Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/070108

Application of Size-Biased Geometric Distribution to Migration Data

Raj Kumar Yadav[1,*] *and Upendra Kumar*[2]

Department of Statistics, UdaiPratap Autonomous College, Varanasi, Uttar Pradesh, India

Received: 23 Oct. 2017, Revised: 21 Dec. 2017, Accepted: 23 Dec. 2017
Published online: 1 Mar. 2018

Abstract: Many mathematical and probabilistic models have been used in social sciences to explain the observed phenomenon in a summarizing form. In migration studies, it was observed that in the area under study all the households are not equally exposed to the risk of migration at a particular point of time and having at least one migrant. The present study shows the application of size biased geometric distribution to describe the observed phenomenon related to number of migrants and it was found that the models describes the phenomenon satisfactorily well.

Keywords: Rural out migration, size biased geometric distribution, chi-square

1 Introduction

Aggregate variation in population with connection to various social demographic and spatial characteristics based on data from districts, states and nation as a whole is called migration. The study of migration taken at community, village, household or individual level is done under micro level studies and it is based on objective and availability of data. In a developing country like India, where still about seventy per cent population lives in villages, migration from rural to urban or rural to rural areas has become a major concern of interest for demographers and social scientists as well as for policy makers.

Generally migration from rural areas is classified in three categories. In first category those household were kept from where only adult males migrate leaving their family in villages. Second kind of households was those from where adult male members migrate with their wives and children. The third kind household was mixture of the above two. The study of migration from rural areas at micro level has been conducted keeping in mind the above classification.

Many mathematical and probabilistic models have been used in social sciences to explain the observed phenomenon in a summarizing form. A model can explain the pattern, trend and volume of the demographic factors and thus is of great importance in study. Many stochastic models have been proposed and modified by the researchers to study the migration process. Different types of models have been proposed for migration studies based on the type of migration discussed above. The distribution of adult male migrants in a household has been explained by negative binomial distribution. After this a number of modifications and new models have been proposed to study the pattern of rural out migration (Hossain, 2000; Singh, 1984; Singh et.al, 2014, 2015, 2016; Sharma, 1988; Yadava and Singh, 1983). Later on several attempts have been made to explain the distribution households according to the total number of migrants based on assumptions (Yadava et. al, 1989, 1994; Pandey, 1993).

Most of the researchers had used the moment method as well as the mle method to estimate the parameters of the models. Some of the researchers had also used the method of mean zero frequently method (equating observed and theoretical zeros cell frequencies and means).

* Corresponding author e-mail: arkphd86@gmail.com

The idea of weighted distribution has been introduced by Fisher (1934) to explain the real life phenomenon more effectively. After some years this idea was extended by Rao (1965) and this area of research gained momentum. The size biased geometric distribution is a particular form of weighted geometric distribution.

In migration studies, it was observed that in the area under study all the households are not equally exposed to the risk of migration at a particular point of time and having at least one migrant. The main objective of the paper is to show the application of size biased geometric distribution to describe the observed phenomenon related to number of migrants. The estimates of the parameters have been estimated by method of maximum likelihood.

2 Data

The above discussed model is verified by the data set collected in a primary survey entitled Migration and its Impact in Rural Varanasi has been conducted during September - October 2015 in four villages of two blocks of Varanasi District by the Department of Statistics, U. P. Autonomous College, Varanasi. Apart from this the data set collected during October 2009 to June 2010 in Northeastern Bihar has also been used to show the applicability of the model(Singh et. al, 2013). One more data set collected during 2000-01 in Varanasi and Chandauli district has also been used for comparison purpose (Shukla, 2007). The geographical as well as social and economic condition of both the places i.e. Varanasi and northeaster Bihar is different. The education level of northeastern Bihar is very low (about 53 per cent) while in Varanasi it is more than 75 percent. Northeastern Bihar is a flood hit area. Varanasi provides many job opportunities to the locals of the surrounding district while in Northeastern Bihar the migration rate of labors is very high.

