On the Pattern of Migration in the Household: An Explanation through Binomial Law

Introduction

Migration, conventionally, being an important part of demography, is least studied as compared to fertility and mortality. Due to decreasing birth and death rate, migration (internal and international) has become a more important concern for demographers and other social scientists. In developing countries particularly in India where about seventy per cent people still live in villages (Census 2011), migration from rural areas has become a major subject of interest for social scientists as well as planners.

Although, macro level studies have their own importance since this approach describes aggregate flow of rate of migration and identify factors influencing out migration (Banerjee, 1986), the behavioral parameters of process can be explained through micro level studies i.e., at the level of household or individual. Micro level studies have important implications for housing policies and also for the development of other sociological models related families and communities (Pryor, 1975; Rossi, 1995). At the micro level, the topic of household is getting prominence in demography to understand demographic process. The recent studies indicate the importance of demographic behaviors of the individual and household level factors (Pressure and Das, 2001). It is seen that a migrant household (with one or more persons involved in the process of migration in relation to do some job outside the village) may have different socio-economic and cultural characteristics through remittances besides providing good ideas, awareness and environments than a non-migrant household (Yadava, 2010).

A household, especially in Indian context is a basic socio-economic unit for the integrated rural development. Household characteristics (age, size, occupation, socio-economic status, etc.) play a vital role in decision of its members to move or not to move. A study of movement process at the household level is also useful for the prediction of future size of the households as well as to study the imbalances in sex-ratio occurred due to such migration. Motivated by the fact that the life time data related to migration may have more biases than data on number of migrants and keeping in view that the study of the distribution of migrants has been a powerful device for explaining changes and variation in the population, in the present study an attempt has been made to study the trends in rural out migration at the household level where size of household is fixed. A probability distribution model has been discussed to describe the phenomenon and it has been applied to the observed distribution of migrants from the households. The study aims to search a suitable model for analyzing the risk of migration in the population and to use it for comparing the risk of migration with respect to space and time.

Construction of models

Consider the households of size n. Let p be the probability of migration of a person in the household and X be the total number of migrated persons out of n persons. Thus X is a random variable which denotes the number of migrants for fixed household size n.

The model is developed under the consideration that each person of the household is either a migrant or non-migrant. Let us define for i^{th} (i=1,2,…, n) person lived in the household, a random variable z_i taking value 1 if the person migrated and 0 otherwise. Thus z_i's are Bernoulli variable. Now if we assume that migration of persons lived in the household are independent of each other and having same probability p then total number of migrated persons X from the household is nothing but sum of independent Bernoulli variables and hence follows a binomial distribution. Therefore, the distribution of X may be given by

$$P[X = x] = \binom{n}{x} p^{x}(1-p)^{n-x} \qquad ; \quad 0 \le p \le 1 \qquad (1)$$

where, $x = 0, 1, 2 \ldots\ldots\ldots\ldots n.$

Model-I

In this model, we consider that the population consist a high proportion of households having no migrants. Due to more observations with zero counts, the frequency of zero cells is inflated and the resulting over dispersion can not be modeled accurately with the simple binomial model. In such scenario an inflated binomial model may be. Assume that the proportion of households prone to the migration be α and $(1-\alpha)$ proportion have no migrants in the household. Therefore the probability density function of zero inflated binomial model is

$$P[X=x]=\begin{cases}(1-\alpha)+\alpha(1-p)^n & for\ x=0\\ \alpha\binom{n}{x}p^x(1-p)^{n-x} & for\ x=1,2,3........n\end{cases} \qquad (2)$$

The zero class data can be partitioned lacking households having no migrants (denoted by X_{00}) and the households have any migrants but no response were recorded (denoted by X_{0l}). X_{00} is estimated by (1-α) and X_{0l} and estimated as (N_0-1+α), where N_0 is the proportion of zeros cell frequency.

