'_1)}
 \end{aligned} \quad (15)$$

$$\begin{aligned}
 \text{Hence } V &= \frac{(b + \mu'_1) (\mu'_1 + 1) \mu'_1}{(b - \mu'_1)} \\
 &\Rightarrow \frac{V}{\mu'_1 (\mu'_1 + 1)} = \frac{(b + \mu'_1)}{(b - \mu'_1)}
 \end{aligned} \quad (16)$$

here we assume that $\frac{V}{\mu'_1 (\mu'_1 + 1)}$ is equal to A .

$$\text{then } A = \frac{(b + \mu'_1)}{(b - \mu'_1)} \quad \text{or} \quad b = \frac{\mu'_1 (A + 1)}{(A - 1)}$$

putting the value of A and μ'_1 from the data the value of b can be estimated. After estimating the value of b , the estimate of a can be obtained by using equation (14).

4. Application of the Models

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 discussed have been applied to other three data sets for different time and space. The other data sets are the same as used by Hossain

(2000) for Bangladesh, Aryal (2011) for Nepal and 'Rural development and population growth (RDPG) survey' conducted in 1978 in Varanasi district.

5. Result and Conclusion

Table 1 shows the estimated values of parameters, mean, variance, observed and expected number of total households according to the number of adult migrants for households of the different region. The value of χ^2 with degree of freedom and p -value are also given in the respective tables. The value of χ^2 shown in the tables clearly indicate that both the models describe the distribution of number of migrants for households. From the table 1, it is clear that for Varanasi and Bangladesh data, both the models works satisfactory well but for Koshi, Bihar and Nepal data both the model works excellently (on the basis of p -value). But according to the p -value, inflated geometric distribution is found better than beta geometric distribution for Nepal and Koshi data and for Varanasi and Bangladesh data beta geometric distribution is found slightly better. According to the model-I the value of parameter α is

about 86 percent households in Koshi, Bihar and 70 percent in Nepal are expected to have adult migrants however in Varanasi it is 35 percent. The value of α is 0.99 in Bangladesh means almost all households having tendency of adult migration. The risk of migration is highest in Bangladesh is 0.72, in Nepal it is 0.67. In Varanasi the risk is 0.69 and in Koshi, Bihar it is 0.54. The value parameters of model-II are used to know the pattern of risk of migration because we assume in the model, the risk of adult migration follows beta distribution. Therefore, the probability curve of Beta distribution has been drawn using the value of parameters for the various regions considered in the study. Figure 1 shows the distribution of risk of adult migration according to different region. For the Varanasi data the distribution of risk of adult migration is highly positively skewed and for Bangladesh it is peaked and slight positively skewed. For Koshi, Bihar and Nepal the risk is almost normal but flat, thus one can draw the conclusion that the pattern of risk of adult migration is different for different region. Also this model facilitates us by providing a technique of understanding the behavior of unobservable (risk of adult migration) through observable (number of adult migration in the household).

Table 1. Observed and expected frequency of the number of households according to the adult migrants.

Number of migrants	Koshi			Varanasi		
	Observed number of households	Expected number of households from		Observed number of households	Expected number of households from	
		Inflated geometric	Beta geometric		Inflated geometric	Beta geometric
0	401	401.0	392.9	1032	1032.0	1027.8
1	147	141.5	155.6	95	88.5	100.6
2	57	65.4	64.6	19	27.8	21.6
3	29	30.2	28.1	10		
4	16	13.9	12.6	2		
5	8	6.5		2	12.7	11.0
6	5		10.2	0		
7	1	5.5		1		
Total	664	664.0	664.0	1161	1161.0	1161.0
χ^2		2.044	3.940		3.679	2.080
p -value		0.73(df=4)	0.27(df=3)		0.06(df=2)	0.15(df=1)
Estimated value of parameters		$p=0.5378$ $\alpha=0.8570$	$a=17.1960$ $b=11.9275$		$p=0.6862$ $\alpha=0.3541$	$a=5.0527$ $b=0.6563$
Mean	0.7365			0.1619		
Variance	1.4471			0.3114		

Table 1. Continued.

Number of migrants	Nepal			Bangladesh		
	Observed number of households	Expected number of households from		Observed number of households	Expected number of households from	
		Inflated geometric	Beta geometric		Inflated geometric	Beta geometric
0	623	623.0	617.3	1941	1941.0	1945.9
1	126	125.9	138.1	542	529.5	524.8
2	42	41.7	37.5	124	146.5	144.9
3	13	13.8	11.7	48	40.5	41.0
4	4			13		
5	2			4		
6	1	6.6	6.4	1	15.5	16.4
7	-			-		
Total	811	811.0	811.0	2673	2673.0	2673.0
χ^2		0.08	1.86		5.28	4.97
p -value		0.78(df=2)	0.39(df=2)		0.06(df=2)	0.08(df=2)
Estimated value of parameters		$p=0.6690$ $\alpha=0.7004$	$a=10.8739$ $b=3.4212$		$p=0.7233$ $\alpha=0.9898$	$a=80.1332$ $b=29.9599$
Mean	0.3465			0.3786		
Variance	0.5717			0.5353		