3 The Model

Suppose x is a non-negative random variable with probability function f(x,θ), $\theta \varepsilon \Omega$, the parameter space. Further, assume that x of X under f(x,θ) enters the investigation records with probability proportional to $\omega(x,\beta)$ Here the recording function $\omega(x,\beta)$ is a non-negative function with parameter β which denotes the recording mechanism. $\omega(x,\beta)$ is also known as weight function. Then x is not an observation on X but on the random variable X^w having the probability function:

$$f^w(x,\beta,\theta) = \frac{\omega(x,\beta) f(x,\theta)}{\omega} \tag{1}$$

Where ω represents the normalizing factor obtained to make the total probability equal to unity by taking $\omega = E[\omega(x,\beta)]$ The random variable X^w is the weighted version of X and its distribution is called the weighted distribution with the weight function ω When $\omega(x,\beta) = x$, $X^w = X$ is called the size biased of X.
Consequently, the distribution of x is called the size-biased distribution with the probability function as:

$$f*(x,\theta) = \frac{x f(x,\theta)}{\mu} \tag{2}$$

Where $\mu = E(x)$ f* is the size biased form of f and the corresponding sighting mechanism is called the size-biased sampling.
When the random variable x follows geometric distribution, the distribution of weighted version of x is called the weighted geometric distribution. Its size biased form is known as the size biased geometric distribution (SBGD).
The probability mass function of geometric distribution is given as (1.3) Here E(x) = q/p And variance =q/p2 Its size biased form denoted by f*(x) is obtained as follows:

$$p(X=x) = \{q^x p \qquad where \quad x = 0,1,2,\ldots \quad 0 \le p \le 1 \tag{3}$$

$$0, \qquad Otherwise \tag{4}$$

Here $E(X) = \frac{q}{p}$
and $V(x) = \frac{q}{p^2}$
Its size biased form denoted by f*(x) is obtained as follows
$p(X=x) = f*(x) = \frac{xf(x)}{E(x)}$
$f*(x) = xp^2q^{x-1} \qquad x = 0,1,2,\ldots$

4 Estimation

Method of Maximum Likelihood:
The likelihood function of (5) is given as

$L = \sum_{i=1}^{n} x_i p^{2n} q^{\sum_{i=1}^{n} xi-n}$

After taking log of both the sides, we have

$$logL = \sum_{i=1}^{n} logx_i + 2nlogp + (\sum x_i - 1)log(1-p) \quad (5)$$

Differentiating with respect to p, we get

$\frac{\theta}{tap}L = 0 + \frac{2n}{p} + \frac{\sum X_i - n}{(1-p)}(-1) = 0$

$\frac{2n}{p} = \frac{\sum X_i - n}{(1-p)}$

or $2n(1-p) = \sum X_i p - np$

or $2n - 2np = \sum X_i p - np$

or $2n = \sum X_i p + np$

or $\hat{p} = \frac{2}{x+1}$

From equation (5), we get

$\overline{x} = E * (x) = 1 + \frac{2q}{p}$

Table 1: Observed and expected number of households having adult male migrants aged 15 and above

Number of Migration	Number of Households	
	observed	Expected
1	97	93
2	42	48
3	16	18
4	7	6
5+	5	2
Total	167	167
P	0.7438	
	4.5065	
Df	2	

5 Application

The suitability of any proposed or modified model can be checked by observing the idea behind its construction. If empirical as well as conceptual aspects of the modification of construction of any model is justifiable the real aim is achieved.

Here the discussed size biased geometric model has been applied on the distribution of households having at least one adult male migrant aged 15 years and above. In table 1, the model has been applied on the primary data collected by the researcher in Varanasi district.

The value of chi-square is not significant at 5 per cent level of significance. This shows that the discussed model provides a reasonable approximation for the adult rural out migration. Figure 2 reveals the same result.

Table 2 portrays that the model is a good approximation for the data of north eastern Bihar also. The value of chi square is not significant at 5 per cent level of significance. The figure 2 gives the same results.