Model-II

In the model 1, we have assumed that probability of migration 'p' from a household is fixed for all. But in reality, 'p' is affected by a number of factors and therefore assumption of p being constant for all households seems to be questionable. Thus, it seems more logical to consider p as a random variable following some distribution *g(p)*. Beta distribution of first kind with parameters (*a*, *b*) is a suitable distribution for risk of migration '*p*', since '*p*' the risk varies from 0 to 1 and beta distribution possess the property of flexibility, and capability of accommodating wide range of variability. The probability density function of Beta distribution is:

$$g(p)=\frac{1}{\beta(a,b)}p^{(a-1)}(1-p)^{(b-1)}; \qquad 0\le p\le 1; \qquad a,b>0 \qquad (3)$$

Thus, the joint distribution of *x* and *p* is given by

$$P[X=x\cap P=p]=P[X=x/p]\times g(p) = \binom{n}{x}p^x(1-p)^{n-x}\frac{1}{\beta(a,b)}p^{(a-1)}(1-p)^{b-1} \qquad (4)$$

and the marginal distribution of *x* is given as

$$P[X=x]=\binom{n}{x}\frac{\beta(a+x,b+n-x)}{\beta(a,b)}; \qquad a,b>0 \qquad (5)$$

where, x = 0, 1, 2……………..*n*

The above distribution (5) is known as beta-binomial distribution and it is natural extension of binomial model under the consideration for random nature of '*p*' in the population. The parameters *a* and *b* are its shape parameter. If someone is interested in getting a single value (like p) for comparing the migration of two places, one may take mean i.e. $\frac{\hat{a}}{\hat{a}+\hat{b}}$ as an estimate of average number of migrants at the household level.

Model-II : Method of Moment (MM)

The moment estimates of the parameters a and b can be obtained as follows

$$E(X) = \frac{na}{(a+b)} \tag{10}$$

$$E(X^2) = \frac{na[n(1+a)+b]}{(a+b)(a+b+1)} \tag{11}$$

As mentioned above replacing $E(X)$ and $E(X^2)$ by $\mu_1^{'}$ and $\mu_2^{'}$ in above equations we get two equations with two unknowns a and b as given below:

$$\mu_1^{'} = \frac{na}{(a+b)} \tag{12}$$

and $$\mu_2^{'} = \frac{na[n(1+a)+b]}{(a+b)(a+b+1)} \tag{13}$$

Substituting the value of $b = \left(\frac{n-\mu_1^{'}}{\mu_1^{'}}\right)a$ from the equation (12) in the above equation and separating the coefficients for a we have

$$a\left[\mu_2^{'} - n\mu_1^{'} + \left(\frac{n-\mu_1^{'}}{\mu_1^{'}}\right)(\mu_2^{'} - \mu_1^{'})\right] = n\mu_1^{'} - \mu_2^{'} \tag{14}$$

or $$a = \frac{n\mu_1^{'2} - \mu_1^{'}\mu_2^{'}}{n(\mu_2^{'} - \mu_1^{'}) - n\mu_1^{'2} + \mu_1^{'2}} \tag{15}$$

after solving this we can get the estimate of a and the using this estimate and equation (12) b can be estimated easily.

We have also used maximum likelihood method to estimate the parameters involved in the models.

Application of the Models

The models has 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. Data have been collected using a multistage random sampling procedure.

This analysis is based on the information collected from 664 households. The households with inadequate and incomplete information have been excluded.

Result and Discussion

In this section we have discussed about the estimate of parameters and fitting of the proposed distributions. Since, in this study two models are proposed for fixed household size so that after obtaining the estimate of parameters for different household sizes, we obtained the estimated frequencies for both the models. Table 1 shows the distribution of households according to their size and the number of migrants. Tables 2-5 show the expected frequencies along with the observed frequencies for household size 5-8 of Koshi effected region in Bihar. Estimate of parameters, the value of χ^2 with degree of freedom along with average risk of migration from a household are given in the respective tables. The value of χ^2 shown in the tables clearly indicate that both the models for distribution of number of migrants for fixed size of households described well. The advantage of model-I is that the parameters involved in the model have physical meaning such that *p* provides the risk of migration at the household level whereas (1-α) gives the proportion of households where migration does not occur. From the tables it can be easily seen that the risk of migration increases with the increased size of households and the proportion of households.