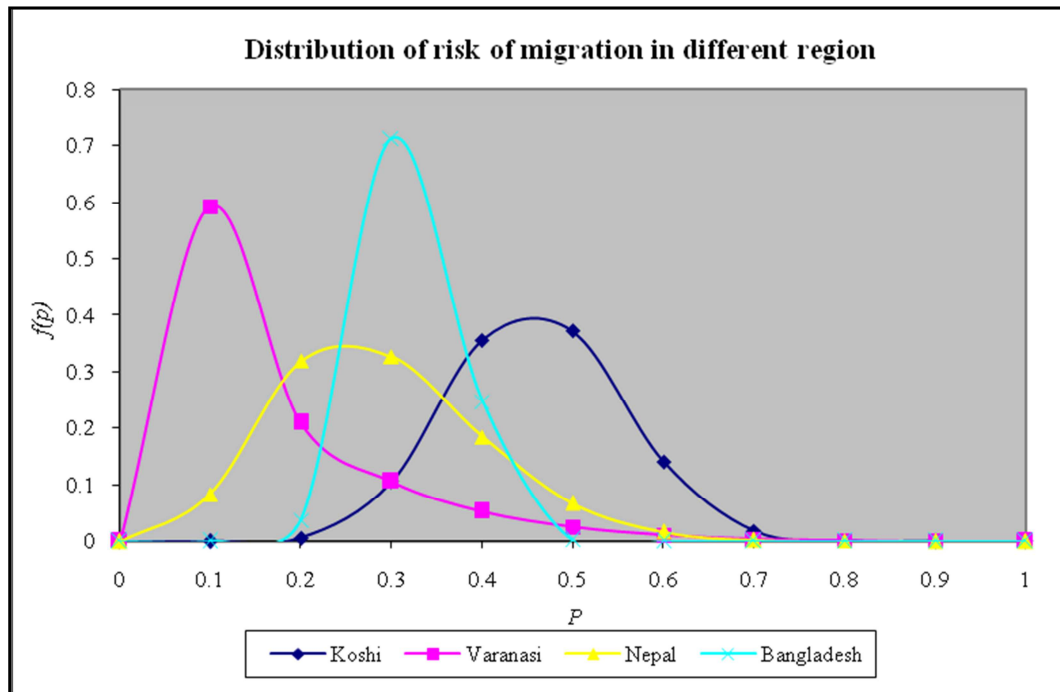


Fig. 1. Distribution of risk of migration for different regions.

References

- [1] Aryal, T. R. (2011): "Inflated Geometric Distribution to Study the Distribution of Rural Out- migrants", *Journal of the Institute of Engineering*, Vol. 8, No. 1, pp. 266-268.
- [2] Gerstenkom, T. (1979): "Moment recurrence relation for the Inflated Negative Binomial, Poisson and Geometric distribution", *Demonstration Mathematica*, Vol. 11, No. 2, pp 389-411.
- [3] Hossain, M. Z. (2000): "Some demographic models and their application with special reference to Bangladesh", unpublished Ph.D. thesis, Banaras Hindu University, India.
- [4] Iwunor, C. C. O. (1995): "Estimating of parameters of the inflated geometric distribution for rural out-migration", *Genus*, Vol. 51, pp. 3-4.
- [5] Kothari, D. K. (1980): "Patterns of rural urban migration", *Australian National University Press*, Canberra.
- [6] Lee, E. S. (1966): "A Theory of Migration", *Demography*, Vol. 3, 47-57.
- [7] Oberai, A. S., Prasad, P. H. & Sardana, M. G. (1989): "Determinants and Consequences of Internal Migration in India", *Oxford University Press*, Delhi.
- [8] Sharma, H. L. (1985): "A probability distribution for rural out migration at micro level", *Rural Demography*, Vol. 12, No. 1 & 2, 63-69.
- [9] Sharma, H. L. (1987): "A Probability Distribution for Out migration", *Janasamkhyā*, Vol. 5, No. 2, 95-101.
- [10] Singh, Brijesh P., Singh, Niraj K. & Roy, Tapan K. (2014): "On the Pattern of Migration in the Household: An Explanation through Binomial Law" in *International Journal of Statistics and Systems*, Volume 9, No. 2, pp. 203-214.
- [11] Singh, Brijesh P., Singh, Niraj K. & Dixit, Shweta (2015): "Estimation of Parameters of Inflated Poisson-Lindley Distribution for Adult Out-migration", in *Journal of Institute of Science and Technology*, Vol. 20, No. 2, pp. 6-10.
- [12] Singh, S. N. (1963) "A note on Inflated Poisson distribution", *Journal of Indian Statistical Association*, Vol. 1, No. 3, pp. 140-144.
- [13] Singh, S. N. & Yadava, K.N.S. (1981): "Trends in rural out-migration at household level", *Rural Demography*, Vol. 8, No. 1, pp. 53-63.
- [14] 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.