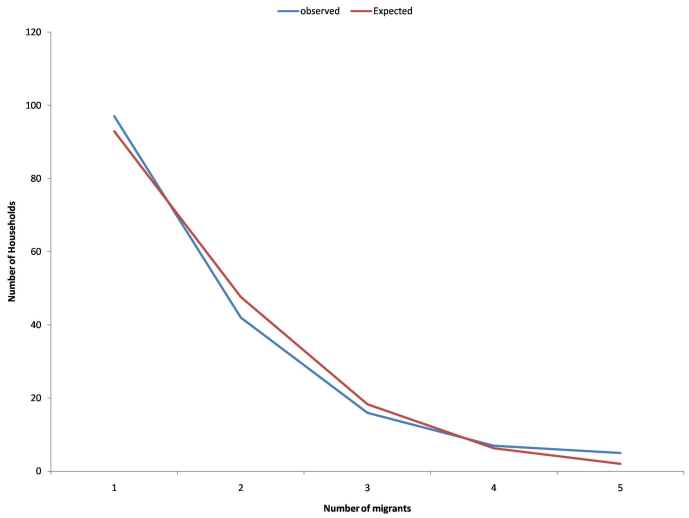

Fig. 1: Observed and expected number of households having adult male migrants aged 15 and above

In table 3 the model has been applied on the data collected during 2001 (Shukla 2007) in Varanasi district. The results show that the model fits satisfactorily well on the data. The value of chi square is not significant at 5 per cent level of significance.

6 Conclusion

The study shows that the discussed model is a good approximation for the adult male rural out migration. The importance of any model is that whenever same condition prevails in any part of society, we should not do same exercise again. We simply apply the model and get the desired estimates. It not only saves our time but also saves the money and other resources. Thus keep this in mind the authors believe that this model can be used for further studies and it would be helpful in policy making as well.

Table 2: Observed number of households having adult male migrants aged 15 and above in north eastern Bihar

Number of Migration	Number of Households	
	observed	
1	95	88
2	41	49
3	15	21
4	12	8
5	6	3
Total	169	169
p df	0.7206 5.87 2	

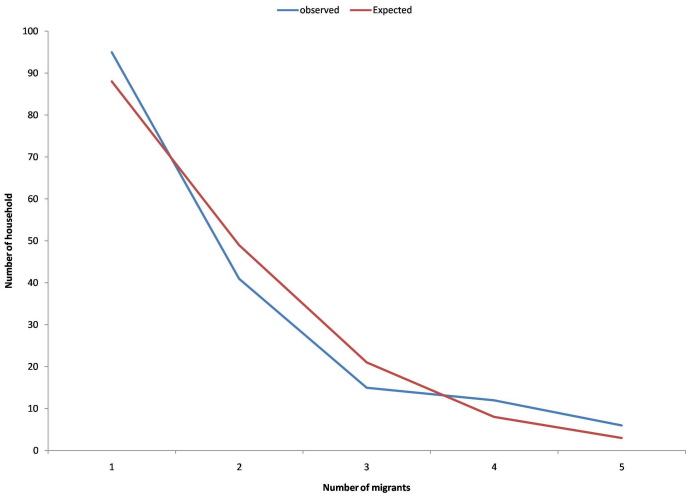

Fig. 2: Observed and expected number of households having adult male migrants aged 15 and above in north eastern Bihar

Number of Migration	Number of households	
	observed	Expected
1	97	94
2	35	45
3	19	16
4+	9	5
Total	160	160
p	0.7619	
	5.82	
df	2	