In the table 2 and 3 where the distribution of migrants among the households of size 5 and 6 are given, the value of χ^2 cannot be calculated due to the degree of freedom comes out to be zero. However, to see the fitting we have drawn the plot of empirical distribution function. Figures 1 and 3 give the plot of inflated binomial distribution for household size 5 and 6 respectively. Figure 2 and 4 give the plot of beta binomial distribution for the household size 5 and 6. From the plot it was found that both the models fit the data satisfactorily well. For the comparison purpose Akaike's Information Criterion (AIC) have been calculated. We see that in both the cases AIC for beta binomial is smaller than inflated binomial. Thus we can say that beta binomial gives a better fit in both the cases i.e., for the household of size 5 and 6.

Another added advantage of use of the model-II may be that as soon as we get the estimate of *a* and *b*, an estimated distribution for *p* for the population can be obtained. It is worthwhile to note that the model provides a way to study the distribution of *p* which cannot be studied otherwise since *p* itself is unobservable. Figure 5 shows the distribution of risk of migration which is not observable directly. For lower household size the risk of migration is left skewed and leptokurtic than higher household size and as size of household increases the distribution of risk of migration becomes flatter and flatter. Figure shows clearly that the location measures of migration for lower household size are less than the higher household size.

Table 1: Distribution of number of migrants according to the household size

Number of migrants	**<=4**	**5**	**6**	**7**	**8**	**9+**	**Total**
0	176	123	58	26	14	4	401
1	29	33	33	23	10	19	147
2	2	9	9	10	11	16	57
3	-	2	2	7	6	12	29
4	-	-	1	1	2	12	16
5	-	-	-	-	-	8	8
6	-	-	-	-	-	5	5
7	-	-	-	-	-	1	1
Total	207	167	103	67	43	77	664

Table 2: Expected & Observed distributions of migrants in the household with size 5

Number of migrants	Observed number of households	Expected number of households	
		Inflated Binomial	**Beta Binomial**
0	123	128.5016	122.5090
1	33	30.4421	34.1797
2	9	8.0563	10.3113
3	2		
Total	167	167	167
$\chi^2_{0.05}(0)$		-	-
mean=	0.3413	p=0.1053	a=0.9363
variance=	0.4045	α =0.5404	b=12.7795

Table 3: Expected & Observed distributions of migrants in the household with size 6

Number of migrants	Observed number of households	Expected number of households	
		Inflated Binomial	**Beta Binomial**
0	58	59.5210	58.8268
1	33	29.0810	31.3147

2	9	14.3980	12.8585
3	2		
4	1		
Total	103	103	103
$\chi^2_{0.05}(0)$		-	-
mean=	0.5922	p=0.1377	a=2.1824
variance=	0.6493	α =0.7168	b=19.9280

Table 4: Expected & Observed distributions of migrants in the household with size 7

Number of migrants	Observed number of households	Expected number of households	
		Inflated Binomial	Beta Binomial
0	26	26.6362	25.5239
1	23	20.4657	23.0947
2	10	13.6030	12.1838
3	7	6.2951	6.1976
4	1		
Total	67	67	67
$\chi^2_{0.05}(1)$		1.7445	0.9249

mean=	1.0149	p=0.1814	a=3.2623
variance=	1.0893	α =0.7994	b=19.2379

Table 5: Expected & Observed distributions of migrants in the household with size 8

Number of migrants	**Observed number of households**	**Expected number of households**	
		Inflated Binomial	**Beta Binomial**
0	14	11.9337	11.9156
1	10	10.1050	14.2983
2	11	10.9028	9.7297
3	6	10.0585	7.0564
4	2		
Total	43	43	43
$\chi^2_{0.05}(1)$		0.7808	1.9488
mean=	1.3488	p=0.2356	a=4.0326
variance=	1.4365	α =0.8178	b=19.8850

Table 6: Expected & Observed proportion of Zeroth cell of the distributions migrants in the household

Household size	Zeroth cell proportion [N_0]	proportion of households having migrants [$X_{00}=(1-\alpha)$]	proportion of households having migrants but respond no (zero migrants) [$X_{0I}=(N_0-1+\alpha)$]
5	0.74	0.46	0.28
6	0.56	0.28	0.28
7	0.39	0.20	0.19
8	0.33	0.18	0.14

Fig 1: Distribution of risk of migration according to the household size