The Nepali Math. Sc. Report
Vol. 23, No. 2, 2004.

A Mixture of Two Displaced Geometric Distributions for Describing the Distribution of the Total Number of Migrants at Micro-Level

TIKA RAM ARYAL

Abstract: The aim of this paper is to study the distribution of the total number of rural outmigrants from a household. Probability model has been used for this purpose. The maximum likelihood estimation technique has been proposed to estimate the parameters involved in the model. The data has been taken from a sample survey of Palpa and Rupandehi Districts of western Nepal. The proportion of households having only one migrant was found to be 0.52381 whereas the average number of migrants from the household was 0.57831. The asymptotic variances and co-variances of the estimators have also been obtained. A mixture of two displaced geometric distribution fit the data sets satisfactorily well. Thus proposed model describe the distribution of the total number of rural out-migrants from rural areas of Rupandehi and Palpa districts of western Nepal.

Key words: model, probability, frequency, maximum likelihood, risk, parameter, logarithm.

1. Introduction

Researchers have given their due attention on the formulation of models and their applications due to its usefulness and applicability in social sciences [1]. The macro-level migration studies have their own importance [2]. These approaches describe aggregate flow/rate of migration, and recognize the factors motivating out-migration [3]. Micro-level (i.e. at the level of household or individual) studies help to classify the behavioural parameter of migration process. There is a rare study in out-migration at micro-level, which may be due to the lack of interest of the researcher as well as unavailability of the data.

Several attempts have been made to documents the pattern of rural out-migration through probability models [2, 4, 7, 11]. The idea of cluster was integrated in the model by Yadava and Singh [10]. It was found that Thomas distribution is well suited to describe the number of migrants from a household.

Yadava and Yadava [9] further extended by assuming the occurrence of migration in cluster varies from household to household and the number of migrants to a cluster follows truncated displaced geometric distribution. Under such assumptions, probability model fitted well to the distribution of male migrants aged 15 years and above. However, these models do not fit the distribution of total number of migrants including their wife and children from a household.

Sharma [5,6] proposed a probability model with some assumptions : (i) the number of male migrants aged 15 years and above follows negative binomial distribution and (ii) the distribution of alive children to a couple be known. However, the prior knowledge about these two distributions is difficult since the distribution of children alive to a couple has not yet been derived theoretically. Singh [8] proposed a probability model for the total number of migrants under the assumption that there are two types of households i.e. the households from where male members aged 15 years and above migrate singly leaving their wives and children at home, and the households where male members migrate with their wives, children and other dependent relatives. Yadav [12] proposed a probability model to describe the distribution of households according to total number of migrants.

Moment techniques or mean-zero frequency method have been used to estimate the parameters involved in their proposed models. About 80 to 85 percent variation in migration is equated through zero'th cell frequencies i.e. all the non-migrant households are counted. About 15 to 20 percent variations are explained by the estimated parameters when mean-zero frequency method is applied [2]. Moment estimates are generally consistent, but they are often less efficient. By taking such limitations, the maximum likelihood estimation technique is applied to estimate the parameters involved in the proposed model. Maximum likelihood method provides standard error of the estimators as well as measures the total variation of the distribution.

The main aim of this paper is to study the distribution of the total number of rural out-migrants through probability model. Maximum likelihood estimate technique has been proposed to estimate the parameter involved in the model. The asymptotic variances and co-variances of the estimators have also been discussed. The suitability of the model has been tested to the data of Palpa and Rupandehi districts of western Nepal.

2. Model

Yadava [12] proposed a probability model to describe the distribution of households according to total number of migrants. Mean-zero frequency method was used to estimate the parameters involved in the model. Here we used maximum likelihood estimation technique to estimate the parameters involved in the model by avoiding the limitation of mean-zero frequency method. In brief, the model along with their assumptions is given below :

(i) At the survey point, let β be the proportion of households which poses at-least one migrants

(ii) Out of β proportion of households, let ξ be the proportion of households, which poses only one migrant at the survey point.