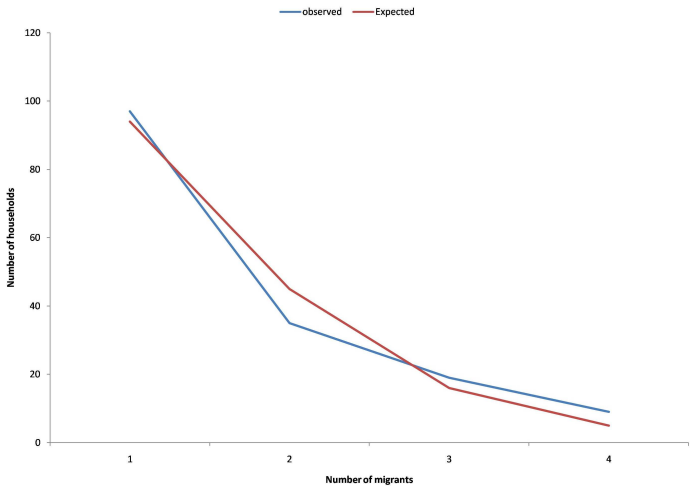

Fig. 3: Observed and expected number of households having adult male migrants aged 15 and above (survey 2001)

References

[1] Hossain M.Z. (2000), Some Demographic Models and Their Application with Special Reference to Bangladesh., , unpublished Ph.D. thesis, Banaras Hindu University, India.
[2] Singh, B. P., Singh, N. K., & Roy Tapan, 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.
[3] Singh, B. P., Singh, N. K., & Dixit, S. (2015). Estimation of Parameters of Inflated Poisson-Lindley Distribution for Adult out-Migration. Journal of Institute of Science and Technology, 20(2), 6-10.
[4] Singh, B. P., & Singh, N. K. (2016).On the Distribution of Risk of Migration and Its Estimation. International Journal of Statistical Distributions and Applications, 2(4), 67-71.
[5] Singh, S. N., & Yadava, K. N. (1981).Trends in Rural Out-migration at Household Level.Rural Demography, 8(1), 53.
[6] Sharma, H. L. (1987). Estimating the Parameters of a Polya-AeppliDistribution. Rural Demography, 15(1-2), 1-6.
[7] Yadava, K. N. S., & Singh, S. R. J. (1983).A Model for the Number of Rural Out-migrants at Household Level. Rural Demography, 10(1-2), 23-33.
[8] Yadava, K. N. S., Tripathi, S., & Singh, V. S. (1994). A Probability Model for the Total Number of Migrants from the Household: An Alternative Approach. Journal of Scientific Research, Banaras Hindu University, 44, 125-134.
[9] Pandey, H. (1993).On Human Out-migration Model.Jansmkhya, 11 (2), 123-127.
[10] Yadava, K. N., Singh, S. K., & Kumar, U. (1989).A Probability Model for the Number of Migrants.Janasamkhya, 7(2), 83-92. Fisher, R. A. (1934). The Effect of Methods of Ascertainment upon the Estimation of Frequencies. Annals of Human Genetics, 6(1), 13-25.
[11] Rao, C. R. (1965). On Discrete Distributions Arising out of Methods of Ascertainment.Sankhy: The Indian Journal of Statistics, Series A, 311-324.
[12] Shukla, K. K., & Yadava, K. N. S. (2007).The Distribution of the Number of Migrants at the Household Level. Journal of Population and Social Studies [JPSS], 14(2), 153-166.
[13] Singh, N. K., Singh, B. P., Yadava, K. N. S., & Singh, B. P. (2013). A Study of Socio Demographic Condition of North eastern Bihar.Recent Research in Science and Technology, 5(4).

Raj Kumar Yadav is a research Scholar in the department of Statistics, U.P. Autonomous College, Varanasi, India. He is pursuing Ph.D. in applied Statistics. His research field is in the area of Demography. He has published three research papers in different journals.

Upendra Kumar is a senior Assistant professor of Statistics in U.P. Autonomous College, Varanasi, India. He obtained his Ph.D. degree in 1991 from Banaras Hindu University, Varanasi, India. He has worked for total literacy mission in Uttar Pradesh under Govt. of India for 10 years. He has worked as project officer in the department of non-formal education for two years. He has 07 years of teaching experience in higher secondary education and more than 09 years of teaching experience in higher education. He has published 12 research papers in various journals.