(iii) Out of $(1-\xi)\beta$ proportion of households, let π be the proportion of households from which only males ≥ 15 years migrate and $(1-\pi)$ be the proportion of households which poses both types of migrants (males ≥ 15 years as well as males with their families).

(iv) The number of migrants from a household follows a mixture of two displaced geometric distributions with π proportion of households from which only males aged 15 years migrates and $(1-\pi)$ be the proposition of households from which both type of migration occur.

(v) Let p_1 and p_2 be the probability of migration of a person from π and $(1-\pi)$ proportions of households respectively.

Under these assumptions, the probability distribution for the total number of migrants, X, is given as

$$
\begin{aligned}
p(X=k) &= 1-\beta \text{ if } k=0 \\
&= \xi\beta \text{ if } k=1 \\
&= (1-\xi)\beta\{\pi p_1 q_1^{k-2} + (1-\pi) p_2 q_2^{k-2}\} \text{ if } k=2,3,\ldots
\end{aligned} \tag{1}
$$

This model involves β, ξ, π, p_1, p_2 parameters, which is difficult to estimate from the observed data set. Assume that $p_1 = p_2 = p$, which is equivalent to the probability of migration from both types of households is same and (1) becomes

$$
\begin{aligned}
p(X=k) &= 1-\beta \text{ for } k=0 \\
&= \xi\beta \text{ for } k=1 \\
&= (1-\xi)\beta\, pq^{k-2} \text{ for } k=2,3,\ldots
\end{aligned} \tag{2}
$$

3. Estimation of Parameters

Model (2) involves ξ, β and p parameters that are estimated by using observed data. Let $X_1, X_2, \ldots X_n$ denote a random sample of size n from the expression (2). Suppose, n_k $(k = 0,1,2,\ldots, m)$ be the number of observations corresponding to the value of k such that $\sum_{k=0}^{m} n_k = n$. Likelihood function for the given sample is

$$L=\prod_{k=0}^{m}[p(X=k)]^{n_k}=(1-\beta)^{n_0}\ (\xi\beta)^{n_1}\ \prod_{k=2}^{m}[(1-\xi)\beta\, pq^{k-2}]^{n_k}$$

(3) $$=(1-\beta)^{n_0}\,\xi^{n_1}\ \beta^{n-n_0}\ (1-\xi)^{n-n_0-n_1}\,p^{n-n_0-n_1}\,q^{\sum_{k=2}^{m}(k-2)n_k}$$

Taking log in (3) and differentiating w.r.t. β, ξ and p respectively and by equating it to zero, then we get,

(4) $$\frac{\delta \log L}{\delta\beta}=-\frac{n_0}{1-\beta}+\frac{n-n_0}{\beta}=0$$

(5) $$\frac{\delta \log L}{\delta\xi}=\frac{n_1}{\xi}-\frac{n-n_0-n_1}{1-\xi}=0$$

(6) $$\frac{\delta \log L}{\delta p}=\frac{n-n_0-n_1}{p}-\frac{\sum_{k=2}^{m}(k-2)n_k}{1-p}=0$$

By solving (4),(5) and (6), the estimators of β, ξ and p is obtained

$$\hat{\beta}=\frac{n-n_0}{n}\ ,\ \hat{\xi}=\frac{n_1}{n-n_0}\ \text{ and }\ \hat{p}=\frac{n-n_0-n_1}{(n-n_0-n_1)+\sum_{k=3}^{m}(k-2)n_k}$$

The second partial derivations of log L is given as

(7) $$\frac{\delta^2 \log L}{\delta\beta^2}=-\frac{n_0}{(1-\beta)^2}-\frac{n-n_0}{\beta^2}$$

(8) $$\frac{\delta^2 \log L}{\delta\xi^2}=-\frac{n_1}{\xi^2}-\frac{(n-n_0-n_1)}{(1-\xi)^2}$$

(9) $$\frac{\delta^2 \log L}{\delta p^2}=-\frac{(n-n_0-n_1)}{p^2}-\frac{\sum_{k=3}^{m}(k-2)n_k}{(1-p)^2}$$

(10) $$\frac{\delta^2 \log L}{\delta\beta\delta\xi}=\frac{\delta^2 \log L}{\delta\beta\delta p}=\frac{\delta^2 \log L}{\delta\beta\delta p}=0$$

Here, $E(n_0)=E[\sum_{i=1}^{n}1_{\{X_i=0\}}]=\sum_{i=1}^{n}1p\{X_i=0\}=\sum_{i=1}^{n}(1-\beta)=n(1-\beta)$, in similar manner we can write $E(n_1)=n\xi\beta$, $E(n_k)=n(1-\xi)\ \beta\, pq^{k-2}$ for $k=2,3,\ldots,m$, $E(n-n_0)=np$, $E(n-n_0-n_1)=n\beta(1-\xi)$ and

$$E\,[\sum_{k=3}^{m}(k-2)n_k] = E\,[n_3 + 2n_4 + 3n_5 + \ldots + (m-2)n_m]$$

$$= n(1-\xi)\,\beta\,pq\,[1 + 2q + 3q^2 + \ldots + (m-2)\,q^m\,]$$

(11) $$= n(1-\xi)\,\beta\,q\left[\frac{1-q^{m-2}}{p} - (m-2)\,q^{m-2}\right] \text{ for small } m$$

(12) $$= \frac{n(1-\theta)\,\beta\,q}{p} \text{ for large } m$$

By using these facts, the expected values of the second partial derivatives are

(13) $$-E\left(\frac{\delta^2 \log L}{\delta\beta^2}\right) = \frac{E(n_0)}{(1-\beta)^2} - \frac{E(n-n_0)}{\beta^2} = \frac{n}{\beta(1-\beta)} = \phi_{11} \text{ (say)}$$

(14) $$-E\left(\frac{\delta^2 \log L}{\delta\xi^2}\right) = \frac{E(n_1)}{\xi^2} + \frac{E(n-n_0-n_1)}{(1-\beta)^2} = \frac{n\beta}{\xi(1-\xi)} = \phi_{22} \text{ (say)}$$

$$-E\left(\frac{\delta^2 \log L}{\delta p^2}\right) = \frac{E(n-n_0-n_1)}{p^2} + \frac{E\,[\sum_{k=3}^{m}(k-2)n_k]}{(1-p)^2}$$

$$= \frac{n\beta\,(1-\xi)q + n(1-\xi)\,\beta\,p[1-q^{m-2}-(m-2)pq^{m-2}]}{p^2 q}$$

(15) $$= \phi_{33} \text{ (a) (say), for small } m$$

and

(16) $$-E\left(\frac{\delta^2 \log L}{\delta p^2}\right) = \frac{n\beta\,(1-\xi)}{p^2 q} = \phi_{33} \text{ (b) (say) for large } m$$

Covariance between the estimators equal zero since

$$E\left(\frac{\delta^2 \log L}{\delta\beta\delta\xi}\right) = E\left(\frac{\delta^2 \log L}{\delta\xi\delta p}\right) = E\left(\frac{\delta^2 \log L}{\delta\beta\delta p}\right) = 0$$

and the asymptotic variances of the estimators can be obtained as:

$$V(\hat{\beta}) = \frac{1}{\phi_{11}}, \quad V(\hat{\xi}) = \frac{1}{\phi_{22}}, \text{ and}$$

(17) $$V(\hat{p}) = \frac{1}{\phi_{11}(a)} \text{ when } m \text{ is small}$$

$$= \frac{1}{\phi_{22}(b)} \text{ when } m \text{ is large,}$$

4. Applications

The probability model discussed in this paper for describing the total number of migrants from a household is fitted using the maximum likelihood estimators to the data collected from rural areas of Palpa and Rupandehi districts of Nepal. These data were collected under a sample survey "Demographic Survey on Fertility and Mobility in Rural Nepal (DSFM, 2000) : A Study of Palpa and Rupandehi Districts" during January-June, 2000. The detail about the sample survey is discussed in Aryal [1].

Table 1: Distribution of Observed and Expected Number of Households According to the Total Number of Male Migrants in Rural, Nepal.

Number of migrants per household	Observed	Expected
0	622	622.00
1	99	99.00
2	27	30.98
3	17	21.02
4	13	14.26
5	11	9.68
6	7	6.57
7	4	7.49
8	11	
Total	811	811.00
χ^2		9.13
d.f.		4
$\hat{\beta}$		0.23305
$\hat{\xi}$		0.52381
$\hat{p}$		0.32143
$V(\hat{\beta})$		0.00022
$V(\hat{\xi})$		0.00132
$V(\hat{p})$		0.00079
Covariances		0.00000
Average Number of migrants per households		0.57831

The observed and expected number of households (along with the variances and co-variances between the estimators) according to the total number of migrants is presented in Table 1. The proportion of households that poses at-least one migrants ($\hat{\beta}$) was 0.23305 and out of which the proportion of households having only one migrant ($\hat{\xi}$) was found to be 0.52381. The probability of migration of a person from households ($\hat{p}$) was found to be 0.32143. The higher proportion of households having only one migrant may be due to higher cost on travel and higher cost of living at the place of destination along with their families. Moreover, most of the migrants move for a certain period of time and after completion of their tenure they have to return back to their home. Further, the higher value of $\hat{\beta}$ indicated that a higher rate of migration was observed in the rural migrant's households.

The average number of migrants per household can be obtained as $\hat{\xi}\hat{\beta} + (1-\hat{\xi})\hat{\beta}\left(1+\frac{1}{\hat{p}}\right)$. The average number of migrants from the household was found to be 0.57831. The chi-square value was found insignificant which confirms that the model fits the data sets reasonably well. Thus the model is a reasonable approximation to describe the distribution of the total number of migrants from a household for Nepal.

5. Conclusions

A mixture of two displaced geometric distribution was found a reasonable approximation to describe the distribution of the total number of migrants from a household at least at the micro-level. The exact variances and co-variance of the estimators for the model has also been computed. The proportion of households having only one migrant was found to be 0.52381. Whereas the average number of migrants from the household was 0.57831. For the development of a more effective and equitable rural and urban policies in the developing countries like Nepal, the policy planners and social researchers may get an idea from this study.

REFERENCES

[1] Aryal, T.R., *Some demographic models and their applications with reference to Nepal*, unpublished Ph.D. thesis in Statistics, Banaras Hindu University, India (2002)

[2] Aryal, T.R., *Probability models for the number of rural out-migrants at micro-level*, *The Nepali Mathematical Sciences Reports*, Vol. 21 (1&2) (2003), 9–18.

[3] Banerjee B., *Rural urban migration and urban labour market*, Himalaya Publishing House, New Delhi, India, (1986).

[4] Iwunor C.C.O. *Estimating of parameters of the inflated geometric distribution for rural out-migration*, GENUS, Vol. LI (3–4), (1995).

[5] Sharma H.L., *A probability distribution for rural out-migration, at micro-level, Rural Demography*, Vol. 5(2), (1985).

[6] Sharma H.L., *A probability distribution for rural out-migration, Janasamkhya-A Journal of Demography*, Vol. 12(1&2), (1987).

[7] Singh S,N., and K.N.S. Yadava, *Trends in rural out-migration at household level, Rural Demography*, Vol. 8, (1981).

[8] Singh S.R.J., *A study of rural out-migration and its effects on fertility*, unpublished Ph.D. thesis in statistics, Banaras Hindu University, India, (1985).

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

[10] Yadava K.N.S. and S.R.J. Singh, *A model for the number of rural out-migrants at household level, Rural Demography*, Vol. 10, (1983).

[11] Yadava K.N.S., S. Tripathi and V.S. Singh, *A probability model for the total number of migrants from the household*: An alternative approach, *Journal of Scientific Research*, Banaras Hindu University, Vol. 44c, (1994).

[12] Yadava S.N., *On some migration and population growth models*, unpublished Ph.D., thesis in Statistics, Banaras Hindu University, India, (1993).

TIKA RAM ARYAL
Central Department of Statistics
Tribhuvan University,
Kirtipur, Kathmandu, Nepal.
E-mail: traryal@rediffmail.com